IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF NEW JERSEY

MONIB ZIRVI, M.D., Ph.D.	)	
Plaintiff,	)	
v.	)	Case No.
	)	
ILLUMINA, INC., THERMO FISHER	)	**PLAINTIFF DEMANDS**
SCIENTIFIC, AKIN GUMP STRAUSS	)	**TRIAL BY JURY**
HAUER & FELD LLP,	)	
LATHAM & WATKINS,	)	
RIP FINST, SEAN BOYLE,	)	
MATTHEW A. PEARSON,	)	
ANGELA VERRECCHIO,	)	
ROGER CHIN, and	)	
DOUGLAS LUMISH	)	
	)	
Defendants.	)	
_____	)	

COMPLAINT

Plaintiff, Dr. MONIB ZIRVI, by and through its undersigned counsel hereby sues Defendants, Illumina, Inc., ThermoFisher Scientific, Akin Gump Strauss Hauer & Feld LLP, Latham & Watkins, Rip Finst, Sean Boyle, Matthew A. Pearson, Angela Verrecchio, Roger Chin and Douglas Lumish, and states the following in support thereof:

THE PARTIES

1. Dr. MONIB ZIRVI (hereinafter "Plaintiff") is an individual who has developed several trade secrets and intellectual property in the field of DNA diagnostics and DNA arrays, including the property at issue in this case and, at all times relevant, has been a practicing

physician-scientist with his principal location at 1 Diamond Hill Road, Berkeley Heights, NJ 07922 in Union County.

2. Defendant Illumina, Inc., is a corporation of the State of Delaware with a principal place of business in the State of California.

3. Defendant ThermoFisher Scientific is a corporation of the State of Delaware with a principal place of business in the State of California.

4. Defendant Attorneys Rip Finst and Sean Boyle of THERMO FISHER SCIENTIFIC INC., Matthew A. Pearson and Angela Verrecchio, of AKIN GUMP STRAUSS HAUER & FELD LLP, Roger Chin and Douglas Lumish of LATHAM & WATKINS are attorneys whose principal places of business are in the States of Pennsylvania and California.

JURISDICTION AND VENUE

5. This Court has exclusive jurisdiction to hear Court I pursuant to 28 U.S.C. § 1338(a): "The district courts shall have original jurisdiction of any civil action arising under any Act of Congress relating to patents…. No State court shall have jurisdiction over any claim for relief arising under any Act of Congress relating to patents …."

6. This Court has supplemental jurisdiction to hear the remaining Counts II–IV under 28 U.S. Code § 1367 (a) (2022) ("… in any civil action of which the district courts have original jurisdiction, the district courts shall have supplemental jurisdiction over all other claims that are so related to claims in the action within such original jurisdiction that they form part of the same case or controversy under Article III of the United States Constitution.")

7. Venue is governed by 28 U.S. Code § 1391(a) (2) proper in this case because the judicial district in which a substantial part of the events or omissions giving rise to the claim occurred in this district, and a substantial part of the property that is the subject of the action is situated in this district.

8. The acts complained of herein were directed against Plaintiff who, at all times relevant, has been a practicing physician-scientist with his principal location at 1 Diamond Hill Road, Berkeley Heights, NJ 07922 in Union County.

GENERAL ALLEGATIONS

I. Factual Background

9. Plaintiff was a graduate student in the laboratory of Dr. Francis Barany at Cornell University Medical College (now known as Weill Cornell Medicine) from August 1994 to January 1999.

10. During that time, the Barany laboratory was developing new technology in DNA diagnostics, including new methods to detect cancer mutations to improve patient cancer care.

11. After graduating from the laboratory, Plaintiff, on his own time, using his equipment, relying on his knowledge of computer programming, independently developed a data set of ZipCode sequences to enable mass production and manufacture of DNA computer chips known as DNA microarrays. The ZipCode sequences he invented were protectable as patentable subject matter and as a trade secret.

12. ZipCode sequences are necessary to identify cancer mutations and other DNA applications, analogous to developing the operating system software for personal computers. (MS-DOS, iOS, and Windows)

13. Plaintiff's intent was to incorporate the Zip Code Operating System (Zip Code Chemistry) into the manufacture of DNA microchips and receive royalties through the licensing of the invention.

14. The Zip Code Operating System invented independently by Plaintiff, and other work done at the lab, was incorporated into Illumina patents, products, and SEC filings. The key

difference is that Illumina, including its founders, renamed the Zirvi-Barany Zip Code Operating System as “Tag Sequences”, “Illumacodes”, “Illumicodes”, or other misleading names. All the while, Illumina’s source code of their software identified them as what the truly were, Plaintiff’s “ZipCodes.”

15. Illumina’s illicit taking became the subject matter of multiple litigations between Cornell University, Cornell Research Foundation, Inc., Life Technologies Corporation, and Applied Biosystems, LLC versus Illumina.

16. Throughout the litigation the Plaintiff was in communication with, and advised extensively by Attorneys for Cornell University, Cornell Research Foundation, Inc., PE Applied Biosystems, LLC, Life Technologies Corporation, and Thermo Fisher Scientific Inc., who represented to the Plaintiff that they were his counsel as well.

17. At various points throughout the Cornell litigation, the Attorneys representing the Plaintiff included, **Rip Finst** and **Sean Boyle** of THERMO FISHER SCIENTIFIC INC., **Matthew A. Pearson** and **Angela Verrecchio**, of AKIN GUMP STRAUSS HAUER & FELD LLP, **Roger Chin** and **Douglas Lumish** of LATHAM & WATKINS. (Dr. Zirvi did not have direct contact with non-party attorneys Dianne B. Elderkin, Rachel J. Elsby, and Jason E. Weil of AKIN GUMP STRAUSS HAUER & FELD LLP).

18. Plaintiff’s Attorneys advised him that his interests were aligned with Cornell University, Cornell Research Foundation, Inc., PE Applied Biosystems, LLC, Life Technologies Corporation, and Thermo Fisher Scientific Inc., and that these attorneys represented his interests.

19. The Attorneys advised Plaintiff to not prepare for his deposition, to not review any documents, including his own patent filings, and not research facts related to Illumina, its founders, employees, and patent filings.

20. Instead, the Attorneys instructed and advised Plaintiff to answer questions with "I don't know", and "I don't remember", all explained to him as the best way to benefit Cornell and his rights against Illumina's illicit taking.

21. Attorneys for the Plaintiff withheld from Plaintiff that ThermoFisher was secretly collaborating with Illumina to develop "Ampliseq for Illumina" during the entire time they were representing Plaintiff – a knowingly deliberate conflict of interest. The secret collaboration, which upon information and belief started at least three years prior, was publicly admitted to having occurred at least a year prior to January 30th, 2018, while the *Cornell v. Illumina* (1:10-cv-00433-LPS) case was still active. See, this Complaint's Exhibits 1 and 2. (respectively, Illumina_SEC_8k_Ampliseq_for_Illumina_01_30_2018_v02Hi copy 2 and see Exhibit 13 of Zirvi v US NIH et al. Case 3:20-cv-07648-MAS-DEA: Bringing together two leaders: AmpliSeqTM for Illumina. https://www.youtube.com/watch?v=lWPaZX1TDa4).

22. Plaintiff subsequently learned that other inventors from Patent application WO 97/31256 were instructed to do the same.

23. Attorneys for the Plaintiff repeatedly requested via phone calls, video teleconferences, and emails for Plaintiff to submit his expert analysis and findings that would strengthen the *Cornell v Illumina* (1:10-cv-00433-LPS) case, and Plaintiff spent hundreds of hours preparing such confidential analysis over a two-year period.

24. Attorneys for ThermoFisher informed Plaintiff at the end of January 2017-beginning of February 2017 that they were unhappy with the performance of Akin Gump, and thus were immediately bringing in Roger Chin and Doug Lumish of Latham & Watkins to take over the Cornell v Illumina case (1:10-cv-00433-LPS). None of these Attorneys informed Plaintiff of

their conflict of interest in they were simultaneously involved in a parallel and related case of Illumina v Life Technologies Corp (Thermo Fisher) Case 3:11-cv-03022-JAH-DHB.

25. Between January 2017 and April 2017, Plaintiff sent Attorneys at ThermoFisher, Latham & Watkins and Akin Gump numerous emails containing bullet-proof evidence of fraud by Illumina which would have proven the claims by Cornell, yet such evidence was kept from the court in the *Cornell v Illumina* case (1:10-cv-00433-LPS). Plaintiff requested to file a declaration to introduce this evidence to the court, yet inexplicably, Attorneys denied this request. (See Exhibit 12 for Draft of Declaration written by Plaintiff and emailed about to Defendants including Rip Finst, Matthew Pearson, Roger Chin and Douglas Lumish shortly prior to the signing of the settlement agreement in *Cornell v Illumina*. Plaintiff specifically requested for their help to edit and submit this to the US District Court in Delaware as a Third Party with an interest.)

26. Unbeknownst to Plaintiff or Cornell at the time, *Cornell v Illumina* (1:10-cv-00433-LPS) was fraudulently settled by Illumina and ThermoFisher, simultaneously with several other lawsuits involving Illumina and ThermoFisher, in April 2017. Tellingly, Attorneys Roger Chin and Doug Lumish were working on the other cases involving Illumina and ThermoFisher but failed to disclose their obvious conflict of interest to either Plaintiff or Cornell. The settlements, while benefiting Illumina and ThermoFisher, completely undermined the rights of the Plaintiff.

27. Plaintiff discovered through a review of public SEC filings that Illumina had filed a First Amendment Agreement to raise funds for many years. In this document, the definition of "Tag Sequences" was redacted. When asked, Matthew Pearson said he knew about the First Amendment Agreement, and stated it had nothing to do with the inventors, including Plaintiff, as well as their rights as inventors under the *Cornell v. Illumina* suit. Matthew Pearson refused to share any details of the First Amendment Agreement, requiring Plaintiff to obtain a copy of the

unredacted First Amendment Agreement through a FOIA request to the SEC, which was received by Plaintiff on May 17, 2017 (Exhibit 3).

28. In the redacted document, everything passed the words " 'Tag Sequence' means " was redacted. In the unredacted document, it continues as follows: "Tag Sequence" means a set of oligonucleotide probes, developed pursuant to the Original Agreement or this First Amendment, which act independently of any target-sequence-specific analytical chemical reactions to allow the physical addressing of the products of a chemical reaction to locations on a solid support, such as the "addressable array-specific portion" of the oligonucleotide probes and their complements described in International Patent Application No. W097/31256 and that are designed for use in the Collaboration Product. The Parties will agree on the selection of Tag Sequences to be used in the Collaboration Product, subject to the approval of the Joint Steering Committee.

29. International Patent Application No. W097/31256 is an invention that was submitted before Illumina even existed, with Dr. Zirvi as a coinventor. In the First Amendment Agreement, it states: "The Parties will share responsibility for defining and developing Tag Sequences for the Collaboration Product which will attempt to avoid third party intellectual property rights or other encumbrances." In other words, this was collusion by Illumina and ThermoFisher to apparently defraud third parties, such as Cornell and the Plaintiff. Dr. Zirvi informed Cornell of the findings in the unredacted First Amendment Agreement.

30. Cornell, when it became aware of the apparent fraud and collusion between these two companies, filed a Rule 60(b)(6) motion to overturn this settlement agreement, attached hereto as Exhibit 4.

31. This motion was vigorously opposed by both Illumina and ThermoFisher, even though ThermoFisher was a Plaintiff allegedly representing Cornell's and Plaintiff's interests in the case against Illumina.

32. Unbeknownst to Plaintiff or Cornell at the time, Illumina and ThermoFisher were secretly working together to develop the "Ampliseq for Illumina" product line during the Cornell v Illumina case. (See Exhibits 1 and 2).

33. Illumina and ThermoFisher's joint venture in the "Ampliseq for Illumina" product line strengthened their duopoly in the DNA sequencing and DNA microarrays market.

34. Illumina has used its monopoly position in DNA sequencing to purchase Grail and claim that they have developed a "multi-cancer early detection" (MCED) test claiming to find 50 different cancers at early stages, without any evidence of long-term prospective population base clinical trial to prove clinical utility for this test.

35. Despite this fact, Illumina / Grail is attempting to unduly influence lawmakers to pass bills (117 HR 1947, To amend title XVIII of the Social Security Act to provide for Medicare coverage of multi-cancer early detection screening tests; 117 S 1873, To amend title XVIII of the Social Security Act to provide for Medicare coverage of multi-cancer early detection screening tests.) requiring Medicare to reimburse for this unproven test, misspending up to $60B in taxpayer funds per year.

36. Moreover, under the guise of this "testing" to detect "early cancer" patients would unwittingly be providing all their private DNA information, which Illumina could resell for their own personal profit regardless of the end user's purpose.

37. At the foundation of the DNA analysis market is Plaintiff's Zip Code Operating System that allows samples to be processed and reliable results obtained.

38. The Plaintiff has not been recognized as an inventor on any of Illumina's patents using the Zip Code Operating System; has not received royalties from Illumina's use of Plaintiff's Zip Code Operating System in their software and products Others have noted: "To paraphrase William Shakespeare's Hamlet, something is rotten in the state of Illumina."[1]

39. Illumina and its founders and employees have incorporated the ZipCode Operating System and utilized Plaintiffs' ZipCode sequences and designs to manufacture and commercialize numerous products. Illumina knowingly applied for these patents without including Plaintiff as an inventor and the patent claims would not have been issued had it not been for Illumina's commercial success using ZipCode sequences and the ZipCode Operating System to determine the location of DNA sequences in the Sentrix arrays, GoldenGate assays and Infinium arrays as demonstrated by the software used to analyze .dmap files associated with these products. This software specifically calls the DNA sequences, used by Illumina in its products, ZipCode (and not by any other name or pseudonym).

40. The patents were issued by the USPTO without Plaintiff as a named inventor solely because Illumina knowingly and fraudulently withheld this usurpation of intellectual property rights from the USPTO patent examiners (See First Amendment Agreement, Exhibit 3) by renaming ZipCode sequences as Addresses, DBL, decoders and adaptors as well as other pseudonym and not referencing Plaintiffs 465 and 4633 ZipCode sets which Kevin Gunderson and Mark Chee and Ilumina purloined as Illumacodes 1-16. (See Probes and Decoder Oligonucleotides patent application, in Expert analysis as filed in Zirvi v. US NIH et al. (Case 3:20-cv-07648-MAS-DEA) as Exhibit 17, see Exhibit 5). Illumina's "Tag Sequences" and Illumicodes and Illumacodes and DBLs and adaptors are all in fact derivatives of Plaintiff's ZipCode sequences and ZipCode

[1] Open letter to Shareholders of Illumina, Inc., March 13th, 2023, Carl Icahn: (https://carlicahn.com/open-letter-to-shareholders-of-illumina-inc/)

Operating System. As others have noted: "Yet these individuals suffer no consequences or remorse. The members of Illumina's management team and board of directors collectively own less than 0.1% of the company's stock yet they feel entitled to take these reckless actions with our money."[2]

41. During the *Cornell v. Illumina* case, Attorneys Roger Chin, Doug Lumish, and Rip Finst presented "legal advice" that included knowingly false statements and deliberate omissions in two PowerPoint presentations (Exhibits 6 and 7). Tellingly, the First Amendment Agreement was never mentioned, let alone discussed in these presentations, thus allowing for a false and misleading legal analysis, including the misleading definition of ZipCodes.

42. At the end of the *Cornell v. Illumina* case, Plaintiff was informed by Attorney Roger Chin that the protection of the purloined 16 Zip Code sequences was a trade-secret case and that Attorney Matthew Pearson had investigated bringing the action as a trade-secret case. See the email from Roger Chin addressed to Plaintiff and inventors (Exhibit 8).

43. Based on Roger Chin's legal advice, Plaintiff Zirvi filed *Zirvi v. Flatley* (Case 1:18-cv-07003-JGK). As part of *Zirvi v. US NIH et al*. (Case 3:20-cv-07648-MAS-DEA) case, Expert Analysis revealed numerous additional instances of apparent fraud (see Exhibit 5), including, but not limited to the dependence of numerous Illumina patent filings and Illumina products on Zip Code sequences (i.e. Sentrix, Infinium Arrays.). These findings were subverted by Attorneys Matthew Pearson, Roger Chin, Doug Lumish, and Rip Finst. As others have noted: "In response to our letter, Illumina's board did what boards do best. They used our money – and yours – to defend themselves against their own shareholders by hiring highly-paid bankers, lawyers and PR firms. In the process, they issued a press release and spoke to sell-side analysts to disseminate the

[2] *Id.*

incumbent board's message. They make several points which are either misleading, improbable or are clear double speak."[3]

44. At the conclusion of *Zirvi v. Flatley* legal analysis of "storm warnings" of fraud and IP theft by Illumina should have commenced by the ThermoFisher Attorneys, who at the time were representing Cornell and all the inventors at the beginning of *Cornell v. Illumina*.

45. The patents that Illumina filed that absolutely depend on ZipCodes to function include but are not limited to the following:

a. 6355431 (the '5431 patent); Detection of nucleic acid amplification reactions using bead arrays Chee; Mark S. (Del Mar, CA), Gunderson; Kevin (Encinitas, CA)

b. 6620584 (the '0584 patent); Combinatorial decoding of random nucleic acid arrays Chee; Mark (Del Mar, CA), Walt; David R. (Lexington, MA)

c. 6770441 (the '0441 patent); Array compositions and methods of making same Dickinson; Todd (San Diego, CA), Coblentz; Kenneth D. (Del Mar, CA), Carlson; Edward (Oceanside, CA)

d. 6812005 (the '2005 patent); Nucleic acid detection methods using universal priming Chee; Mark S. (Del Mar, CA), Auger; Steven R. (Norwell, MA), Stuelpnagel; John R. (Encinitas, CA)

e. 6858394 (the '8394 patent); Composite arrays utilizing microspheres Fan; Jian-Bing (San Diego, CA), Stuelpnagel; John R. (Encinitas, CA), Chee; Mark S. (Del Mar, CA)

f. 6890741 (the '0741 patent); Multiplexed detection of analytes Fan; Jian-Bing (San Diego, CA), Stuelpnagel; John R. (Encinitas, CA), Chee; Mark S. (Del Mar, CA)

[3] Icahn Responds to Illumina's Obfuscations; March 15th, 2023, Carl Icahn: (https://carlicahn.com/icahn-responds-to-illuminas-obfuscations/)

g. 6913884 (the '3884 patent); Compositions and methods for repetitive use of genomic DNA Chee; Mark S. (Del Mar, CA), Czarnik; Anthony W. (San Diego, CA), Stuelpnagel; John R. (Encinitas, CA)

h. 7033754 (the '3754 patent); Decoding of array sensors with microspheres Chee; Mark S. (Del Mar, CA), Czarnik; Anthony W. (San Diego, CA), Stuelpnagel; John R. (Encinitas, CA)

i. 7060431 (the '0431 patent); Method of making and decoding of array sensors with microspheres Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

j. 7166431 (the '6431 patent); Combinatorial decoding of random nucleic acid arrays Chee; Mark S. (Del Mar, CA), Walt; David R. (Lexington, MA)

k. 7226734 (the '6734 patent); Multiplex decoding of array sensors with microspheres Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

l. 7361488 (the '1488 patent); Nucleic acid detection methods using universal priming Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

m. 7455971 (the '5971 patent); Multiplex decoding of array sensors with microspheres Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

n. 7510841 (the '0841 patent); Methods of making and using composite arrays for the detection of a plurality of target analytes Stuelpnagel; John (Encinitas, CA), Chee; Mark (Del Mar, CA), Auger; Steven (Norwell, MA)

o. 7563576 (the '3576 patent); Combinatorial decoding of random nucleic acid arrays Chee; Mark S. (Del Mar, CA), Walt; David R. (Lexington, MA)

p. 7582420 (the '2420 patent); Multiplex nucleic acid reactions Shen; Mun-Jui Richard (Poway, CA), Oliphant; Arnold (Poway, CA), Butler; Scott L. (San Diego,

CA), Stuelpnagel; John E. (Encinitas, CA), Chee; Mark S. (Del Mar, CA), Kuhn; Kenneth M. (San Diego, CA), Fan; Jian-Bing (San Diego, CA)

q. 7611869 (the '1869 patent); Multiplexed methylation detection methods Fan; Jian-Bing (San Diego, CA)

r. 7612020 (the '2020 patent); Composite arrays utilizing microspheres with a hybridization chamber Fan; Jian-Bing (San Diego, CA), Chee; Mark S (Del Mar, CA)

s. 7670810 (the '0810 patent); Methods and compositions for whole genome amplification and genotyping Shen; Mun-Jui Richard (Poway, CA), Oliphant; Arnold (Poway, CA), Butler; Scott L. (San Diego, CA), Stuelpnagel; John E. (Encinitas, CA), Chee; Mark S. (Del Mar, CA), Kuhn; Kenneth M. (San Diego, CA), Fan; Jian-Bing (San Diego, CA)

t. 7776531 (the '6531 patent); Compositions and methods for stabilizing surface bound probes Fan; Jian-Bing (San Diego, CA)

u. 7803537 (the '3537 patent); Parallel genotyping of multiple patient samples Fan; Jian-Bing (San Diego, CA), Chee; Mark S (Del Mar, CA)

v. 7899626 (the '9626 patent); System and method of measuring methylation of nucleic acidsi Kruglyak; Semyon (San Diego, CA), Bibikova; Marina (San Diego, CA), Chudin; Eugene (Kirkland, WA)

w. 7901897 (the '1897 patent); Methods of making arrays Kain; Robert (San Diego, CA)

x. 7914988 (the '4988 patent); Gene expression profiles to predict relapse of prostate cancer Chudin; Eugene (Kirkland, WA), Lozach; Jean (San Diego, CA), Fan; Jian-Bing (San Diego, CA), Bibikova; Marina (San Diego, CA)

y. 7955794 (the '5794 patent); Multiplex nucleic acid reactions Shen; Min-Jui Richard (San Diego, CA), Oliphant; Arnold (Poway, CA), Butler; Scott L. (San

Diego, CA), Stuelpnagel; John R. (Encinitas, CA), Chee; Mark S. (Del Mar, CA), Kuhn; Kenneth M. (San Diego, CA), Fan; Jian-Bing (San Diego, CA)

z. 7960119 (the '0119 patent); Combinatorial decoding of random nucleic acid arrays Chee; Mark S. (Encinitas, CA), Walt; David R. (Lexington, MA)

aa. 8003354 (the '3354 patent); Multiplex nucleic acid reactions Oliphant; Arnold (Poway, CA), Stuelpnagel; John R. (Encinitas, CA), Chee; Mark S. (Del Mar, CA), Butler; Scott L. (San Diego, CA), Fan; Jian-Bing (San Diego, CA), Shen; Min-Jui Richard (Poway, CA)

bb. 8076063 (the '6063 patent); Multiplexed methylation detection methods Fan; Jian-Bing (San Diego, CA)

cc. 8080380 (the '0380 patent); Use of microfluidic systems in the detection of target analytes using microsphere arrays Chee; Mark S. (Encinitas, CA), Dickinson; Todd A. (San Diego, CA), Gunderson; Kevin (Encinitas, CA), O'Neil; Don (San Juan Capistrano, CA), Stuelpnagel; John R. (Encinitas, CA)

dd. 8110363 (the '0363 patent); Expression profiles to predict relapse of prostate cancer Chudin; Eugene (Kirkland, WA), Lozach; Jean (San Diego, CA), Fan; Jian-Bing (San Diego, CA), Bibikova; Marina (San Diego, CA)

ee. 8150626 (the '0626 patent); Methods and compositions for diagnosing lung cancer with specific DNA methylation patterns Fan; Jian-Bing (San Diego, CA), Bibikova; Marina (San Diego, CA)

ff. 8150627 (the '0627 patent); Methods and compositions for diagnosing lung cancer with specific DNA methylation patterns Fan; Jian-Bing (San Diego, CA), Bibikova; Marina (San Diego, CA)

gg. 8206917 (the '6917 patent); Combinatorial decoding of random nucleic acid arrays Chee; Mark S. (Encinitas, CA), Walt; David R. (Lexington, MA)

hh. 8288103 (the '8103 patent); Multiplex nucleic acid reactions Oliphant; Arnold (Sunnyvale, CA), Stuelpnagel; John R. (Encinitas, CA), Chee; Mark S.

(Encinitas, CA), Butler; Scott L. (Sandwich, GB), Fan; Jian-Bing (San Diego, CA), Shen; Min-Jui Richard (Poway, CA)

ii. 8440407 (the '0407 patent); Gene expression profiles to predict relapse of prostate cancer Chudin; Eugene (Kirkland, WA), Lozach; Jean (San Diego, CA), Fan; Jian-Bing (San Diego, CA), Bibikova; Marina (San Diego, CA)

jj. 8460865 (the '0865 patent); Multiplex decoding of array sensors with microspheres Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

kk. 8481268 (the '1268 patent); Use of microfluidic systems in the detection of target analytes using microsphere arrays Chee; Mark S. (Encinitas, CA), Dickinson; Todd A. (San Diego, CA), Gunderson; Kevin (Encinitas, CA), O'Neil; Don (San Juan Capistrano, CA), Stuelpnagel; John R. (Encinitas, CA)

ll. 8486625 (the '6625 patent); Detection of nucleic acid reactions on bead arrays Chee; Mark S. (Encinitas, CA), Walt; David R. (Lexington, MA)

mm. 8541207 (the '1207 patent); Preservation of information related to genomic DNA methylation Gormley; Niall Anthony (Nr. Saffron Walden, GB), Smith; Geoffrey Paul (Nr Saffron Walden, GB), Bentley; David (Nr Saffron Walden, GB), Rigatti; Roberto (Nr Saffron Walden, GB), Luo; Shujun (Hayward, CA)

nn. 8563246 (the '3246 patent); Combinatorial decoding of random nucleic acid arrays Chee; Mark S. (Encinitas, CA), Walt; David R. (Lexington, MA)

oo. 8628952 (the '8952 patent); Array kits and processing systems Lebl; Michal (San Diego, CA), Perbost; Michel (San Diego, CA), DeRosier; Chad F. (San Diego, CA), Nibbe; Mark J. (San Diego, CA), Burgett; Steve R. (San Diego, CA), Heiner; David L. (San Diego, CA)

pp. 8741630 (the '1630 patent); Methods of detecting targets on an array Dickinson; Todd (San Diego, CA), Coblentz; Kenneth D. (San Diego, CA), Carlson; Edward (San Diego, CA)

qq. 8795967 (the '5967 patent); Multiplex decoding of array sensors with microspheres Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

rr. 8883424 (the '3424 patent); Use of microfluidic systems in the detection of target analytes using microsphere arrays Chee; Mark S. (Encinitas, CA), Dickinson; Todd A. (San Diego, CA), Gunderson; Kevin (Encinitas, CA), O'Neil; Don (San Juan Capistrano, CA)

ss. 8895268 (the '5268 patent); Preservation of information related to genomic DNA methylation Kester; Henri A. (San Diego, CA)

tt. 8906626 (the '6626 patent); Multiplex nucleic acid reactions Oliphant; Arnold (Sunnyvale, CA), Stuelpnagel; John R. (Santa Barbara, CA), Chee; Mark S. (Encinitas, CA), Butler; Scott L. (Sandwich, GB), Fan; Jian-Bing (San Diego, CA), Shen; Min-Jui Richard (Poway, CA)

uu. 9045796 (the '5796 patent); Methods and compositions for whole genome amplification and genotyping Gunderson; Kevin (Encinitas, CA), Stuelpnagel; John R. (Encinitas, CA), Chee; Mark S. (Encinitas, CA), Fan; Jian-Bing (San Diego, CA)

vv. 9163283 (the '3283 patent); Combinatorial decoding of random nucleic acid arrays Chee; Mark S. (Encinitas, CA), Walt; David R. (Lexington, MA)

ww. 9279148 (the '9148 patent); Detection of nucleic acid reactions on bead arrays Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

xx. 9289766 (the '9766 patent); Use of microfluidic systems in the detection of target analytes using microsphere arrays Chee; Mark S. (Encinitas, CA), Dickinson; Todd A. (San Diego, CA), Gunderson; Kevin (Encinitas, CA), O'Neil; Don (San Juan Capistrano, CA)

yy. 9399795 (the '9795 patent); Multiplex decoding of array sensors with microspheres Chee; Mark S. (Del Mar, CA), Stuelpnagel; John R. (Encinitas, CA), Czarnik; Anthony W. (San Diego, CA)

zz. 9441267 (the '1267 patent); Detection of nucleic acid reactions on bead arrays Oliphant; Arnold (Sunnyvale, CA), Stuelpnagel; John R. (Santa Barbara, CA), Chee; Mark S. (Encinitas, CA), Butler; Scott L. (Sandwich, GB), Fan; Jian-Bing (San Diego, CA), Shen; Min-Jui Richard (Poway, CA).

46. But for the intentional acts and negligence of the attorneys representing Plaintiff in the Cornell case the Plaintiff would have been able to enforce his intellectual property rights against Illumina.

47. But for the conspiring of all Defendants, Plaintiff Zirvi would have received the recognition and royalties on a series of patents (See: *Czarnik v. Illumina*, Case 1:05-cv-00400-JJF).

48. Plaintiff asserts that the failure to name Plaintiff as an inventor on the patents filed by the assignee has caused significant reputational harm. Plaintiff's contributions to the development of the technology are significant and cannot be denied. The claimant has been acknowledged as a key contributor to the development of the technology. (See Exhibit 9 Affidavit of Dr. Francis Barany) However, the failure to name the claimant as an inventor on the patents has undermined the claimant's reputation and standing in the scientific community.

49. The Plaintiff's claim is based on 35 U.S.C. § 256, which provides for correction of inventorship in patents. Plaintiff believes that the assignee erred in failing to name Plaintiff as an inventor on the patents and that this error should be corrected.

50. Exhibit 10 is the original provisional application 60/180,810 submitted by Jian-Bing Fan for Illumina on Feb. 07, 2000. This provisional matured into the following patents which incorporate inventive concepts and steps of the Plaintiff: the '9148 patent, the '1267 patent, the

'3354 patent, the '8103 patent, the '6626 patent, the '0536 patent, the '2005 patent, the '1488 patent, the '3754 patent, the '0431 patent, the '5796 patent, the '6734 patent, the '5971 patent, the '0865 patent, the '9795 patent, the '5794 patent, the '0741 patent, the '1869 patent, and the '6063 patent.

51. In the 60/180,810 provisional application JB Fan claims inventorship of Zip Code arrays (Universal Arrays) and zip-code sequences, as well as their combined use with PCR and ligation reactions. On the cover sheet, the box at the bottom was knowingly left unchecked for "Additional inventors are being named on separately numbered sheets attached hereto." The provisional application contains an NIH grant application within, wherein the last Figure (page 50) titled "Detection of Alternative Splicing on Zip-code Arrays and Comparison with RT-PCR" shows the use of Plaintiff's Zip Code Array invention and completely dependent on Zip-code sequences to work. Tellingly, this smoking gun figure is missing from these issued patents, including US 6,812,005 patent ('2005), as well as US 7,955,794 patent ('5794). (See Exhibit 11)

52. When Defendant Matthew Pearson was asked as to why Jian Bing Fan was not deposed in *Cornell v Illumina*, he stated that Illumina had taken him off the deposed witness list, even though a subpoena had been served and accepted by Illumina. Illumina stated that they "no longer knew where he was." Plaintiff was able to track Jian Bing Fan's fraudulent "exodus" to China with evidence of the assistance and knowledge of Illumina. (Exhibit 13) Despite being presented with all this information, Defendants Rip Finst, Matthew Pearson, Roger Chin and Douglas Lumish inexplicably made no attempt to contact or depose such a key witness, yet alone notify the Court of this obstruction of justice. This is further evidence of legal malpractice as other key witnesses including Mark Chee and David R. Walt were also not deposed. This is well below

the acceptable standard of legal practice in an intellectual property case involving multiple billions of dollars of revenue.

53. Illumina has used US 7,955,794 patent to extract a $26.7 million judgment from Roche (Verinata Health, Inc. v. Ariosa Diagnostics, Inc., Case No. 12-cv-05501-SI), showing the value of the Patents derived from the Plaintiff's inventive steps and contributions. The Defendants have all aided and abetted Illumina in its journey towards "a new low in corporate governance."[4]

54. An analysis of the merits of legal malpractice claims has been performed by Jeff Oster (See Affidavit of Merit, Exhibit 14).

COUNT I – FAILURE TO NAME PLAINTIFF ON PATENT

55. Plaintiff repeats and re-alleges all of the paragraphs 1-54 above as if fully set forth herein.

56. The Defendant, Illumina, has obtained the following issued patents which incorporate inventive steps derived from Plaintiff's intellectual property and copyrights. The '5431 patent, the '0584 patent, the '0441 patent, the '2005 patent, the '8394 patent, the '0741 patent, the '3884 patent, the '3754 patent, the '0431 patent, the '6431 patent, the '6734 patent, the '1488 patent, the '5971 patent, the '0841 patent, the '3576 patent, the '2420 patent, the '1869 patent, the '2020 patent, the '0810 patent, the '6531 patent, the '3537 patent, the '9626 patent, the '1897 patent, the '4988 patent, the '5794 patent, the '0119 patent, the '3354 patent, the '6063 patent, the '0380 patent, the '0363 patent, the '0626 patent, the '0627 patent, the '6917 patent, the '8103 patent, the '0407 patent, the '0865 patent, the '1268 patent, the '6625 patent, the '1207 patent, the '3246 patent, the '8952 patent, the '1630 patent, the '5967 patent, the '3424 patent, the '5268 patent, the '6626 patent, the '5796 patent, the '3283 patent, the '9148 patent, the '9766 patent, the

[4] Icahn to Commence Investigation into Value Destruction by Illumina's Board (https://carlicahn.com/icahn-to-commence-investigation-into-value-destruction-by-illuminas-board/)

'9795 patent and the '1267 patent. These issued patents generally relate to the concepts of detecting a mutation, generic aberrations, copy number variations, or a cancer, etc., using ZipCodes and the ZipCode Operating System via universal addressable DNA arrays to analyze a sample.

57. Plaintiff contributed novel concepts and work to the inventions claimed in the '5431 patent, the '0584 patent, the '0441 patent, the '2005 patent, the '8394 patent, the '0741 patent, the '3884 patent, the '3754 patent, the '0431 patent, the '6431 patent, the '6734 patent, the '1488 patent, the '5971 patent, the '0841 patent, the '3576 patent, the '2420 patent, the '1869 patent, the '2020 patent, the '0810 patent, the '6531 patent, the '3537 patent, the '9626 patent, the '1897 patent, the '4988 patent, the '5794 patent, the '0119 patent, the '3354 patent, the '6063 patent, the '0380 patent, the '0363 patent, the '0626 patent, the '0627 patent, the '6917 patent, the '8103 patent, the '0407 patent, the '0865 patent, the '1268 patent, the '6625 patent, the '1207 patent, the '3246 patent, the '8952 patent, the '1630 patent, the '5967 patent, the '3424 patent, the '5268 patent, the '6626 patent, the '5796 patent, the '3283 patent, the '9148 patent, the '9766 patent, the '9795 patent and the '1267 patent. These contributions include, for example, applications of ZipCode sequences sets to decode the physical location or addresses of various DNA sequences in random bead arrays such as Sentrix and Infinium arrays.

58. David R. Walt and Mark Chee and Jian Bing Fan and John R. Stuelpnagel, and others while employed at Illumina, received and used 20 of ZipCodes invented by the Plaintiff and confidential information via the Joint Development Agreement and Joint Steering Committee with Perkin Elmer Applied Biosystems and produced products including Sentrix arrays and GoldenGate assay and derivative arrays using multiple ZipCodes per bead such as Infinium arrays, including to conceive claimed inventions of the '5431 patent, the '0584 patent, the '0441 patent, the '2005 patent, the '8394 patent, the '0741 patent, the '3884 patent, the '3754 patent, the '0431 patent, the

'6431 patent, the '6734 patent, the '1488 patent, the '5971 patent, the '0841 patent, the '3576 patent, the '2420 patent, the '1869 patent, the '2020 patent, the '0810 patent, the '6531 patent, the '3537 patent, the '9626 patent, the '1897 patent, the '4988 patent, the '5794 patent, the '0119 patent, the '3354 patent, the '6063 patent, the '0380 patent, the '0363 patent, the '0626 patent, the '0627 patent, the '6917 patent, the '8103 patent, the '0407 patent, the '0865 patent, the '1268 patent, the '6625 patent, the '1207 patent, the '3246 patent, the '8952 patent, the '1630 patent, the '5967 patent, the '3424 patent, the '5268 patent, the '6626 patent, the '5796 patent, the '3283 patent, the '9148 patent, the '9766 patent, the '9795 patent and the '1267 patent.

59. There is a dispute as to the correct naming of inventors on the '5431 patent, the '0584 patent, the '0441 patent, the '2005 patent, the '8394 patent, the '0741 patent, the '3884 patent, the '3754 patent, the '0431 patent, the '6431 patent, the '6734 patent, the '1488 patent, the '5971 patent, the '0841 patent, the '3576 patent, the '2420 patent, the '1869 patent, the '2020 patent, the '0810 patent, the '6531 patent, the '3537 patent, the '9626 patent, the '1897 patent, the '4988 patent, the '5794 patent, the '0119 patent, the '3354 patent, the '6063 patent, the '0380 patent, the '0363 patent, the '0626 patent, the '0627 patent, the '6917 patent, the '8103 patent, the '0407 patent, the '0865 patent, the '1268 patent, the '6625 patent, the '1207 patent, the '3246 patent, the '8952 patent, the '1630 patent, the '5967 patent, the '3424 patent, the '5268 patent, the '6626 patent, the '5796 patent, the '3283 patent, the '9148 patent, the '9766 patent, the '9795 patent and the '1267 patent.

60. Plaintiff requests that the Court issue a declaratory judgment correcting the inventorship of the patents at issue in this claim and adding Plaintiff Zirvi as a co-inventor pursuant to 35 USC § 256.

61. Plaintiff had to hire the undersigned attorneys and agree to pay attorney's fees to the firm to represent him in this matter.

PRAYER FOR RELIEF COUNT I

62. The Plaintiff seeks relief under 35 U.S.C. § 256 and requests that the inventorship of the patents in question be corrected to include the Plaintiff as an inventor. In addition, Plaintiff seeks damages for reputational harm resulting from not being named as an inventor on the patents. The claimant requests damages to compensate for both the economic harm and the reputational harm caused by the assignee's failure to name the claimant as an inventor on the patents.

WHEREFORE, Plaintiff prays for judgment as follows:

(a) An order directing the Director of the United States Patent and Trademark Office to issue a certificate correcting inventorship of: the '5431 patent, the '0584 patent, the '0441 patent, the '2005 patent, the '8394 patent, the '0741 patent, the '3884 patent, the '3754 patent, the '0431 patent, the '6431 patent, the '6734 patent, the '1488 patent, the '5971 patent, the '0841 patent, the '3576 patent, the '2420 patent, the '1869 patent, the '2020 patent, the '0810 patent, the '6531 patent, the '3537 patent, the '9626 patent, the '1897 patent, the '4988 patent, the '5794 patent, the '0119 patent, the '3354 patent, the '6063 patent, the '0380 patent, the '0363 patent, the '0626 patent, the '0627 patent, the '6917 patent, the '8103 patent, the '0407 patent, the '0865 patent, the '1268 patent, the '6625 patent, the '1207 patent, the '3246 patent, the '8952 patent, the '1630 patent, the '5967 patent, the '3424 patent, the '5268 patent, the '6626 patent, the '5796 patent, the '3283 patent, the '9148 patent, the '9766 patent, the '9795 patent and the '1267 patent, by including Plaintiff Zirvi as co-inventor of the claims is the patents;

(b) An order directing the Defendants to correct inventorship of any foreign or international patent based on patents in paragraph (a) by including Plaintiff Zirvi as co-inventor of the claims of those patents and pending applications;

(c) A declaration that this case be deemed exceptional pursuant to 35 USC section 285, as appropriate and provided by law;

(d) Reasonable attorney's fees, costs, and expenses, as appropriate and provided by law; and

(e) Such other and further relief as the Court deems just and proper,

COUNT II – LEGAL MALPRACTICE

63. Plaintiff repeats and re-alleges all of the paragraphs 1-62 above as if fully set forth herein.

64. Defendant Attorneys were negligent in their representation of Plaintiff in the Cornell case and subsequent settlement thereof when they failed to represent Plaintiff during settlement negotiations or failed to inform Plaintiff during the Cornell case that they were not going to represent Plaintiff's interest.

65. The Defendant Attorneys' failure to represent, or inform Plaintiff to retain his own representation, was the proximate cause of damages, including, failing to be compensated or recognized under the Cornell case settlement agreement, and causing Plaintiff to lose rights of enforcement against Illumina in an individual matter because of a delay him bringing such action against Illumina, based upon representations that those claims would be resolved on behalf of the Plaintiff in the Cornell case.

66. Plaintiff's damages include lost compensation for the intellectual property that was incorporated into a series of Illumina patents and products since the time of Plaintiff's invention. In addition, damages include all expenses incurred trying to enforce intellectual property rights against Illumina that the Federal Court of Appeals ultimately ruled were time-barred, even though Plaintiff's counsel never advised of the potential that his claims could be barred. Indeed, Plaintiff's counsel advised that there was no need to pursue individual claims, or representation because Plaintiff was represented and all ThermoFisher, Cornell, and the inventors' interests were all aligned.

67. The damage suffered by Plaintiff are actual and ascertainable damages that would not have been incurred by the Plaintiff but for the malpractice of Defendant Attorneys.

68. In doing the acts alleged herein, Defendants acted with fraud, oppression and/or malice entitling Plaintiff to punitive damages according to proof at trial.

PRAYER FOR RELIEF COUNT II

WHEREFORE, Plaintiff's pray for judgment as follows:

69. A declaratory judgment that Defendants engaged in fraudulent conduct in violation of federal and state law;

70. An award of compensatory and consequential damages, including but not limited to the costs and expenses associated with investigating and prosecuting this matter;

71. An award of punitive damages to punish Defendants for their egregious, intentional and malicious conduct;

72. An award of attorney's fees and other costs incurred in connection with this lawsuit;

73. An order requiring Defendants to cease their fraudulent conduct and to take all necessary steps to remedy the harm caused to Plaintiff and other affected parties;

74. An order compelling Defendants to provide a full and accurate accounting of all funds and assets obtained through their fraudulent conduct;

75. An order enjoining Defendants from engaging in similar fraudulent conduct in the future;

76. An order requiring Defendants to disgorge any ill-gotten profits obtained through their fraudulent conduct;

77. Any further relief that this Honorable Court deems just and proper.

COUNT III – FRAUD

78. Plaintiff repeats and re-alleges all of the paragraphs 1-77 above as if fully set forth herein.

79. Defendants (ThermoFisher and the Attorneys) made material misrepresentations, and misrepresentations by omission, to Plaintiff. During the *Cornell Case* the Defendants stated that they would represent Plaintiff and in fact did represent Plaintiff in preparing and defending his deposition.

80. Defendants made false statements that Plaintiff's interest and ThermoFisher's interest were aligned and therefore he didn't need his own attorney.

81. Defendants made false statements that even if Plaintiff retained his own counsel, that counsel would not be allowed to participate in the litigation.

82. Defendants intended that Plaintiff rely on these statements so that he would not hire his own attorneys to protect his interest in the intellectual property plaintiff had created and Illumina incorporated intuits into its own patents and products.

83. Defendants intended that Plaintiff relied on these statements so that he would not hire his own attorney to be involved in the negotiation of a settlement agreement between ThermoFisher and alumina, that directly affected Plaintiff's rights in his intellectual property.

84. Plaintiff reasonably relied upon the assertions of ThermoFisher and its lawyers regarding the alignment of interest and the ThermoFisher lawyers' ability to represent Plaintiff's interests during the litigation and settlement of the Cornell case.

85. As a result of Plaintiff's reasonable reliance under the false statement of ThermoFisher and its lawyers Plaintiff lost his right to be a party too, and compensated under, the settlement agreement in the Cornell case which included his intellectual property.

86. The Plaintiff suffered further damage in a subsequent lawsuit against Illumina to enforce his rights when it was determined that the Defendants' fraudulent and purposeful delay in Plaintiff asserting claims against Illumina time-barred his recovery of rightful royalties.

87. In doing the acts alleged herein, Defendants acted with fraud, oppression and/or malice entitling Plaintiff to punitive damages according to proof at trial.

PRAYER FOR RELIEF COUNT III

WHEREFORE, Plaintiff's pray for judgment as follows:

88. Compensatory damages in an amount to be determined at trial, plus prejudgment interest;

89. Punitive damages in an amount to be determined at trial;

90. Costs, disbursements, and reasonable attorneys' fees incurred in this action under applicable law;

91. Other relief that this Honorable Court deems just and proper.

COUNT IV – CIVIL CONSPIRACY

Plaintiff alleges that Defendants: (1) a combination of two or more persons; (2) acting in concert to commit an unlawful act or commit a lawful act by unlawful means; (3) an agreement between the parties to inflict a wrong against or injury upon another; and (4) an overt act resulting in damages. See, Banco Popular N. Am. v. Gandi, 184 N.J. 161, 183 (2005). "Under New Jersey law, a claim for civil conspiracy cannot survive without a viable underlying tort…." Id. at 177-78.

92. Plaintiff re-alleges and reincorporates paragraphs 1 through 91 as though fully set forth herein.

93. Defendants in combination with one another, acting in concert, conspired to commit an unlawful act that deprived the Plaintiff of his intellectual property rights and did so while pretending to be engaged in a legitimate fight within the confines of a court case over the intellectual property.

94. Defendants Illumina and Thermo Fisher had an agreement between each other to inflict harm upon Plaintiff by constructing an agreement that would make it impossible for Plaintiff the enforce his right to his intellectual property.

95. Defendants ThermoFisher and their counsel Rip Finst, Sean Boyle, Roger Chin, and Douglas Lumish would feign to be averse to Illumina while at the same time working together to usurp Plaintiff's Intellectual Property Rights for their own use and benefit.

96. Defendants ThermoFisher and their counsel Rip Finst, Sean Boyle, Matthew A. Pearson, Angela Verrecchio, Roger Chin and Douglas Lumish failed to represent Plaintiff's interests in the settlement agreement of April 2017.

97. Defendants ThermoFisher and their counsel Rip Finst, Sean Boyle, Roger Chin, and Douglas Lumish would overtly use a series of court proceedings and settlements to damage the Plaintiff and secure his Intellectual Property for themselves.

98. Defendants' Thermo Fisher and their counsel Rip Finst, Sean Boyle, Matthew A. Pearson, Roger Chin, and Douglas Lumish acts would provide Illumina and Thermo Fisher with a duopoly that has made the companies billions of dollars in profit and market cap over the last decade and has deprived the Plaintiff of the same during that time.

99. In doing the acts alleged herein, Defendants acted with fraud, oppression and/or malice entitling Plaintiff to punitive damages according to proof at trial.

PRAYER FOR RELIEF COUNT IV

WHEREFORE, Plaintiff's pray for judgment as follows:

100. Compensatory damages in an amount to be determined at trial, plus prejudgment interest;

101. Punitive damages in an amount to be determined at trial;

102. Costs, disbursements, and reasonable attorneys' fees incurred in this action under applicable law;

103. Other relief that this Honorable Court deems just and proper.

JURY DEMAND

The Plaintiff demands trial by a jury on all the triable issues of this complaint.

Respectfully submitted,

LORIUM LAW
Counsel for Plaintiffs
101 NE 3rd Ave, Suite 1800
Fort Lauderdale, FL 33301
Telephone: (954) 462-8000
Facsimile: (954) 462-4300

By: /s/ Joseph D. Garrity
JOSEPH D. GARRITY, ESQ.
Florida Bar No. 87531
Primary Email
jgarrity@loriumlaw.com
Secondary Service Email
GLT@loriumlaw.com

Soliman & Associates, PC
Counsel for Plaintiffs
923 Haddonfield Rd., Ste. 300
Cherry Hill, NJ 08002
Telephone: (856) 324-8313
Facsimile: (856) 324-9080

By: /s/ Ahmed Soliman
AHMED SOLIMAN, ESQ.
NJ Bar No. 016572010
Primary Email
soliman@solimanlegal.com

EXHIBIT 1

UNITED STATES

SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

Form 8-K

Current Report

Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934

Date of Report (Date of earliest event reported): January 30, 2018

Illumina, Inc.

(Exact name of registrant as specified in its charter)

001-35406
(Commission File Number)

Delaware	**33-0804655**
(State or other jurisdiction of incorporation)	(I.R.S. Employer Identification No.)

5200 Illumina Way, San Diego, CA 92122
(Address of principal executive offices) (Zip code)

(858) 202-4500
(Registrant's telephone number, including area code)

N/A
(Former name or former address, if changed since last report)

Check the appropriate box below if the Form 8-K filing is intended to simultaneously satisfy the filing obligation of the registrant under any of the following provisions:

☐ Written communications pursuant to Rule 425 under the Securities Act (17 CFR 230.425)

☐ Soliciting material pursuant to Rule 14a-12 under the Exchange Act (17 CFR 240.14a-12)

☐ Pre-commencement communications pursuant to Rule 14d-2(b) under the Exchange Act (17 CFR 240.14d-2(b))

☐ Pre-commencement communications pursuant to Rule 13e-4(c) under the Exchange Act (17 CFR 240.13e-4(c))

Indicate by check mark whether the registrant is an emerging growth company as defined in Rule 405 of the Securities Act of 1933 (§230.405 of this chapter) or Rule 12b-2 of the Securities Exchange Act of 1934 (§240.12b-2 of this chapter).

Emerging growth company ☐

If an emerging growth company, indicate by check mark if the registrant has elected not to use the extended transition period for complying with any new or revised financial accounting standards provided pursuant to Section 13a of the Exchange Act. ☐

TABLE OF CONTENTS

Table of Contents

Item 2.02 Results of Operations and Financial Condition.

On January 30, 2018, Illumina, Inc. (the "Company") issued a press release announcing financial results for the fourth quarter and fiscal year ended December 31, 2017. The full text of the Company's press release is attached hereto as Exhibit 99.1.

The information furnished pursuant to this Item 2.02 shall not be deemed to be "filed" for purposes of Section 18 of the Securities Exchange Act of 1934 (the "Exchange Act") or otherwise subject to the liabilities under that Section and shall not be deemed to be incorporated by reference into any filing of the Company under the Securities Act of 1933 or the Exchange Act.

Table of Contents

Item 9.01 Financial Statements and Exhibits.

(d) Exhibits.

99.1 Press release dated January 30, 2018, announcing Illumina, Inc.'s financial results for the fourth quarter and fiscal year ended December 31, 2017.

Table of Contents

SIGNATURES

Pursuant to the requirements of the Securities Exchange Act of 1934, the Registrant has duly caused this report to be signed on its behalf by the undersigned hereunto duly authorized.

ILLUMINA, INC.

Date: January 30, 2018

By: /s/ SAM A. SAMAD

Sam A. Samad

Senior Vice President and Chief Financial Officer

Table of Contents

Exhibit Index

Exhibit Number	Description
99.1	Press release dated January 30, 2018, announcing Illumina, Inc.'s financial results for the fourth quarter and fiscal year ended December 31, 2017.

Investors:
Jacquie Ross, CFA
858-882-2172
ir@illumina.com

Media:
Eric Endicott
858-882-6822
pr@illumina.com

Illumina Reports Financial Results for Fourth Quarter and Fiscal Year 2017

San Diego -- (BUSINESS WIRE) - January 30, 2018 - Illumina, Inc. (NASDAQ:ILMN) today announced its financial results for the fourth quarter and fiscal year 2017.

Fourth quarter 2017 results:

- Revenue of $778 million, a 26% increase compared to $619 million in the fourth quarter of 2016
- GAAP net income attributable to Illumina stockholders for the quarter of $68 million, or $0.46 per diluted share, compared to $124 million, or $0.84 per diluted share, for the fourth quarter of 2016; GAAP provision for income taxes for the fourth quarter of 2017 includes our provisional estimate of the one-time transition tax as a result of U.S. tax reform
- Non-GAAP net income attributable to Illumina stockholders for the quarter of $212 million, or $1.44 per diluted share, compared to $126 million, or $0.85 per diluted share, for the fourth quarter of 2016 (see the table entitled “Itemized Reconciliation Between GAAP and Non-GAAP Net Income Attributable to Illumina Stockholders” for a reconciliation of GAAP and non-GAAP financial measures)
- Cash flow from operations of $294 million compared to $262 million in the fourth quarter of 2016
- Free cash flow (cash flow from operations less capital expenditures) of $218 million for the quarter, compared to $180 million in the fourth quarter of 2016

Gross margin in the fourth quarter of 2017 was 69.7% compared to 67.7% in the prior year period. Excluding amortization of acquired intangible assets, non-GAAP gross margin was 70.9% for the fourth quarter of 2017 compared to 69.5% in the prior year period.

Research and development (R&D) expenses for the fourth quarter of 2017 were $137 million compared to $130 million in the prior year period. Excluding restructuring charges, non-GAAP R&D expenses as a percentage of revenue were 17.4%, including 0.7% attributable to Helix. This compares to 21.0% in the prior year period, including 2.6% attributable to GRAIL and Helix.

Selling, general and administrative (SG&A) expenses for the fourth quarter of 2017 were $175 million compared to $146 million in the prior year period. Excluding amortization of acquired intangible assets and restructuring

charges, SG&A expenses as a percentage of revenue were 22.1%, including 1.2% attributable to Helix. This compares to 23.4% in the prior year period, including 1.7% attributable to GRAIL and Helix.

Depreciation and amortization expenses were $40 million and capital expenditures for free cash flow purposes were $76 million during the fourth quarter of 2017. At the close of the quarter, the company held $2.1 billion in cash, cash equivalents and short-term investments, compared to $1.6 billion as of January 1, 2017.

Fiscal 2017 results:

- Revenue of $2,752 million, a 15% increase compared to $2,398 million in fiscal 2016
- GAAP net income attributable to Illumina stockholders of $726 million, or $4.92 per diluted share, compared to $463 million, or $3.07 per diluted share, in fiscal 2016; GAAP provision for income taxes for fiscal 2017 includes the provisional estimate of the one-time transition tax referenced previously
- Non-GAAP net income attributable to Illumina stockholders of $591 million, or $4.00 per diluted share, compared to $503 million, or $3.33 per diluted share, in fiscal 2016 (see the table entitled “Itemized Reconciliation Between GAAP and Non-GAAP Net Income Attributable to Illumina Stockholders” for a reconciliation of these GAAP and non-GAAP financial measures)
- Cash flow from operations of $875 million compared to $779 million in fiscal 2016
- Free cash flow (cash flow from operations less capital expenditures) of $565 million, compared to $519 million in fiscal 2016

Gross margin for fiscal 2017 was 66.4% compared to 69.5% in the prior year. Excluding amortization and impairment of acquired intangible assets, non-GAAP gross margin was 68.4% for fiscal 2017 compared to 71.3% in the prior year period.

Research and development (R&D) expenses for fiscal 2017 were $546 million compared to $504 million in the prior year. Excluding restructuring charges and an impairment of in-process research and development, non-GAAP R&D expenses as a percentage of revenue were 19.6%, including 1.0% attributable to GRAIL and Helix. This compares to 21.0% in the prior year period, including 1.9% attributable to GRAIL and Helix.

Selling, general and administrative (SG&A) expenses for fiscal 2017 were $674 million compared to $584 million in the prior year period. Excluding amortization of acquired intangible assets, restructuring charges, performance-based compensation related to GRAIL Series B financing, and acquisition related gain, SG&A expenses as a percentage of revenue were 23.9%, including 1.7% attributable to GRAIL and Helix. This compares to 24.0% in the prior year period, including 1.2% attributable to GRAIL and Helix.

“With 26% revenue growth in the fourth quarter, and 15% for the full year, our 2017 results demonstrate customers’ growing demand across both our sequencing and array portfolios,” said Francis deSouza, President and

CEO. “From our NovaSeq and the recently launched iSeq, to our clinical portfolio that includes the VeriSeq NIPT Solution, NextSeqDx and MiSeqDx, Illumina is well-positioned to continue to drive sequencing innovation and unlock the power of the genome.”

Updates since our last earnings release:

- Launched the iSeq™ 100 Sequencing System, a flexible benchtop sequencer priced at $19,900 designed to provide a fast and easy-to-use system with unmatched accuracy
- Announced availability of AmpliSeq for Illumina, developed in partnership with Thermo Fisher Scientific
- Introduced the NextSeq™ 550Dx instrument, Illumina’s second FDA regulated CE-IVD market platform, to deliver the power of high-throughput next-generation sequencing (NGS) to the clinical laboratory
- Partnered with KingMed Diagnostics to develop novel oncology and hereditary disease testing applications utilizing Illumina’s NGS technology in China
- Appointed Gary S. Guthart to the company’s Board of Directors
- Appointed Aimee Hoyt to the position of Senior Vice President and Chief People Officer
- Repurchased $75 million of common stock in the fourth quarter under the previously announced share repurchase program

Financial outlook and guidance

The non-GAAP financial guidance discussed below reflects certain pro forma adjustments to assist in analyzing and assessing our core operational performance. Please see our Reconciliation of Non-GAAP Financial Guidance included in this release for a reconciliation of the GAAP and non-GAAP financial measures.

For fiscal 2018, the company is projecting 13% to 14% revenue growth, GAAP earnings per diluted share attributable to Illumina stockholders of $4.14 to $4.24 and non-GAAP earnings per diluted share attributable to Illumina stockholders of $4.50 to $4.60.

Quarterly conference call information

The conference call will begin at 2:00 pm Pacific Time (5:00 pm Eastern Time) on Tuesday, January 30, 2018. Interested parties may access the live teleconference through the Investor Relations section of Illumina’s web site under the “company” tab at www.illumina.com. Alternatively, individuals can access the call by dialing 888-771-4371, or 1-847-585-4405 outside North America, both with passcode 46251622.

A replay of the conference call will be available from 4:30 pm Pacific Time (7:30 pm Eastern Time) on January 30, 2018 through February 6, 2018 by dialing 888-843-7419, or 1-630-652-3042 outside North America, both with passcode 46251622.

Statement regarding use of non-GAAP financial measures

The company reports non-GAAP results for diluted net income per share, net income, gross margins, operating expenses, operating margins, other income, and free cash flow in addition to, and not as a substitute for, or superior to, financial measures calculated in accordance with GAAP. The company's financial measures under GAAP include substantial charges such as amortization of acquired intangible assets, non-cash interest expense associated with the company's convertible debt instruments that may be settled in cash, and others that are listed in the itemized reconciliations between GAAP and non-GAAP financial measures included in this press release. Management has excluded the effects of these items in non-GAAP measures to assist investors in analyzing and assessing past and future operating performance. Additionally, non-GAAP net income attributable to Illumina stockholders and diluted earnings per share attributable to Illumina stockholders are key components of the financial metrics utilized by the company's board of directors to measure, in part, management's performance and determine significant elements of management's compensation.

The company encourages investors to carefully consider its results under GAAP, as well as its supplemental non-GAAP information and the reconciliation between these presentations, to more fully understand its business. Reconciliations between GAAP and non-GAAP results are presented in the tables of this release.

Use of forward-looking statements

This release contains forward-looking statements that involve risks and uncertainties, such as our financial outlook and guidance for fiscal 2018 and expectations regarding the launch of new products. Among the important factors that could cause actual results to differ materially from those in any forward-looking statements are: (i) challenges inherent in developing, manufacturing, and launching new products and services, including expanding manufacturing operations and reliance on third-party suppliers for critical components; (ii) the timing and mix of customer orders among our products and services; (iii) the impact of recently launched or pre-announced products and services on existing products and services; (iv) our ability to further develop and commercialize our instruments and consumables and to deploy new products, services, and applications, and expand the markets, for our technology platforms; (v) our ability to manufacture robust instrumentation and consumables; (vi) the success of products and services competitive with our own; (vii) our ability to successfully identify and integrate acquired technologies, products, or businesses; (viii) our expectations and beliefs regarding future conduct and growth of the business and the markets in which we operate; and (ix) the application of generally accepted accounting principles, which are highly complex and involve many subjective assumptions, estimates, and judgments, together with other factors detailed in our filings with the Securities and Exchange Commission, including our most recent filings on Forms 10-K and 10-Q, or in information disclosed in public conference calls, the date and time of which are released beforehand. We undertake no obligation, and do not intend, to update these forward-looking statements, to review or confirm analysts' expectations, or to provide interim reports or updates on the progress of the current quarter.

About Illumina

Illumina is improving human health by unlocking the power of the genome. Our focus on innovation has established us as the global leader in DNA sequencing and array-based technologies, serving customers in the research, clinical and applied markets. Our products are used for applications in the life sciences, oncology, reproductive health, agriculture and other emerging segments. To learn more, visit www.illumina.com and follow @illumina.

###

Illumina, Inc.
Condensed Consolidated Balance Sheets
(In millions)

	December 31, 2017	January 1, 2017
ASSETS	(unaudited)	
Current assets:		
Cash and cash equivalents	$ 1,225	$ 735
Short-term investments	920	824
Accounts receivable, net	411	381
Inventory	333	300
Prepaid expenses and other current assets	91	78
Total current assets	2,980	2,318
Property and equipment, net	931	713
Goodwill	771	776
Intangible assets, net	175	243
Deferred tax assets	88	123
Other assets	312	108
Total assets	$ 5,257	$ 4,281
LIABILITIES AND STOCKHOLDERS' EQUITY		
Current liabilities:		
Accounts payable	$ 160	$ 138
Accrued liabilities	432	342
Build-to-suit lease liability	144	223
Long-term debt, current portion	10	2
Total current liabilities	746	705
Long-term debt	1,182	1,056
Other long-term liabilities	360	206
Redeemable noncontrolling interests	125	44
Stockholders' equity	2,844	2,270
Total liabilities and stockholders' equity	$ 5,257	$ 4,281

Illumina, Inc.
Condensed Consolidated Statements of Income
(In millions, except per share amounts)
(unaudited)

	Three Months Ended		Years Ended	
	December 31, 2017	January 1, 2017	December 31, 2017	January 1, 2017
Revenue:				
Product revenue	$ 659	$ 525	$ 2,289	$ 2,032
Service and other revenue	119	94	463	366
Total revenue	778	619	2,752	2,398
Cost of revenue:				
Cost of product revenue **(a)**	172	151	679	534
Cost of service and other revenue **(a)**	55	38	208	155
Amortization of acquired intangible assets	9	11	39	43
Total cost of revenue	236	200	926	732
Gross profit	542	419	1,826	1,666
Operating expense:				
Research and development **(a)**	137	130	546	504
Selling, general and administrative **(a) (b)**	175	146	674	584
Legal contingencies	—	—	—	(9)
Total operating expense	312	276	1,220	1,079
Income from operations	230	143	606	587
Other (expense) income, net	(6)	(9)	437	(26)
Income before income taxes	224	134	1,043	561
Provision for income taxes	166	26	365	133
Consolidated net income	58	108	678	428
Add: Net loss attributable to noncontrolling interests	10	16	48	35
Net income attributable to Illumina stockholders	$ 68	$ 124	$ 726	$ 463
Net income attributable to Illumina stockholders for earnings per share **(c)**	$ 68	$ 124	$ 725	$ 454
Earnings per share attributable to Illumina stockholders:				
Basic	$ 0.47	$ 0.84	$ 4.96	$ 3.09
Diluted	$ 0.46	$ 0.84	$ 4.92	$ 3.07
Shares used in computing earnings per common share:				
Basic	146	147	146	147
Diluted	148	148	148	148

(a) Includes stock-based compensation expense for stock-based awards:

	Three Months Ended		Years Ended	
	December 31, 2017	January 1, 2017	December 31, 2017	January 1, 2017
Cost of product revenue	$ 3	$ 3	$ 12	$ 9
Cost of service and other revenue	—	—	2	2
Research and development	13	9	51	42
Selling, general and administrative	25	15	99	76
Stock-based compensation expense before taxes **(1)**	$ 41	$ 27	$ 164	$ 129

(1) Includes stock-based compensation of $0.9 million and $3.5 million for Helix for the three months and year ended December 31, 2017, respectively, and $10.1 million for GRAIL for the year ended December 31, 2017. This compares to stock-based compensation of $0.5 million and $2.4 million for GRAIL and Helix for the three months and year ended January 1, 2017, respectively.

(b) Headquarter relocation expense of $0.4 million and $1.5 million was reclassified to selling, general and administrative expense for the three months and year ended January 1, 2017, respectively, to conform to the current period presentation.

(c) Amount reflects the additional losses attributable to the common shareholders of GRAIL and Helix for earnings per share purposes.

Illumina, Inc.

Condensed Consolidated Statements of Cash Flows

(In millions)

(unaudited)

	Three Months Ended		Years Ended	
	December 31, 2017	January 1, 2017	December 31, 2017	January 1, 2017
Net cash provided by operating activities **(a)**	$ 294	$ 262	$ 875	$ 779
Net cash used in investing activities	(315)	(173)	(214)	(515)
Net cash used in financing activities **(a)**	(109)	(145)	(176)	(296)
Effect of exchange rate changes on cash and cash equivalents	1	(4)	5	(2)
Net (decrease) increase in cash and cash equivalents	(129)	(60)	490	(34)
Cash and cash equivalents, beginning of period	1,354	795	735	769
Cash and cash equivalents, end of period	$ 1,225	$ 735	$ 1,225	$ 735
Calculation of free cash flow:				
Net cash provided by operating activities **(a)**	$ 294	$ 262	$ 875	$ 779
Purchases of property and equipment **(b)**	(76)	(82)	(310)	(260)
Free cash flow **(c)**	$ 218	$ 180	$ 565	$ 519

(a) Excess tax expense of $19 million and tax benefit of $91 million related to stock-based compensation for the three months and year ended January 1, 2017, respectively, was reclassified from cash used in financing activities to cash provided by operating activities as a result of the retrospective application of ASU 2016-09 adopted in Q1 2017.

(b) Excludes property and equipment recorded under build-to-suit lease accounting, which are non-cash expenditures, of $19 million and $79 million for the three months and year ended December 31, 2017, respectively, and $25 million and $193 million for the three months and year ended January 1, 2017, respectively.

(c) Free cash flow, which is a non-GAAP financial measure, is calculated as net cash provided by operating activities reduced by purchases of property and equipment. Free cash flow is useful to management as it is one of the metrics used to evaluate our performance and to compare us with other companies in our industry. However, our calculation of free cash flow may not be comparable to similar measures used by other companies.

Illumina, Inc.
Results of Operations - Non-GAAP
(In millions, except per share amounts)
(unaudited)

ITEMIZED RECONCILIATION BETWEEN GAAP AND NON-GAAP EARNINGS PER SHARE ATTRIBUTABLE TO ILLUMINA STOCKHOLDERS:

	Three Months Ended		Years Ended	
	December 31, 2017	January 1, 2017	December 31, 2017	January 1, 2017
GAAP earnings per share attributable to Illumina stockholders - diluted	**$ 0.46**	**$ 0.84**	**$ 4.92**	**$ 3.07**
Amortization of acquired intangible assets	0.07	0.08	0.30	0.33
Non-cash interest expense **(a)**	0.05	0.05	0.20	0.20
Restructuring **(b)**	0.03	—	0.03	—
Gain on deconsolidation of GRAIL **(c)**	—	—	(3.07)	—
Impairments **(d)**	—	—	0.15	—
Performance-based compensation related to GRAIL Series B financing **(e)**	—	—	0.03	—
Equity-method investment gain **(f)**	—	—	(0.01)	—
Acquisition related gain **(g)**	—	—	(0.01)	—
Legal contingencies **(h)**	—	—	—	(0.06)
Contingent compensation expense **(i)**	—	—	—	0.01
Headquarter relocation	—	—	—	0.01
Deemed dividend **(j)**	—	—	—	(0.01)
Incremental non-GAAP tax expense **(k)**	(0.05)	(0.07)	0.80	(0.17)
Tax benefit related to cost-sharing arrangement **(l)**	—	(0.05)	—	(0.05)
U.S. Tax Reform **(m)**	1.01	—	1.01	—
Excess tax benefit from share-based compensation **(n)**	(0.13)	—	(0.35)	—
Non-GAAP earnings per share attributable to Illumina stockholders - diluted **(o)**	$ 1.44	$ 0.85	$ 4.00	$ 3.33

ITEMIZED RECONCILIATION BETWEEN GAAP AND NON-GAAP NET INCOME ATTRIBUTABLE TO ILLUMINA STOCKHOLDERS:

GAAP net income attributable to Illumina stockholders	**$ 68**	**$ 124**	**$ 726**	**$ 463**
Amortization of acquired intangible assets	10	12	45	49
Non-cash interest expense **(a)**	8	8	30	30
Restructuring **(b)**	4	—	4	—
Gain on deconsolidation of GRAIL **(c)**	—	—	(453)	—
Impairments **(d)**	—	—	23	—
Performance-based compensation related to GRAIL Series B financing **(e)**	—	—	4	—
Equity-method investment gain **(f)**	—	—	(2)	—
Acquisition related gain **(g)**	—	—	(1)	—
Legal contingencies **(h)**	—	—	—	(9)
Contingent compensation expense **(i)**	—	—	—	2
Headquarter relocation	—	—	—	1
Incremental non-GAAP tax expense **(k)**	(7)	(11)	117	(26)
Tax benefit related to cost-sharing arrangement **(l)**	—	(7)	—	(7)
U.S. Tax Reform **(m)**	150	—	150	—

Excess tax benefit from share-based compensation **(n)**	(21)	—	(52)	—
Non-GAAP net income attributable to Illumina stockholders **(o)**	$ 212	$ 126	$ 591	$ 503

All amounts in tables are rounded to the nearest millions, except as otherwise noted. As a result, certain amounts may not recalculate using the rounded amounts provided.

(a) Non-cash interest expense is calculated in accordance with the authoritative accounting guidance for convertible debt instruments that may be settled in cash.
(b) Amount consists primarily of employee costs related to the restructuring that occurred in Q4 2017.
(c) Amount represents the gain recognized as a result of the deconsolidation of GRAIL in Q1 2017. The $150 million tax effect of the gain is included in incremental non-GAAP tax expense. Subsequent to the transaction, our remaining interest is treated as a cost-method investment.
(d) Amount represents impairment of an acquired intangible asset and in-process research and development of $18 million and $5 million, respectively.
(e) Amount represents performance-based stock which vested as a result of the financing, net of attribution to noncontrolling interest.
(f) Equity-method investment gain represents mark-to-market adjustments from our investment in Illumina Innovations Fund I, L.P.
(g) Acquisition related gain consists of change in fair value of contingent consideration.
(h) Legal contingencies for 2016 represent a reversal of previously recorded expense related to the settlement of patent litigation.
(i) Contingent compensation expense relates to contingent payments for post-combination services associated with an acquisition.
(j) Amount represents the impact of a deemed dividend, net of Illumina's portion of the losses incurred by GRAIL's common stockholders resulting from the company's common to preferred share exchange with GRAIL. The amount was added to net income attributable to Illumina stockholders for purposes of calculating Illumina's consolidated earnings per share. The deemed dividend, net of tax, was recorded through equity.
(k) Incremental non-GAAP tax expense reflects the tax impact related to the non-GAAP adjustments listed above.
(l) Tax benefit related to cost-sharing arrangement refers to the exclusion of stock compensation from prior period cost-sharing charges as a result of a tax court ruling.
(m) In accordance with the Tax Cuts and Jobs Act enacted on December 22, 2017 (U.S. Tax Reform), amount primarily consists of a provisional estimate of the one-time transition tax on earnings of certain foreign subsidiaries that were previously tax deferred.
(n) Excess tax benefits from share-based compensation are recorded as a discrete item within the provision for income taxes on the consolidated statement of income pursuant to ASU 2016-09, which was previously recognized in additional paid-in capital on the consolidated statement of stockholders' equity.
(o) Non-GAAP net income attributable to Illumina stockholders and diluted earnings per share attributable to Illumina stockholders exclude the effect of the pro forma adjustments as detailed above. Non-GAAP net income attributable to Illumina stockholders and diluted earnings per share attributable to Illumina stockholders are key components of the financial metrics utilized by the company's board of directors to measure, in part, management's performance and determine significant elements of management's compensation. Management has excluded the effects of these items in these measures to assist investors in analyzing and assessing our past and future core operating performance.

Illumina, Inc.

Results of Operations - Non-GAAP (continued)

(Dollars in millions)

(unaudited)

ITEMIZED RECONCILIATION BETWEEN GAAP AND NON-GAAP RESULTS OF OPERATIONS AS A PERCENT OF REVENUE:

	Three Months Ended				Years Ended			
	December 31, 2017		January 1, 2017		December 31, 2017		January 1, 2017	
GAAP gross profit	**$ 542**	**69.7 %**	**$ 419**	**67.7 %**	**$ 1,826**	**66.4 %**	**$ 1,666**	**69.5 %**
Amortization of acquired intangible asset	9	1.2 %	11	1.8 %	39	1.4 %	43	1.8 %
Impairment of acquired intangible asset	—	—	—	—	18	0.6 %	—	—
Non-GAAP gross profit **(a)**	$ 551	70.9 %	$ 430	69.5 %	$ 1,883	68.4 %	$ 1,709	71.3 %
GAAP research and development expense	**$ 137**	**17.7 %**	**$ 130**	**21.0 %**	**$ 546**	**19.8 %**	**$ 504**	**21.0 %**
Restructuring **(b)**	(2)	(0.3)%	—	—	(2)	(0.1)%	—	—
Impairment of in-process research and development	—	—	—	—	(5)	(0.1)%	—	—
Non-GAAP research and development expense	$ 135	17.4 %	$ 130	21.0 %	$ 539	19.6 %	$ 504	21.0 %
GAAP selling, general and administrative expense	**$ 175**	**22.5 %**	**$ 146**	**23.6 %**	**$ 674**	**24.6 %**	**$ 584**	**24.4 %**
Amortization of acquired intangible assets	(1)	(0.1)%	(1)	(0.2)%	(6)	(0.2)%	(6)	(0.2)%
Restructuring **(b)**	(2)	(0.3)%	—	—	(2)	(0.1)%	—	—
Performance-based compensation related to GRAIL Series B financing **(c)**	—	—	—	—	(10)	(0.4)%	—	—
Acquisition related gain **(d)**	—	—	—	—	1	—	—	—
Contingent compensation expense **(e)**	—	—	—	—	—	—	(2)	(0.1)%
Headquarter relocation	—	—	—	—	—	—	(1)	(0.1)%
Non-GAAP selling, general and administrative expense	$ 172	22.1 %	$ 145	23.4 %	$ 657	23.9 %	$ 575	24.0 %
GAAP operating profit	**$ 230**	**29.6 %**	**$ 143**	**23.1 %**	**$ 606**	**22.0 %**	**$ 587**	**24.5 %**
Amortization of acquired intangible assets	10	1.3 %	12	2.0 %	45	1.6 %	49	2.0 %
Restructuring **(b)**	4	0.5 %	—	—	4	0.1 %	—	— %
Legal contingencies **(f)**	—	—	—	—	—	—	(9)	(0.4)%
Impairments	—	—	—	—	23	0.9 %	—	—
Performance-based compensation related to GRAIL Series B financing **(c)**	—	—	—	—	10	0.4 %	—	—
Acquisition related gain **(d)**	—	—	—	—	(1)	—	—	—
Contingent compensation expense **(e)**	—	—	—	—	—	—	2	0.1 %
Headquarter relocation	—	—	—	—	—	—	1	0.1 %
Non-GAAP operating profit **(a)**	$ 244	31.4 %	$ 155	25.1 %	$ 687	25.0 %	$ 630	26.3 %
GAAP other (expense) income, net	**$ (6)**	**(0.8)%**	**$ (9)**	**(1.4)%**	**$ 437**	**15.9 %**	**$ (26)**	**(1.1)%**
Non-cash interest expense **(g)**	8	1.1 %	8	1.2 %	30	1.2 %	30	1.3 %
Equity-method investment gain **(h)**	—	—	—	—	(2)	(0.1)%	—	—
Gain on deconsolidation of GRAIL **(i)**	—	—	—	—	(453)	(16.5)%	—	—
Non-GAAP other income (expense), net **(a)**	$ 2	0.3 %	$ (1)	(0.2)%	$ 12	0.5 %	$ 4	0.2 %

All amounts in tables are rounded to the nearest millions, except as otherwise noted. As a result, certain amounts may not recalculate using the rounded amounts provided.

(a) Non-GAAP gross profit, included within non-GAAP operating profit, is a key measure of the effectiveness and efficiency of manufacturing processes, product mix and the average selling prices of our products and services. Non-GAAP operating profit, and non-GAAP other income (expense), net, exclude the effects of the pro forma adjustments as detailed above. Management has excluded the effects of these items in these measures to assist investors in analyzing and assessing past and future operating performance.
(b) Amount consists primarily of employee costs related to the restructuring that occurred in Q4 2017.
(c) Amount represents performance-based stock which vested as a result of the financing.
(d) Acquisition related gain consists of change in fair value of contingent consideration.
(e) Contingent compensation expense relates to contingent payments for post-combination services associated with an acquisition.
(f) Legal contingencies for 2016 represent a reversal of previously recorded expense related to the settlement of patent litigation.
(g) Non-cash interest expense is calculated in accordance with the authoritative accounting guidance for convertible debt instruments that may be settled in cash
(h) Equity-method investment gain represents mark-to-market adjustments from our investment in Illumina Innovations Fund I, L.P
(i) Amount represents the gain recognized as a result of the deconsolidation of GRAIL in Q1 2017. Subsequent to the transaction, our remaining interest is treated as a cost-method investment.

Illumina, Inc.
Reconciliation of Non-GAAP Financial Guidance

Our future performance and financial results are subject to risks and uncertainties, and actual results could differ materially from the guidance set forth below. Some of the factors that could affect our financial results are stated above in this press release. More information on potential factors that could affect our financial results is included from time to time in the public reports filed with the Securities and Exchange Commission, including Form 10-K for the fiscal year ended January 1, 2017 filed with the SEC on February 13, 2017, and Form 10-Q for the fiscal quarters ended April 2, 2017, July 2, 2017, and October 1, 2017. We assume no obligation to update any forward-looking statements or information.

	Fiscal Year 2018
GAAP diluted earnings per share attributable to Illumina stockholders	**$4.14 - $4.24**
Amortization of acquired intangible assets	0.24
Non-cash interest expense **(a)**	0.21
Restructuring **(b)**	0.02
Incremental non-GAAP tax expense **(c)**	(0.11)
Non-GAAP diluted earnings per share attributable to Illumina stockholders	$4.50 - $4.60

(a) Non-cash interest expense is calculated in accordance with the authoritative accounting guidance for convertible debt instruments that may be settled in cash.
(b) Amount consists primarily of employee severance and retention costs related to the restructuring that occurred in Q4 2017.
(c) Incremental non-GAAP tax expense reflects the tax impact related to the non-GAAP adjustments listed above.

EXHIBIT 2

Bringing together two leaders: AmpliSeqTM for Illumina

https://www.youtube.com/watch?v=lWPaZX1TDa4

January 8, 2018 - Illumina unexpectedly announces collaboration with Thermo Fisher with a joint product: “Ampliseq for Illumina”.

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 8-K

CURRENT REPORT
PURSUANT TO SECTION 13 OR 15 (d) OF THE SECURITIES EXCHANGE ACT OF 1934

Date of report (Date of earliest event reported): January 8, 2018

Illumina, Inc.

(Exact name of registrant as specified in its charter)

But internal innovation is only part of the story. Our partnership strategy ensures that our customers will have access to best-in-class technology. With that in mind, I am very pleased to announce an exciting partnership with Thermo Fisher. This collaboration deeply integrates the leading and most trusted sequencers with the leading amplicon library prep technology. Thermo Fisher has 20 years of experience in PCR amplification chemistries that has enabled them to develop an amplicon library prep that is simple, fast and robust. The technology has been widely adopted in the oncology space because of the high-quality data achievable with low input and degraded samples, such as FFPE tissue.

But until now, those customers have not been able to take advantage of the accuracy and power of Illumina sequencers. This agreement provides direct access to AmpliSeq for our RUO customers, delivering the best of sequencing with the best amplicon protocol. Ahead of today's launch, we have been working together to adapt and optimize AmpliSeq specifically for Illumina sequencers, creating a seamless user experience that is fast and easy.

In addition to a collection of ready-to-order panels, Illumina customers will also be able to access AmpliSeq algorithms to create custom amplicon assays utilizing DesignStudio, our online assay design protocol, and they will have access to future panels and new assay types that are introduced as part of the AmpliSeq product line.

Amplicon sequencing is an important entry point for new NGS customers. The combination of our proven technologies further removes barriers to adoption, bringing more new-to-sequencing customers to our platform. Of course, we also have an installed base of more than 9,000 desktop systems. By giving our existing customers access to AmpliSeq chemistry, we're enabling them to do even more with their systems.

January 8, 2018 - Illumina CEO Francis deSouza admits secret collaboration with Thermo Fisher "started the conversation clearly well over a year ago"… meaning prior to the settlement agreement.

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 8-K

CURRENT REPORT
PURSUANT TO SECTION 13 OR 15 (d) OF THE SECURITIES EXCHANGE ACT OF 1934

Date of report (Date of earliest event reported): January 8, 2018

Illumina, Inc.
(Exact name of registrant as specified in its charter)

Tycho W. Peterson - JP Morgan Chase & Co, Research Division - Senior Analyst

Can you talk about just that Thermo agreement, how that came about? (inaudible) press releases in Illumina (inaudible), so just a bit of speculation on that part? I mean, what's the back story?

Francis A. deSouza - Illumina, Inc. - CEO, President & Director

So the question is about the partnership we have with Thermo and what's the back story or how did that come about. The way it came about is as we talk to our customers around what they are looking for, one of the things that we heard consistently is that they liked certain aspects of AmpliSeq. And so they would tell us, look, they love the low sample they put in AmpliSeq, for example. They like the ability to work with degraded samples like FFPE tissue, and AmpliSeq has a following in the oncology community in terms of people like that assay. And so we spent a lot of time internally thinking, look, we want to make sure our customers have the best experience in our sequencer. So what is great for our customers is ultimately good for Illumina.

And so we started the conversation clearly well over a year ago. And initially, it was one of those, like, "Are you sure -- are we sure we want to do this? Are they sure they want to do this?" But it was always -- so the true north for us was what's the best thing for our customers? And if you keep looking at that, the truth is the best thing for our customers is to make the best amplicon technology available on the best sequencer. And I think a lot of credit on both sides. Thermo had the same thinking, which is, yes, there are parts of our portfolio we compete, but this is clearly good for AmpliSeq, it's good for customers and it's good for us. And so that was the thinking behind it. I think we're out of time.

January 30, 2018 - Illumina highlights collaboration with Thermo Fisher joint product: “Ampliseq for Illumina” in SEC report.

SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

Form 8-K

Current Report

Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934

Date of Report (Date of earliest event reported): January 30, 2018

Illumina, Inc.
(Exact name of registrant as specified in its charter)

Updates since our last earnings release:

- Launched the iSeq™ 100 Sequencing System, a flexible benchtop sequencer priced at $19,900 designed to provide a fast and easy-to-use system with unmatched accuracy
- Announced availability of AmpliSeq for Illumina, developed in partnership with Thermo Fisher Scientific
- Introduced the NextSeq™ 550Dx instrument, Illumina’s second FDA regulated CE-IVD market platform, to deliver the power of high-throughput next-generation sequencing (NGS) to the clinical laboratory

Bringing together two leaders: AmpliSeqTM for Illumina®

1,572 views 8 0 SHARE

Illumina Inc
Published on Jan 9, 2018

SUBSCRIBE 7.5K

1. I'm Joydeep Goswami, President of Clinical Oncology and NGS at Thermo Scientific

Bringing together two leaders: AmpliSeqTM for Illumina®

1,572 views 8 0 SHARE

Illumina Inc
Published on Jan 9, 2018

SUBSCRIBE 7.5K

2. My name is Mark Van Oene. I'm Illumina's Chief Commercial officer

Bringing together two leaders: AmpliSeqTM for Illumina®

1,572 views 8 0 SHARE

Illumina Inc
Published on Jan 9, 2018

SUBSCRIBE 7.5K

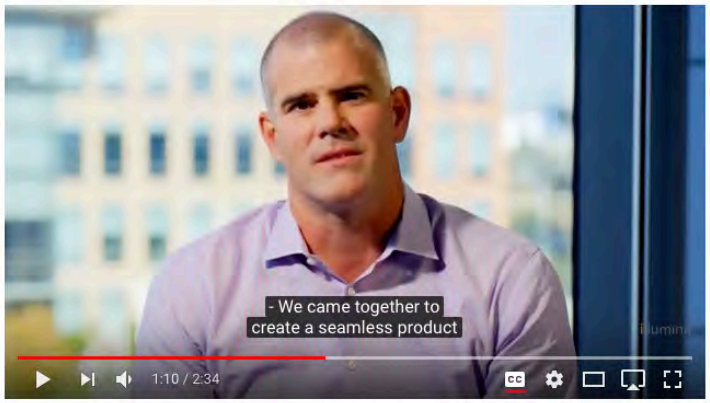
3. (Goswami, TMO) – Today I am really proud to be announcing a partnership with Illumina to bring the power of AmpliSeq library prep across all of Illumina's sequencing platforms…

4. (Oene, ILMN) – For both of us our customers are our number one priority. Our customers need that flexibility to choose the best solutions for their research. I'm really excite to bring AmpliSeq technologies for Illumina sequencers to market.

5. (Oene, ILMN) – This gives our customers access to the high value AmpliSeq chemistries. We've removed the barriers for them to integrate that with their existing technologies.

6. (Goswami, TMO) – Over several years customers on Illumina's platforms have asked for AmpliSeq to be enabled on their instruments and part of that has been the ability of Ampliseq to be able to handle both DNA and RNA in a single workflow.

7. (Goswami, TMO) – To be able to handle complex FFPE samples especially where the quantity of DNA and RNA may be quite small.



8. (Oene, ILMN) – We came together to create a seamless product and so Illumina next gen sequencing customers can order directly through Illumina now and know that we will have that seamless support from the Illumina organization as well…

9. (Oene, ILMN) – Customers regularly ask me if they can use this library prep approach on our sequencers. So I'm really excited to give them the flexibility to do this and add this to their workflows on their systems.

10. (Goswami, TMO) – We're really excited to be able to bring this jointly to our customers. This partnership will have a huge impact by enabling customers across multiple platforms use Ampliseq library prep to be able to conduct their experiments in NGS, but also then be able to more easily compare, share, and collaborate with those results and findings and insights across geographies and platforms.

11. (Oene, ILMN) – We have a wide variety of tools that exist today but customers continue to ask about AmpliSeq and they love the simplicity of this and the performance characteristics of it.

12. (Oene, ILMN) – So really excited to add this to our portfolio of products for them to drive the innovation and understanding of genetics.

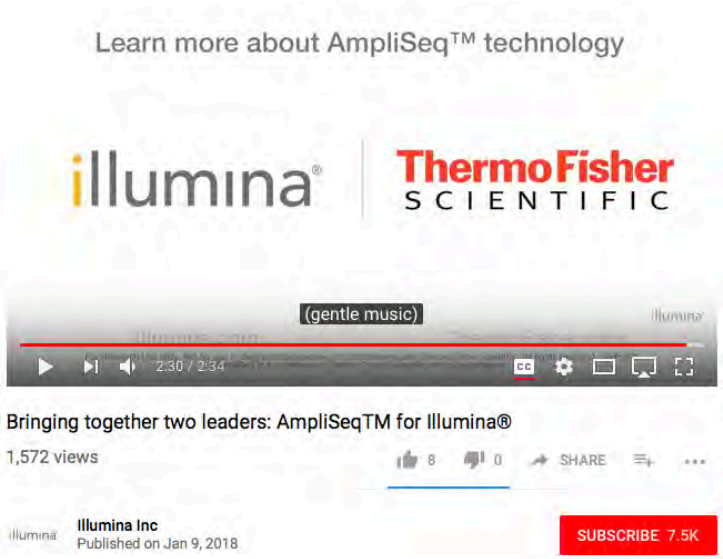


EXHIBIT 3

EXHIBIT 10.13

First Amendment to Joint Development Agreement

This first amendment ("First Amendment"), entered into as of March 27, 2001, ("First Amendment Date"), is by and between ILLUMINA, INC., a Delaware corporation located at 9390 Towne Centre Drive, Suite 200, San Diego, CA 92121-3015 ("Illumina"), and PE CORPORATION (NY), a New York corporation, through its PE Biosystems Group, now known as its Applied Biosystems Group, located at 850 Lincoln Centre Drive, Foster City, CA 94404 ("PEB").

Recitals

The parties have entered into a Joint Development Agreement ("Original Agreement") having an Effective Date of November 8, 1999.

This First Amendment, entered into pursuant to Section 13.7 of the Original Agreement, serves to amend the Original Agreement by providing for the development and commercialization of Tag Sequence technology.

NOW, THEREFORE, in consideration of the mutual obligations in this First Amendment and in the Original Agreement, and for other good and valuable consideration, the receipt and sufficiency of which are acknowledged, the parties to this First Amendment agree as follows:

Amendment

1. Definitions: Any words with initial capitalization that are used and not defined in this First Amendment will have the meanings set forth in the Original Agreement.

2. Add new Section 1.33 immediately after Section 1.32 of the Original Agreement as follows:

1.33 "Tag Sequence" means a set of oligonucleotide probes, developed pursuant to the Original Agreement or this First Amendment, which act independently of any target-sequence-specific analytical chemical reactions to allow the physical addressing of the products of a chemical reaction to locations on a solid support, such as the "addressable array-specific portion" of the oligonucleotide probes and their complements described in International Patent Application No. WO97/31256, and that are designed for use in the Collaboration Product. The Parties will agree on the selection Tag Sequences to be used in the Collaboration Product, subject to the approval of the Joint Steering Committee. ***

25817_1.DOC

* Certain confidential portions of this Exhibit were omitted by means of blackout of the text (the "Mark"). This Exhibit has been filed separately with the Secretary of the Commission without the Mark pursuant to the Company's Application Requesting Confidential Treatment under Rule 24b-2 under the 1934 Act.

CONFIDENTIAL TREATMENT REQUESTED

3. Add the following to Section 3.1 of the Original Agreement immediately before the last sentence of Section 3.1:

> The Parties will share responsibility for defining and developing Tag Sequences for the Collaboration Product which will attempt to avoid third party intellectual property rights or other encumbrances.

4. Add new Sections 4.1.5 and 4.1.6 immediately after Section 4.1.4 of the Original Agreement as follows:

4.1.5 Manufacture of Tag Sequences. It is understood that the Tag Sequences comprise complementary sets of oligonucleotides, one set of which will reside on the Assembled Arrays, and a second, complementary set which will be provided as part of the Reagents. In the Collaboration Field, Illumina will manufacture Tag Sequences for the Assembled Arrays and for use in decoding arrays; and PEB will manufacture Tag Sequences for the Reagents. ***

4.1.6 Use of Tag Sequences. The Parties agree that Tag Sequences will be used in the Collaboration Field by the Parties and purchasers of Collaboration Product. Tag Sequences may not be used by third parties outside the Collaboration Field unless the use is approved by the Joint Steering Committee.

5. Add new Section 6.1.4 immediately after Section 6.1.3 of the Original Agreement as follows:

6.1.4. The Parties agree that any Intellectual Property Rights concerning Tag Sequences conceived after the Effective Date of the Original Agreement, whether Collaboration PEB Intellectual Property, Collaboration Joint Intellectual Property, or Collaboration Illumina Intellectual Property, including the methods by which such Tag Sequences are designed, selected or made, as well as any compositions directed to such Tag Sequences, shall be deemed Collaboration Joint Intellectual Property.

6. Continued Effect of Original Agreement. All provisions of the Original Agreement except as modified by this First Amendment, shall remain in full force and effect and are hereby reaffirmed. Other than as stated in this First Amendment, this First Amendment will not operate as a waiver of any condition or obligation imposed on the Parties under the Original Agreement.

7. Interpretation of Amendment. In the event of any conflict, inconsistency, or incongruity between any provision of this First Amendment and any provision of

the Original Agreement, the provisions of this First Amendment will govern and control.

8. Entire Agreement. This First Amendment, along with the Original Agreement constitute the sole agreements between the Parties relating to the subject matter hereof and supersede all previous writings and understandings. Confidential disclosures made under previously executed Confidentiality Agreements between the Parties will remain subject to the terms of those Confidentiality Agreements. No Party has been induced to enter into this Agreement by, nor is any Party relying on, any representation or warranty outside those expressly set forth in this Agreement.

9. Counterparts. This First Amendment may be executed in any number of counterparts, and each counterpart will deemed an original instrument, but all counterparts together will constitute one agreement.

[Signature page follows.]

The Parties, through their authorized officers, have executed this First Amendment as of the First Amendment Date.

ILLUMINA, INC.	PE CORPORATION (NY), THROUGH ITS APPLIED BIOSYSTEMS GROUP
By: John R. Stuelpnagel	By: Elaine J. Heron
Name: John R Stuelpnagel	Name:
Title: Vice President, Business Development	Title: Vice President, Applera Corporation General Manager, Molecular Biology Divisi Applied Biosystems
Date: 4/2/01	Date: 3/29/01

EXHIBIT 4

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

CORNELL UNIVERSITY, CORNELL RESEARCH FOUNDATION, INC., LIFE TECHNOLOGIES CORPORATION AND APPLIED BIOSYSTEMS, LLC, Plaintiffs, v. ILLUMINA, INC., Defendant.	C.A. No. 10-433-LPS-MPT ORAL ARGUMENT REQUESTED REDACTED PUBLIC VERSION

PLAINTIFFS CORNELL UNIVERSITY AND CORNELL RESEARCH FOUNDATION, INC.'S OPENING BRIEF IN SUPPORT OF MOTION TO VACATE STIPULATION OF DISMISSAL PURSUANT TO RULE 60 AND TO RESCIND SETTLEMENT DOCUMENTS [redacted] OR, ALTERNATIVELY, FOR LEAVE TO CONDUCT DISCOVERY

Dated: June 21, 2017

LECLAIRRYAN
Andrew L. Cole (No. 5712)
800 North King Street, Suite 303
Wilmington, DE 19801
(302) 394-6817

Counsel for Plaintiffs Cornell University and Cornell Research Foundation, Inc.

TABLE OF CONTENTS

PAGE

TABLE OF AUTHORITIES

PAGE

CASES

RULES

NATURE AND STAGE OF THE PROCEEDINGS

Plaintiffs Cornell University and Cornell Research Foundation, Inc. (referred to collectively as "Cornell" and individually as "Cornell University" and "CRF") are plaintiffs in this action, jointly with Life Technologies Corporation and Applied Biosystems, LLC (collectively "Life Tech"). The plaintiffs brought this lawsuit (the "Action") against defendant Illumina, Inc. ("Illumina") alleging infringement of certain patents. The parties entered into a Settlement Agreement effective April 14, 2017 (the "Settlement Agreement"), and a stipulation of dismissal pursuant to the Settlement Agreement [D.I. 598] (the "Dismissal") was entered on April 24, 2017. Cornell now files this opening brief in support of its Motion to Vacate the Dismissal Pursuant to Rule 60 and to Rescind Settlement Documents [redacted] or, alternatively, for Leave to Conduct Discovery.

SUMMARY OF THE ARGUMENT

1. Pursuant to Rule 60 of the Federal Rules of Civil Procedure, Cornell seeks vacatur of the Dismissal entered in this Action, [redacted]

2. As the Court is aware, Cornell and Life Tech are co-plaintiffs in this Action directed against specific Illumina products (the "Accused Products"), that plaintiffs believe infringe certain patents owned by CRF and exclusively licensed to Life Technologies Corporation (the "Asserted Patents"). [redacted]

[redacted]

3. [redacted]

[redacted]

[redacted]

4. [redacted]

[redacted]

5. [redacted]

[redacted]

STATEMENT OF FACTS

A. The Parties, This Action, [redacted]

[redacted]

[redacted]

As the Court is aware, this Action was filed in 2010 with Cornell and Life Tech as co-plaintiffs. Dorn ¶ 9. Throughout this Action, the co-plaintiffs have been represented by outside counsel selected by Life Tech. Dorn ¶ 13. [redacted]

As discussed further below, Latham & Watkins was also counsel for Life Tech in a patent infringement lawsuit brought by Illumina against Life Tech in California. *Id.* ¶ 14. [redacted]

This Action is directed against specified Illumina products which plaintiffs assert infringe 15 patents licensed to Life Technologies Corporation [D.I. 298/299], and which will be referenced herein as the "CRF patents." Dorn ¶¶ 2, 10; Kelly ¶¶ 1, 6; Li ¶¶ 1, 6. In addition, Cornell believes [redacted]

B. The Settlement Discussions

[redacted]

[redacted]

[redacted]

C. The Dismissal Of The Action

Pursuant to the Settlement Agreement, on April 18, 2017, the Dismissal was submitted to the Court along with the parties' joint proposed stipulation to vacate the claim construction order and decisions in this Action. [D.I. 595 and 596]. The Court held a telephonic hearing on April 24, 2017 to discuss the Dismissal and the proposed joint stipulation to vacate the claim construction order and decisions. Dorn ¶ 47. During that teleconference, Delaware counsel of record for Life Tech and Cornell represented that in the context of the settlement, Cornell was seeking to preserve its rights, including by vacatur of the Court's claim construction order and decisions, to avoid "a possible preclusive effect on them in any other litigation unrelated to Illumina or the case that is before Your Honor." [D.I. 600, Tr. at 4]. *Id.* The Court entered the Dismissal that same day [D.I. 598], but declined to vacate the claim construction order and

decisions. *Id.*

[redacted]

D. [redacted]

[redacted]

[redacted] On that date, Illumina and Life Tech filed a Joint Motion to Dismiss a patent lawsuit brought by Illumina against Life Tech in the Southern District of California. *Id.*; Kelly ¶ 25. *See Illumina, Inc. v. Life Technologies Corp. et al.*, No. 11-CV-03022 (S.D. Cal.) (the "Illumina California Action"), stating that the: "parties have reached a Settlement Agreement to resolve this dispute without further litigation." (Case No. 11-CV-03022, D.I. 91 at 2). *Id.* Latham & Watkins was counsel of record for Life Tech in the Illumina California Action. Dorn ¶ 51. [redacted]

ARGUMENT

A. Standards Under Rule 60 And Delaware Law On Rescinding A Contract [redacted]

1. Rule 60 Standards

Rule 60(b) provides that a party may file a motion for relief from a final judgment for numerous reasons, including:

[redacted] or (6) any other reason that justifies relief. Fed. R. Civ. P. 60(b) (2017). A motion filed pursuant to Rule 60(b) is addressed to the sound discretion of the trial court guided by accepted legal principles applied in light of all relevant circumstances. *Pierce Assoc. Inc. v. Nemours Found.*, 865 F.2d 530, 548 (3d Cir. 1988); *Lasky v. Cont'l Prods. Corp.*, 804 F.2d 250, 256 (3d Cir. 1986).

[redacted]. As a result, Cornell seeks relief under Rule 60(b)(6), which "is a catch-all provision that allows relief for any other reason justifying relief from the operation of the judgment." *United States v. Witco Corp.*, 76 F. Supp. 2d 519, 527 (D. Del. 1999) (quotation marks omitted). A motion for relief under Rule 60(b)(6) is appropriate "when the movant shows any ... reason justifying relief from the operation of the judgment other than the more specific circumstances set out in Rules 60(b)(1)-(5)." *Gonzalez v. Crosby*, 545 U.S. 524, 528 (2005) (citation and quotation marks omitted). Cases under Rule 60(b) have required "a movant seeking relief under Rule 60(b)(6) to show 'extraordinary circumstances' justifying the reopening of a final judgment." *Id.* at 535 (citation and quotation marks omitted).

2. Jurisdiction Over Settlement Documents

"It is well settled that a federal court has the inherent power to enforce and to consider challenges to settlements entered into in cases originally filed therein." *Fox v. Consol. Rail Corp.*, 739 F.2d 929, 932 (3d Cir. 1984) (quoting *Pearson v. Ecological Sci. Corp.*, 522 F.2d 171 (5th Cir. 1975); *cf. Washington Hosp. v. White*, 889 F.2d 1294, 1298-99 (3d Cir. 1989). The Court can do this without inquiring into, or requiring, an independent grounds for subject matter

jurisdiction for the enforcement/challenge suit. *See Aro Corp. v. Allied Witan Co.*, 531 F.2d 1368, 1371 (6th Cir. 1976). Accordingly, this Court is empowered to consider both vacating the Dismissal and rescinding the agreements entered into in connection with the settlement.

3. [redacted]

[redacted]

B. The Dismissal Should Be Vacated [redacted]

1. [redacted]

[redacted]

a. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

b. [redacted]

[redacted]

[1] Further, although not directly relevant to this motion, [redacted]

[redacted]

2. [redacted]

a. [redacted]

[redacted] Exhibit B-1 to Dorn Exhibit B is the same as the Dismissal entered in this Action. *Id.* Thus, there can be no dispute that the Dismissal was submitted pursuant to the Settlement Agreement. [redacted]

[redacted]

[2] [redacted]

[redacted]

b. Extraordinary Circumstances Are Present And Cornell Cannot Obtain Relief Under Any Other Provision of Rule 60

To vacate a judgment under Rule 60(b)(6), the movant needs to show "extraordinary circumstances." *Gonzalez,* 545 U.S. at 535. The circumstances here are certainly extraordinary.

[redacted]

[redacted]

C. If The Court Believes Further Evidence Is Needed To Support The Relief Requested, Cornell Should Be Granted Leave To Conduct Discovery

[redacted] However, Cornell has had no opportunity to seek discovery on Life Tech's conduct to obtain further evidence. Cornell has presented more than sufficient proofs to support the relief sought in this motion; however, if the Court believes more evidence should be presented, Cornell respectfully requests that the Court grant it the opportunity to conduct discovery into the issues presented in this motion. [redacted]

Courts have repeatedly granted requests for discovery in the context of Rule 60 motions. [redacted] Many other courts have found that discovery is warranted when there is a dispute on the facts relating to relief requested under Rule 60. *See. e.g., Sheng v. Starkey Labs., Inc.*, 53 F.3d 192, 194 (8th Cir. 1995) (holding that evidentiary hearing had to be held to resolve both Rule 60 motions); *Southerland v. Irons*, 628 F.2d 978, 979 (6th Cir.

1980) (court scheduled evidentiary hearing on Rule 60 motion); *Midwest Franchise Corp. v. Metromedia Rest. Grp., Inc.*, 177 F.R.D. 438, 441 (N.D. Iowa 1997) (holding that plaintiff had made necessary showing to warrant post-trial deposition of witness pursuant to Rule 60(b)(2) motion); [redacted]

CONCLUSION

For the above reasons, the Dismissal should be vacated and the Settlement Agreement and the [redacted] or in the alternative, Cornell should be granted leave to conduct discovery before the Court rules on this motion.

OF COUNSEL:
LECLAIRRYAN, A Professional Corporation
Andrew P. Zappia
70 Linden Oaks, Suite 210
Rochester, NY 14625
(585) 270-2100
Andrew.Zappia@leclairryan.com

Counsel for Plaintiffs Cornell University and Cornell Research Foundation, Inc.

LECLAIRRYAN

By: s/ Andrew L. Cole
Andrew L. Cole. (No. 5712)
800 North King Street, Suite 303
Wilmington, DE 19801
(302) 394-6817
Andrew.Cole@leclairryan.com

Counsel for Plaintiffs Cornell University and Cornell Research Foundation, Inc.

June 21, 2017

EXHIBIT 5

Illumina and PE Biosystems colluded to defraud Zip Code Inventors

Background and Chronology for Attempts by Illumina Employees to Re-Patent Zip Code Inventors Intellectual Property

Timeline for RICO Complaint

- January, 2018 – Illumina CEO Francis deSouza admits secret collaboration with Thermo Fisher "started the conversation clearly well over a year ago"… meaning prior to the April 2017 settlement agreement (see below). Discovery will reveal if this secret collaboration started over 3 years earlier in 2015, i.e. prior to the Markman and bulk of the *Cornell v Illumina* case (1:10-cv-00433-LPS).

- April, 2017 – Life Technologies (ThermoFisher) and Illumina fraudulently colluded to induce settlement of *Cornell v Illumina* in Delaware (1:10-cv-00433-LPS). This eventually was contested by a Rule 60(b)(6) Motion filed by Cornell once they became aware of Illumina's and ThrmoFisher's deception. This triggered a series of FOIA requests on Illumina SEC filings and NIH Grants in April and May 2017 by Plaintiff Zirvi.

- January, 2015 – Illumina in-house attorney, William Noon, files a highly unusual FOIA request to the NIH for copies of Illumina's own NIH SBIR Grants:
  - 1R21HG001911-01
  - 1R44HG002003-01
  - 1R43CA081952-01
  - 1R43CA083398-01

January 8, 2018 - Illumina unexpectedly announces collaboration with Thermo Fisher with a joint product: “Ampliseq for Illumina”.

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 8-K

CURRENT REPORT
PURSUANT TO SECTION 13 OR 15 (d) OF THE SECURITIES EXCHANGE ACT OF 1934

Date of report (Date of earliest event reported): January 8, 2018

Illumina, Inc.

(Exact name of registrant as specified in its charter)

But internal innovation is only part of the story. Our partnership strategy ensures that our customers will have access to best-in-class technology. With that in mind, I am very pleased to announce an exciting partnership with Thermo Fisher. This collaboration deeply integrates the leading and most trusted sequencers with the leading amplicon library prep technology. Thermo Fisher has 20 years of experience in PCR amplification chemistries that has enabled them to develop an amplicon library prep that is simple, fast and robust. The technology has been widely adopted in the oncology space because of the high-quality data achievable with low input and degraded samples, such as FFPE tissue.

But until now, those customers have not been able to take advantage of the accuracy and power of Illumina sequencers. This agreement provides direct access to AmpliSeq for our RUO customers, delivering the best of sequencing with the best amplicon protocol. Ahead of today's launch, we have been working together to adapt and optimize AmpliSeq specifically for Illumina sequencers, creating a seamless user experience that is fast and easy.

In addition to a collection of ready-to-order panels, Illumina customers will also be able to access AmpliSeq algorithms to create custom amplicon assays utilizing DesignStudio, our online assay design protocol, and they will have access to future panels and new assay types that are introduced as part of the AmpliSeq product line.

Amplicon sequencing is an important entry point for new NGS customers. The combination of our proven technologies further removes barriers to adoption, bringing more new-to-sequencing customers to our platform. Of course, we also have an installed base of more than 9,000 desktop systems. By giving our existing customers access to AmpliSeq chemistry, we're enabling them to do even more with their systems.

January 8, 2018 - Illumina CEO Francis deSouza admits secret collaboration with Thermo Fisher "started the conversation clearly well over a year ago"… meaning prior to the settlement agreement.

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 8-K

CURRENT REPORT
PURSUANT TO SECTION 13 OR 15 (d) OF THE SECURITIES EXCHANGE ACT OF 1934

Date of report (Date of earliest event reported): January 8, 2018

Illumina, Inc.
(Exact name of registrant as specified in its charter)

Tycho W. Peterson - JP Morgan Chase & Co, Research Division - Senior Analyst

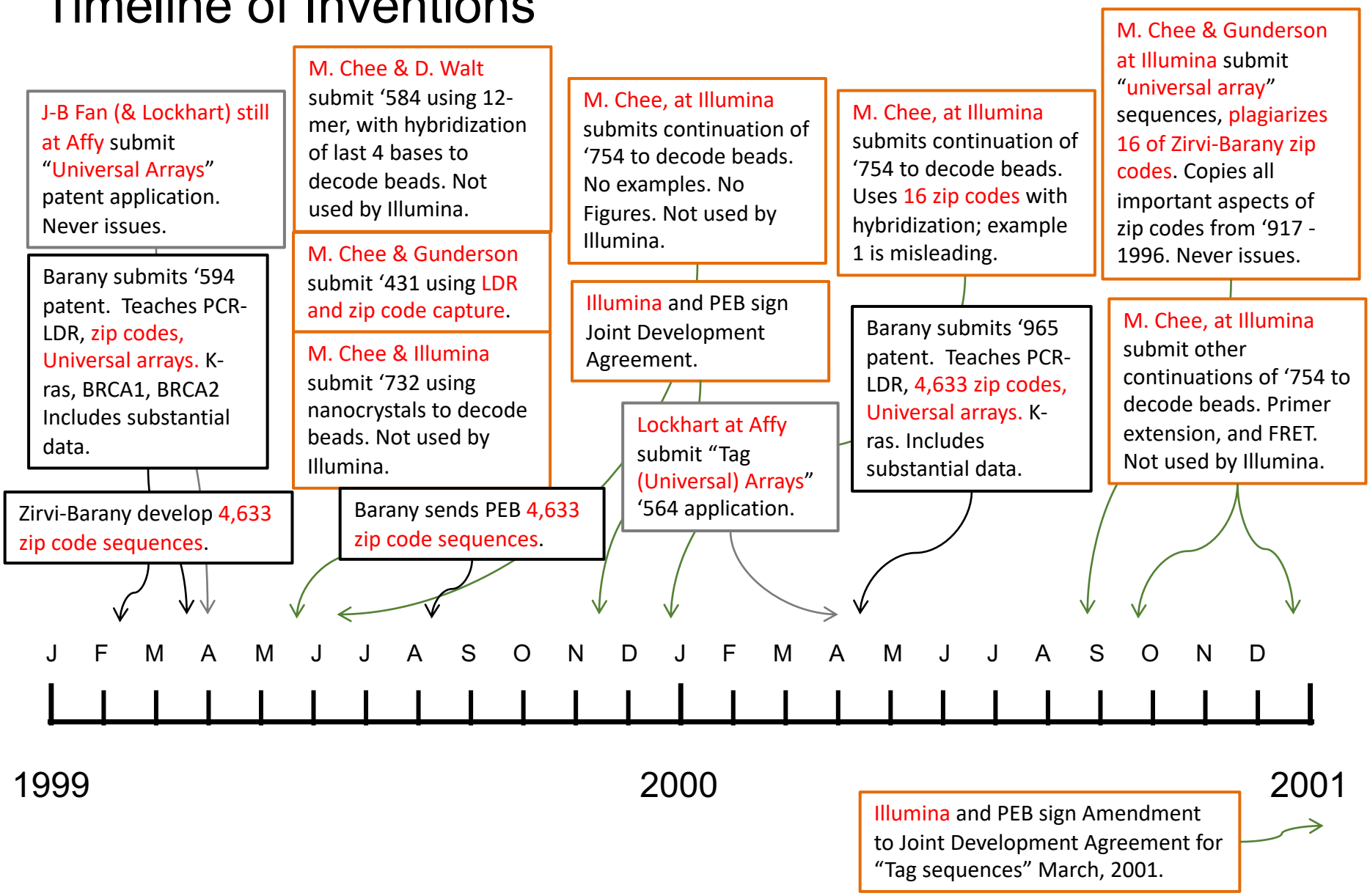
Can you talk about just that Thermo agreement, how that came about? (inaudible) press releases in Illumina (inaudible), so just a bit of speculation on that part? I mean, what's the back story?

Francis A. deSouza - Illumina, Inc. - CEO, President & Director

So the question is about the partnership we have with Thermo and what's the back story or how did that come about. The way it came about is as we talk to our customers around what they are looking for, one of the things that we heard consistently is that they liked certain aspects of AmpliSeq. And so they would tell us, look, they love the low sample they put in AmpliSeq, for example. They like the ability to work with degraded samples like FFPE tissue, and AmpliSeq has a following in the oncology community in terms of people like that assay. And so we spent a lot of time internally thinking, look, we want to make sure our customers have the best experience in our sequencer. So what is great for our customers is ultimately good for Illumina.

And so we started the conversation clearly well over a year ago. And initially, it was one of those, like, "Are you sure -- are we sure we want to do this? Are they sure they want to do this?" But it was always -- so the true north for us was what's the best thing for our customers? And if you keep looking at that, the truth is the best thing for our customers is to make the best amplicon technology available on the best sequencer. And I think a lot of credit on both sides. Thermo had the same thinking, which is, yes, there are parts of our portfolio we compete, but this is clearly good for AmpliSeq, it's good for customers and it's good for us. And so that was the thinking behind it. I think we're out of time.

Timeline of Inventions



Timeline of Inventions

M. Chee, becomes Director of Genomic Research at Affy
Was aware of "Tag (Universal) Arrays", patent application. Never issues.

D. Walt of Tufts, M. Chee, and J. Stuelpnagel form Illumina
No mention of universal array or zip codes.

D. Walt of Tufts submits '884 patent claims dye encoding of beads. Not used by Illumina. No mention of universal array or zip codes.

M. Chee, at Illumina submits '754 application to decode beads. Example relates to enzyme-based sensor. Not used by Illumina.

M. Chee, at Illumina submits '431 application to decode microsphere array sensors. Example relates to enzyme-based sensor. Not used by Illumina.

K. Gunderson & M. Chee, from time at Affy submit '913 application, sequencing by ligation on array. Never issues.

M. Chee, at Illumina submits continuation of '754 to decode beads. Not used by Illumina.

J F M A M J J A S O N D J F M A M J J A S O N D

1997 1998 1999

Timeline of Inventions

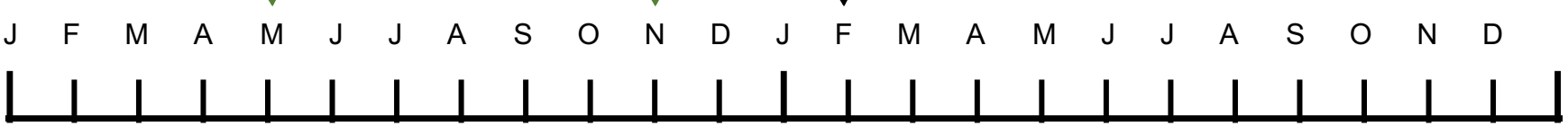
D. Walt admits he is not co-inventor of zip code chemistry or universal arrays.

D. Walt of Tufts publishes use of beads and a hexagonal fiber-optic layout to detect DNA. Relies on pre-hybridization with known sequences to mRNA. Does not use zip codes.

Barany collaborates with D. Walt of Tufts. Barany explains zip code chemistry, ligase detection, and their use on addressable arrays for encoding and decoding DNA.

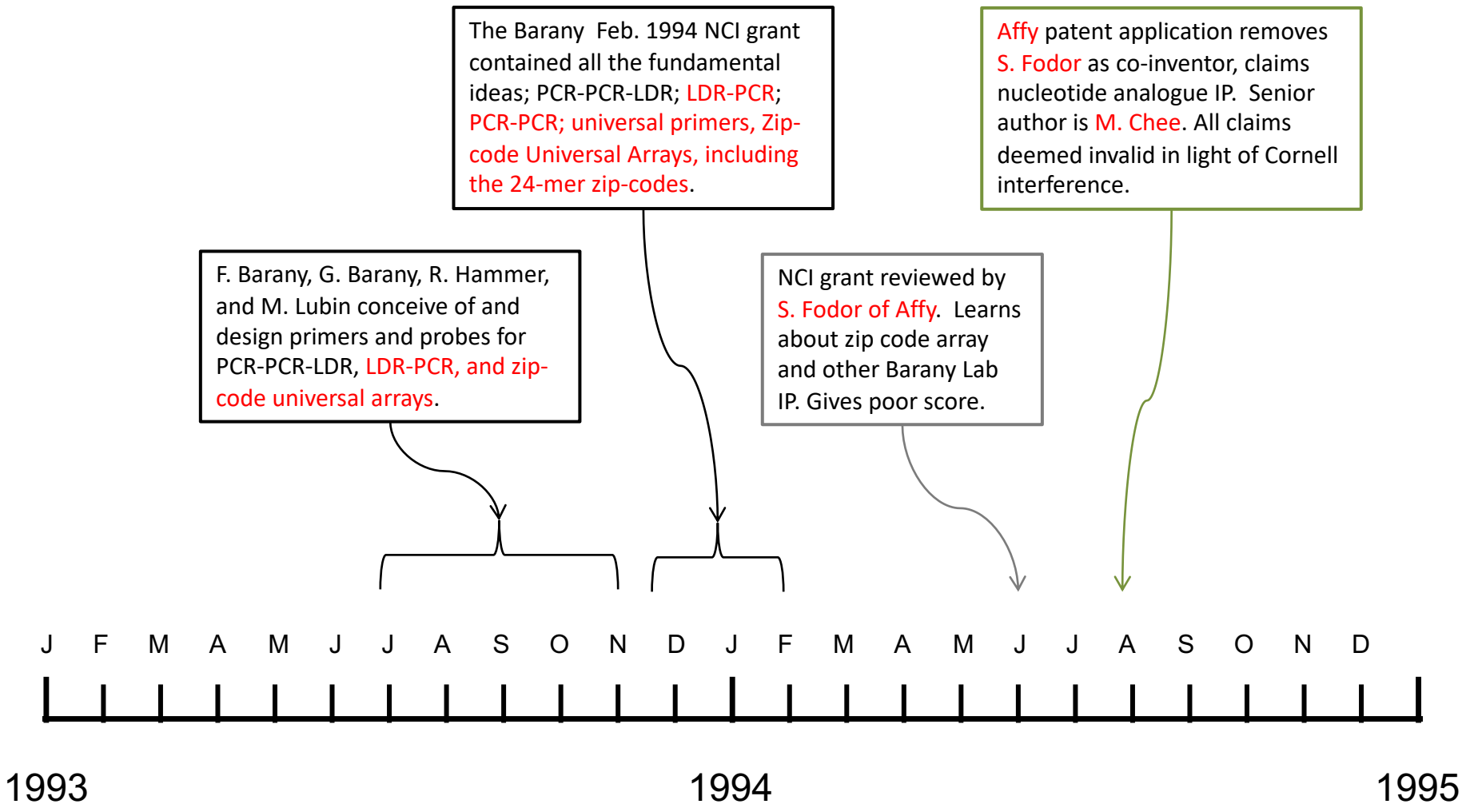
Barany submits '917 series patent. Teaches PCR-LDR, zip codes, Universal arrays. Includes substantial data.

Barany submits '470 series patent. Teaches PCR-PCR-LDR; LDR-PCR; PCR-PCR; zip code capture; zip-code primers. Includes substantial data.



1995 1996 1997

Timeline of Inventions



Concluding Remarks on the Timeline

1. Affymetrix employee Steve Fodor had *detailed knowledge* about the confidential Barany 1994 NCI grant submission, which described zip code chemistry, zip code primers, combining PCR with ligase detection reactions (LDR)and universal arrays.

2. David Walt, founder of Illumina (with Mark Chee and others, April 1998) reads draft of Barany's '917 patent application, and acknowledges to Mike Goldman on November 22nd, 1995 that he is ***not*** an inventor of zip code arrays.

3. Mark Chee, Kevin Gunderson, and J-B. Fan all worked at Affymetrix, and knew of zip code arrays *prior* to joining Illumina.

4. Affymetrix employees, including J-B. Fan, submit patent applications on March 26th, 1999, and April 6th, 2000 (as well as publications) using Universal arrays of 20-mer zip code sequences, they termed "Tag arrays" and "Tag Sequences".

5. David Walt and Mark Chee submit patent applications on May 20th, 1999, both for Illumina and Tufts describing other ways to encode for bead-arrays, i.e. nano-crystals, or sequencing by hybridization of 12-mers. In fact, Illumina failed to commercialize 7 of their own approaches for decoding which bead is in a given location.

6. Mark Chee, Kevin Gunderson, and J-B. Fan, submit several patent applications and publications using Universal arrays, with 24-mer zip codes AFTER the Nov. 9th, 1999 collaboration with ABI (PE-Biosystems) was announced.

7. Mark Chee and Kevin Gunderson had access to the confidential Zirvi-Barany *unpublished* 465 zip-code list – which was never published – from PEB some time between February 1999, (when Zirvi-Barany designed them and provided to PEB shortly thereafter,) and August 25th, 2000, when Illumina tried to re-patent Zirvi-Barany's zip codes.

8. Mark Chee and other Illumina employees file provisional application 60172106 on December 23rd, 1999 (claimed in both US7033754 and US7226734) where in Example 1, they claim to demonstrate identification of "16 zip codes" through 4 decoding steps, but knowingly and willfully conceal zip code sequences as well as hybridization and washing conditions used.

9. Mark Chee and Kevin Gunderson submit a patent application on August 25th, 2000 where they claimed to "generate a list of about 4,000 zip-codes, with special properties", which they call "Illumacodes" – such special properties were invented by Barany Lab in the '917 application of 1996. In Table 2, the first 16 zip codes are EXACTLY the same sequence as 16 of the first 52 Zirvi-Barany 24-mer zip-codes of the *unpublished* 465 list, and are in the identical order as that list, and *thus were plagiarized*.

Illumina and PE Biosystems Collaboration Requires Barany Lab's Zip Code Chemistry (covered by WO 97/31256 patent) for David Walt's Bead Arrays to Work

Joint Development Agreement, November 9th, 1999

JOINT DEVELOPMENT AGREEMENT

This Joint Development Agreement ("Agreement") dated as of the _____day of November, 1999 ("Effective Date") is by and between ILLUMINA, INC., a California corporation, located at 9390 Towne Centre Drive, Suite 200, San Diego, CA 92121- 3015 ("Illumina"), and PE CORPORATION, a Delaware corporation, through its PE Biosystems Group, located at 850 Lincoln Centre Drive, Foster City, CA 94404 ("PEB").

1.22. "Assembled Array" means an array of microspheres having chemical functionality attached thereto distributed on a patterned substrate, as generally described in U.S Patent Application No.08/818,199.

1.23. "Zip Code Chemistry" means a nucleic acid sequence detection method employing a sequence-specific hybridization pull-out step subsequent to a chemical or enzymatic polynucleotide ligation reaction, as generally described in International Patent Application No. WO 97/31256.

Illumina makes <u>randomly</u> "*Assembled* Arrays", invented by Dr. David Walt, while Dr. Francis Barany and team invented "Zip Code Chemistry" to make "*Addressable* Arrays" . On their own, randomly assembled bead arrays do not work for DNA analysis like SNP genotyping. The zip code chemistry of Barany Lab's IP not only works on traditional DNA arrays, but is also <u>the</u> enabling technology that allows the random bead arrays to work for DNA analysis.

Barany teaches Walt about ligation assays, zip codes, and Universal arrays:

Date: Mon, 1 May 1995 17:13:04 -0400
From: Francis Barany <barany@mail.med.cornell.edu>
To: dwalt@pearl.tufts.edu
Cc: barany@maroon.tc.umn.edu, chammer@sn01.sncc.lsu.edu,
dbergstr@mace.cc.purdue.edu, hjblok@mail.med.cornell.edu
Subject: Derivatized Glass

Dear David,

It was a pleasure to speak with you last week. You will be receiving a FedEx package from me some time this week which will include a confidential copy of our zip code approach, and some special glass slides for making the first test systems. These glass slides are a special thickness and will be used in a Perkin Elmer In-Situ PCR machine. I haven't received the slides yet, but as soon as I do, they will be sent to you.

I have discussed with my co-workers what type of functionality we would like on the polyethylene surfaces you can place on specific regions of glass using your optic fiber system. The early consensus is a COOH group so we can link it to our Amino-link oligonucleotides using water soluble carbocarbodi-imide activation. I'm not quite sure of the efficiency of this reaction, so I was wondering if you could make an activated group directly on your surface, so I could just react the amino-link oligo directly to your group. Perhaps you have some other suggestions. The COOH group will ultimately be used for building arrays using PNA oligomers on the surface.

I would also like to provide you with the phone numbers and E.mail addresses of the other collaborators on the zip code capture team:

Francis Barany, 212-746-6509, barany@mail.med.cornell.edu
George Barany, 612-625-1028, barany@maroon.tc.umn.edu
Donald Bergstrom, 317-494-6275, dbergstr@mace.cc.purdue.edu
Herman Blok, 212-746-6507, hjblok@mail.med.cornell.edu
Robert Hammer, 504-388-4025, chammer@sn01.sncc.lsu.edu

I'm looking forward to collaborating with you.

Francis

TUFTS UNIVERSITY

Department of Chemistry

November 22, 1995

Francis Barany, Ph.D.
Cornell University Medical College
Department of Microbiology
1300 York Avenue, Box 62
New York, NY 10021

Michael L. Goldman
Nixon, Hargrave, Devans & Doyle LLP
Clinton Square
Post Office Box 1051
Rochester, NY 14603-1051

Dear Francis:

Thank you for sending a copy of the patent DETECTION OF SINGLE BASE DIFFERENCES USING THE LIGASE DETECTION REACTION WITH ADDRESSABLE ARRAYS for my review. After going through the patent, we have found several protocols and/or references to work that my laboratory has contributed. Since we are not inventors on the patent we respectfully ask that any work that we have contributed to the patent be removed. Because we use these techniques in much of our research we do not wish to limit the scope of interest to which we can apply these techniques. Here is a list of places in the patent that we feel should be deleted:

pg. 25, line 34 - pg. 28, line 2
pg. 31, line 19
pg. 31, line 35
pg. 53, line 25 - pg. 55, line 26
pg. 56, line 36 - pg. 57, line 2
claims 33, 84, 110, 130 remove 3-aminopropyl methacrylate amide
claims 34, 85, 111, 131 remove 2-hydroxyethyl methacrylate
remove claims 115
delete Figures 20, 21

Thank you for your understanding in this matter. Please contact me if you have any questions.

Sincerely,

David R. Walt

62 Talbot Ave.
Medford, Massachusetts 02155
Telephone (617) 627-3441
Facsimile (617) 627-3443

David Walt admits that he did not invent zip codes and Universal arrays.

Illumina and PE Biosystems Collude to Defraud Zip Code Inventors

First Amendment to Joint Development Agreement, March 27th, 2001

Recitals

The parties have entered into a Joint Development Agreement ("Original Agreement") having an Effective Date of November 8, 1999.

This First Amendment, entered into pursuant to Section 13.7 of the Original Agreement, serves to amend the Original Agreement by providing for the development and commercialization of Tag Sequence technology.

CONFIDENTIAL TREATMENT REQUESTED

2. **Add** new Section 1.33 immediately after Section 1.32 of the Original Agreement as follows:

1.33 "Tag Sequence" means a set of oligonucleotide probes, developed pursuant to the Original Agreement or this First Amendment, which act independently of any target-sequence-specific analytical chemical reactions to allow the physical addressing of the products of a chemical reaction to locations on a solid support, such as the "addressable array-specific portion" of the oligonucleotide probes and their complements described in International Patent Application No. WO97/31256, and that are designed for use in the Collaboration Product. The Parties will agree on the selection Tag Sequences to be used in the Collaboration Product, subject to the approval of the Joint Steering Committee.

***

Illumina and PE Biosystems, deliberately colluded to provide "for the development and commercialization of Tag Sequence technology", which is really the "Zip Code Chemistry" invented in the Barany Lab. There is one and only one logical reason to redact the definition of "Tag Sequence" in this SEC document: To prevent the true inventors from finding out that Illumina and PE Biosystems were colluding to develop and commercialize sequences that infringed on the inventor's intellectual property. This was part of a deliberate strategy to re-name the inventors's zip codes with other names to obfuscate their origin.

First Amendment to Joint Development Agreement, March 27th, 2001

CONFIDENTIAL TREATMENT REQUESTED

3. Add the following to Section 3.1 of the Original Agreement immediately before the last sentence of Section 3.1:

The Parties will share responsibility for defining and developing Tag Sequences for the Collaboration Product which will attempt to avoid third party intellectual property rights or other encumbrances.

4. Add new Sections 4.1.5 and 4.1.6 immediately after Section 4.1.4 of the Origina Agreement as follows:

4.1.5 Manufacture of Tag Sequences. It is understood that the Tag Sequences comprise complementary sets of oligonucleotides, one set of which will reside on the Assembled Arrays, and a second, complementary set which will be provided as part of the Reagents. In the Collaboration Field, Illumina will manufacture Tag Sequences for the Assembled Arrays and for use in decoding arrays; and PEB will manufacture Tag Sequences for the Reagents.

***

Further, Illumina and PE Biosystems, deliberately colluded to develop "Tag Sequences for the Collaboration, which will attempt to avoid third party intellectual property rights or other encumbrances", in other words, to defraud the rightful inventors of royalties for the "Zip Code Chemistry". As this sentence was (accidentally) left un-redacted, when we asked about this section, Dr. Barany were falsely told by the lawyers that this did not refer to the true inventors intellectual property. The redacted section (4.1.5) clearly shows that Illumina and PE colluded to divide the responsibilities for manufacturing Tag sequences on assembled arrays (Illumina) and Tag sequences for reagents (PE). Since Tag sequences are synonymous to zip codes (section 1.33), uncovering this redacted section demonstrates that Illumina knew all along that zip codes work on both the array and in solution.

First Amendment to Joint Development Agreement, March 27th, 2001

5. **Add new Section 6.1.4 immediately after Section 6.1.3 of the Original Agreement as follows:**

6.1.4. **The Parties agree that any Intellectual Property Rights concerning Tag Sequences conceived after the Effective Date of the Original Agreement, whether Collaboration PEB Intellectual Property, Collaboration Joint Intellectual Property, or Collaboration Illumina Intellectual Property, including the methods by which such Tag Sequences are designed, selected or made, as well as any compositions directed to such Tag Sequences, shall be deemed Collaboration Joint Intellectual Property.**

The Parties, through their authorized officers, have executed this First Amendment as of the First Amendment Date.

ILLUMINA, INC.	PE CORPORATION (NY), THROUGH ITS APPLIED BIOSYSTEMS GROUP
By: John R. Stuelpnagel	By: Elaine J. Heron
Name: John R Stuelpnagel	Name:
Title: Vice President, Business Development	Title: Vice President, Applera Corporation General Manager, Molecular Biology Divisi Applied Biosystems
Date: 4/2/01	Date: 3/29/01

Illumina and PE Biosystems, deliberately colluded to jointly own “any Intellectual Property rights concerning Tag Sequences”,… “including the methods by which such Tag Sequences are designed, selected or made, as well as any compositions directed to such Tag sequences.” In other words, Illumina and PE Biosystems acknowledge the high value of “Tag sequences”, which per the redacted section 1.33 “Tag sequences mean” is really the intellectual property of WO97/31256, i.e. the “Zip Code Chemistry” invented by the Barany Lab.

Illumina's GoldenGate Assay and Tag Sequences Infringe on the Zip Code inventors Intellectual Property

Illumina – PE-Biosystems settlement agreement, August 18th, 2004

*CONFIDENTIAL
TREATMENT REQUESTED*

EXHIBIT 1

Description of the GoldenGate™ Assay

Illumina's GoldenGate™ assay involves all of the following steps:

Hybridization of at least two oligonucleotide probes to a target nucleic acid material, where the target material may be DNA, cDNA, RNA, or artificial oligonucleotide template;

Extension and ligation of one probe to the other while the probes are hybridized to the target material;

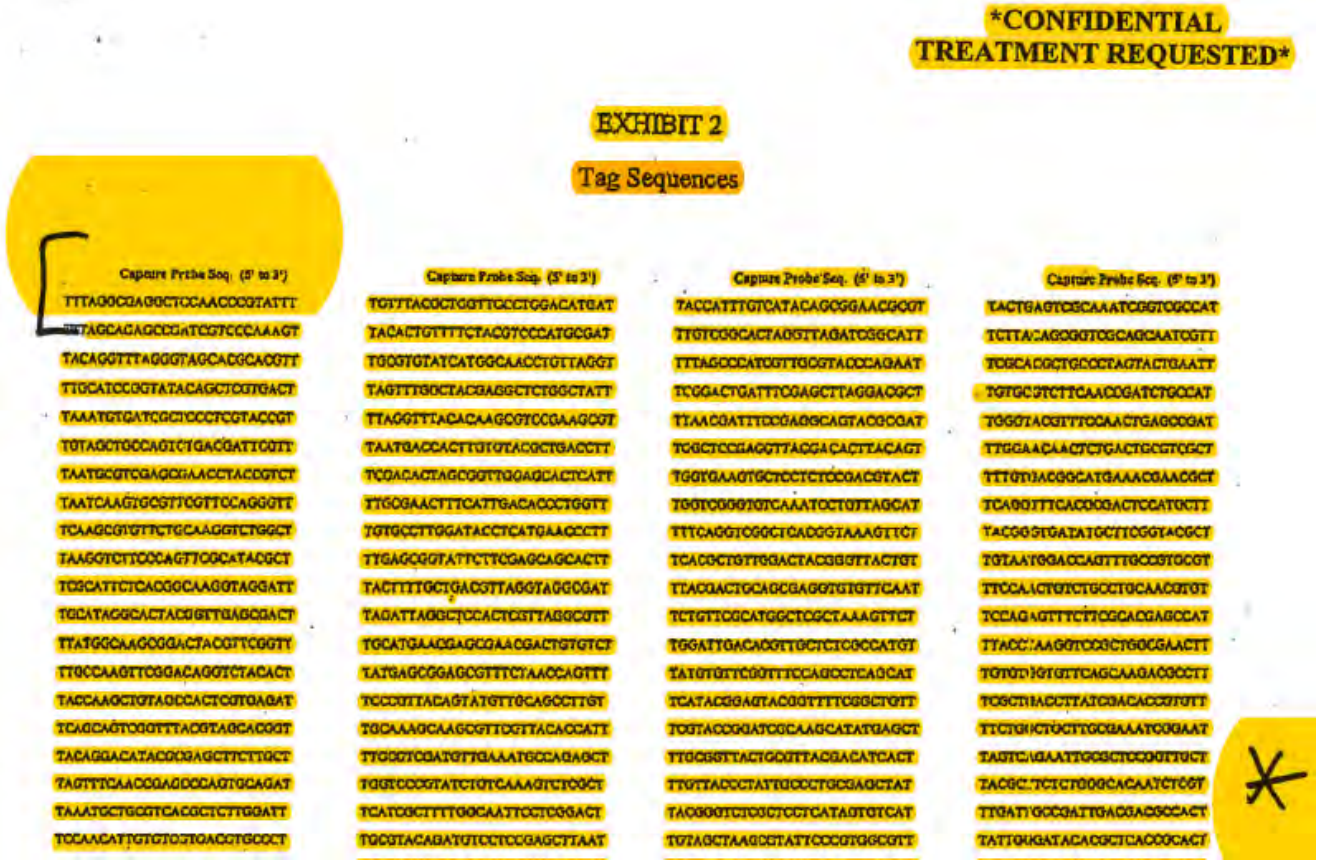
Eluting the extended and ligated probes from the target material;

Universal PCR, rolling circle, random priming, T4/Eberwine, strand displacement, TMA (transcripted mediated amplification), LCR (ligase chain reaction), MDA (multiple displacement amplification) or SPIA amplification of the extended and ligated probes to generate labeled amplicons; and

Detection of the amplicons on an Illumina BeadArray.

In this redacted section, Illumina admits that their "GoldenGate™" assay, uses Barany's LDR-PCR-Zip Code Chemistry technology, which is covered in the '917, '470, and '293 series patents. Again, the rightful inventors were harmed by deliberately redacting this section. Thus, Illumina deliberately set out to defraud the true inventors of royalties due from patents and trade secrets covering Barany's "Zip code chemistry", "Ligase Detection", and "Addressable Arrays" inventions.

Illumina – PE-Biosystems settlement agreement, August 18th, 2004



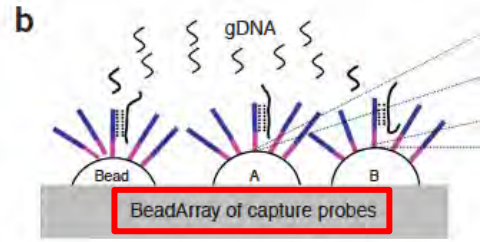
Illumina and/or PE Biosystems, deliberately redacted the actual "Tag Sequences" from the 2004 settlement agreement, to deprive the true inventors of rightful royalties. The recent FOIA request revealed "Tag sequences" be 24 to 26 + 2 bases in length (addition of an extra "T" to both the 3' and 5' end does not materially affect the performance characteristics of these sequences); Tag sequences differ from each other by 25% or more within those 24-26 bases, and are designed to operate under uniform hybridization conditions, in other words, they literally infringe upon the intellectual property of WO97/31256, i.e. the "Zip Code Chemistry" invented by the Barany Lab.

Illumina Plagiarizes the Zirvi-Barany Zip Codes

Zip-code Array, Zip, Decoder sequence, IllumaCode, Illumicode, Probe sequence, BeadArray capture probes, Adapter Sequence,

TABLE 1

IllumaCode ID	Decoder (5'-3')
17	GGCTGGTTCGGCCCGAAAGCTTAG
18	GTTCCCAGTGAAGCTGCGATCTGG
19	TACTTGGCATGGAATCCCTTACGC
20	ACTAGCATATTTCAGGGCACCGGC

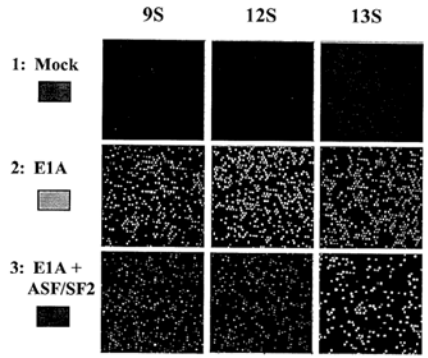


Detection of Alternative Splicing on Zip-Code Arrays and Comparison with RT-PCR

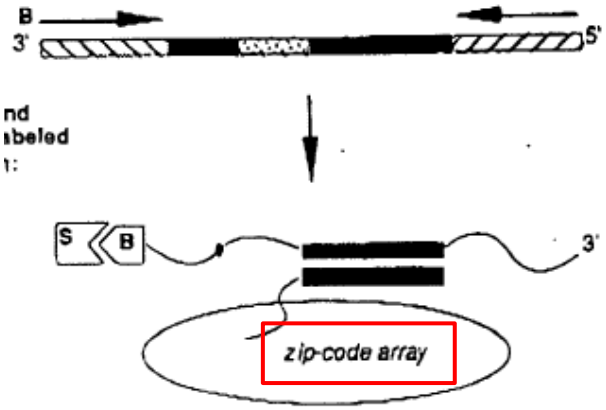
TABLE 2

IllumaCode ID	Decoder Sequence (5'-3')	Probe Sequence (5'-3')
1	TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTAGAAGCAGCGGCGAA
2	TTCGAAGCGCACGTCCCTTTTCAA	TTGAAAAGGGACGTGCGCTTCGAA
3	AACGCGTGGGGAATGGGACATCAA	TTGATGTCCCATTCCCCACGCGTT
4	CCGTCGCATACCGGCTACGATCAA	TTGATCGTAGCCGGTATGCGACGG
5	ATGGCCGTGCTGGGGACAAGTCAA	TTGACTTGTCCCCAGCACGGCCAT
6	TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA
7	CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG
8	CCGTTTGCGGTCGTCCTTGCTCAA	TTGAGCAAGGACGACCGCAAACGG
9	TTCGCTTTCGTGGCTGCACTTCAA	TTGAAGTGCAGCCACGAAAGCGAA
10	GTCCAACGCGCAACTCCGATTCAA	TTGAATCGGAGTTGCGCGTTGGAC
11	TTGCCGCACCGTCCGTCATCTCAA	TTGAGATGACGGACGGTGCGGCAA
12	CATCGTCCCTTTCGATGGGATCAA	TTGATCCCATCGAAAGGGACGATG
13	GCACGGGAGCTGACGACGTGTCAA	TTGACACGTCGTCAGCTCCCGTGC
14	AGACGCACCGCAACAGGCTGTCAA	TTGACAGCCTGTTGCGGTGCGTCT
15	CGTGTAGGGGTCCCGTGCTGTCAA	TTGACAGCACGGGACCCCTACACG
16	CATCGCTGCAAGTACCGCACTCAA	TTGAGTGCGGTACTTGCAGCGATG
17	GGCTGGTTCGGCCCGAAAGCTTAG	CTAAGCTTTCGGGCCGAACCAGCC
18	GTTCCCAGTGAAGCTGCGATCTGG	CCAGATCGCAGCTTCACTGGGAAC
19	TACTTGGCATGGAATCCCTTACGC	GCGTAAGGGATTCCATGCCAAGTA
20	ACTAGCATATTTCAGGGCACCGGC	GCCGGTGCCCTGAAATATGCTAGT

A. Array Images



5' to 3'; note the direction from right to left): an upstream universal 18 nt priming site (U), a unique 20 nt zip-code sequence (Zip) to target the oligo to a specific address on the array, a 20



In these Illumina patent applications, Illumina uses the terms "Zip-code Array", "Zip", "Decoder sequence", "IllumaCode", "Probe sequence", and "BeadArray capture probes" to describe Zirvi-Barany's (*exact*) 24-mer zip code sequences.

Zip code, cZip, Decoders, IllumaCode, IlluniCode, Capture Probe, Bead identifier, Target, Universal Tag Sequence

Gali Steinberg
Katie Stromsborg
Lynette Thomas
David Barker
Chanfeng Zhao
Illumina, Inc.,
9885 Towne Centre Drive,
San Diego, CA 92121

Strategies for Covalent Attachment of DNA to Beads

Received 15 June 2003;
accepted 1 September 2003
Published online 17 February 2004 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.20006

Y functional group
X 5' modification group

FIGURE 3 A general oligo immobilization scheme.

MATERIALS AND METHODS

Chemical Reagents

Silica beads in water (3 μm diameter, ~5×10^{10} beads/g) were obtained from Bangs Laboratories, Inc., Carmel, IN. Silanization reagents, 3-aminopropyltrimethoxysilane, 3-mercaptopropyltrimethoxysilane, and 3-glycidoxypropyltrimethoxysilane were obtained from United Chemical Technologies (UCT), Bristol, PA. Succinimidyl 4-hydrazinonicotinate acetone hydrazone, and 5′-aldehyde modified oligonucleotide 5′-(CHO) TTT GAA AAG CCT ACA CGA CGG CGA A-3′ (capture probe) were obtained from Solulink, San Diego, CA. All other oligonucleotides 5′-NH_2-TTT GAA AAG GGA CGT GCG CTT CGA A-3′ (capture probe), 5′-HS-TTT GAA AAG GGA CGT GCG CTT CGA A-3′ (capture probe), 5′-FAM-TTT CGC CGT CGT GTA GGC TTT TCA A-3′ (target), 5′-FAM-TTT CGA AGC GCA CGT CCC TTT TCA A-3′ (target), were obtained from Operon Technologies, Alameda, CA, or synthesized in house using the Oligator™ technology. All oligonucleotides were high performance liquid chromatography (HPLC) purified. All other reagents were obtained from NovaBiochem, Aldrich or Sigma. All solutions were prepared with OmniPur™ (sterile, nuclease free) water from Merck.

Complementary Oligonucleotide (Target) Hybridization. Hybridization of complementary oligo (target) to beads containing immobilized probes was carried out in solution and fluorescence-activated cell sorter (FACS) was used to measure hybridization intensity.

Hybridization was done in solution as follows: about 1 mg of probe-immobilized beads were suspended in 30 μL hybridization buffer (0.1*M* potassium phosphate, 1*M* sodium chloride, 0.1% Tween-20, and 5% ethanol, pH 7.6), followed by addition of 15–20 μL of ~1 m*M* target (in 0.05*M* borate buffer, pH 8.5) and then shaken for 2 h. The beads were washed 2–3 times with hybridization buffer to remove excess target oligonucleotides. For the release of target from the beads, the beads were incubated with 50 μL of 0.1*M* sodium hydroxide solution for 10 min.

In this Illumina publication describing chemistry for Infinium arrays, Illumina uses the terms “Capture Probe”, and “target” to describe Zirvi-Barany’s *exact* 24-mer zip code sequences (with an extra T on the 5’ end).

Zip code, cZip, Decoders, IllumaCode, IllumiCode, Capture Probe, Target, Address 1, Address 2, Universal Tag Sequence

Bioconjugate Chem. 2006, *17*, 841–848

Synthetic Modification of Silica Beads That Allows for Sequential Attachment of Two Different Oligonucleotides

Gali Steinberg-Tatman, Michael Huynh, David Barker, and Chanfeng Zhao*

Illumina Inc., 9885 Towne Centre Drive, San Diego, California 92121. Received January 19, 2006; Revised Manuscript Received February 22, 2006

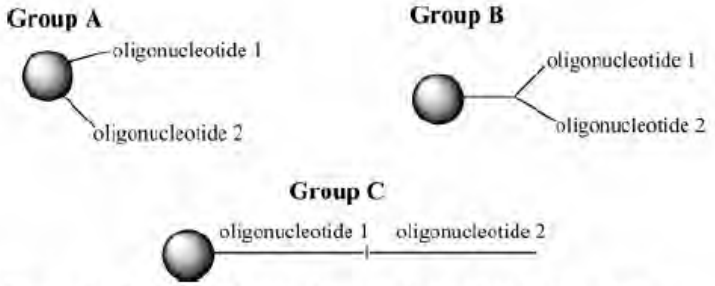


Figure 1. Potential schemes for two oligonucleotide attachment.

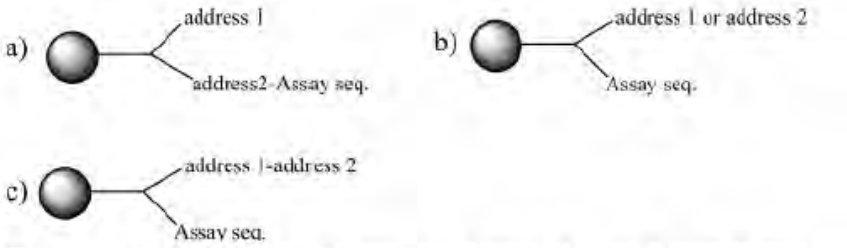


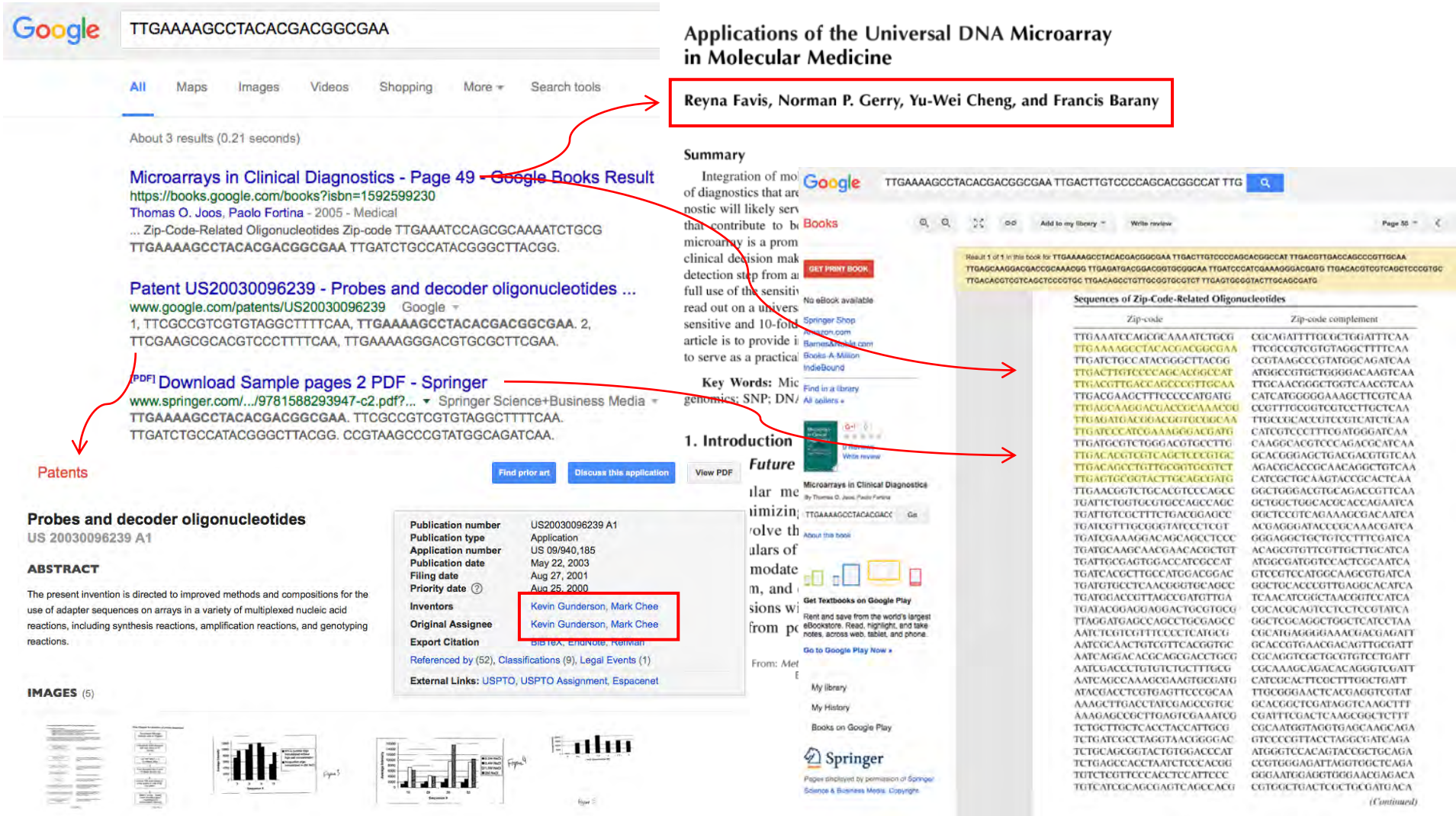
Figure 3. Detailed potential encoding schemes of group B from Figure 1.

Table 1. Hybridization Intensities of First and Second Sequences Resulting from the Different Activation/Attachment Chemistries for the Second Sequence

1st Sequence: S9 5′ (NH_2) TTTGATGTCCCATTCCCCACGCGTT					
2nd Sequence: S13 5′ (NH_2 or CHO) TTTGATCGTAGCCGGTATGCGACGG					
immobilization chemistries for 2nd attachment	5A	5B	5C	5D	5E
hybridization intensity of the 1st sequence before the 2nd sequence was immobilized	4978	5100	5154	5120	4978
hybridization intensity of 1st sequence after the 2nd sequence was immobilized	3925	2640	2729	2852	2670
hybridization intensity of the 2nd sequence	808	205	457	430	2950
1st Sequence: S13 5′ (NH_2) TTTGATCGTAGCCGGTATGCGACGG					
2nd Sequence: S9 5′ (NH_2 or CHO) TTTGATGTCCCATTCCCCACGCGTT					
immobilization chemistries for 2nd attachment	5A	5B	5C	5D	5E
hybridization intensity of the 1st sequence before the 2nd sequence was immobilized	3095	2872	2950	2930	3093
hybridization intensity of 1st sequence after the 2nd sequence was immobilized	2488	1500	1580	1550	1650
hybridization intensity of the 2nd sequence	1400	108	800	820	4800

In this Illumina publication describing chemistry for Infinium arrays, Illumina uses the terms “Capture Probe”, “target”, “Address 1” and “Address 2” to describe Zrivi-Barany’s *exact* 24-mer zip code sequences (with an extra T on the 5’ end).

A closer look at one of the 16 zip codes: TTGAAAAGCCTACACGACGGCGAA



Chances of 9 new "Illumacode" zip code sequences matching exactly to Zirvi-Barany's earlier work?
One in 10^{130} or a Googol (1×10^{100}) times the total number of stars in the Universe (1×10^{29})!

The 16 zip codes, plagiarized from Zirvi-Barany's unpublished document in the identical order:

The 16 zip codes in original unpublished 465 order

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	1	3	8	77.6	TTGAAATCCAGCGCAAAATCTGCG		1	4	31	21	4	29		CGCAGATTTTGCGCTGGATTTCAA		75.6
2	2	5	15	77.2	TTGAAAAGCCTACACGACGGCGAA		1	6	26	30	35	16		TTCGCCGTCGTGTAGGCTTTTCAA	1	77.3
3	97	6	18	79.1	TTGAAAAGGGACGTGCGCTTCGAA		1	6	34	33	17	16		TTCGAAGCCGCACGTCCCTTTTCAA	2	77.7
4	3	7	22	76.5	TTGATCTGCCATACGGGCTTACGG		1	8	15	35	17	35		CCGTAAGCCCGTATGGCAGATCAA		76.5
5	98	8	25	79.5	TTGATCTGCACGCGTTGTCGCGTA		1	8	30	12	33	18		TACCGCACAACGCGTCGACGTCAA		79.1
6	99	9	29	80.3	TTGATGTCCCATTCCCCACGCGTT		1	9	15	28	30	12		AACGCCTGGGGAATGGGACATCAA	3	78.9
7	66	12	42	81.5	TTGATCGTGGACACGGCACGCTCA		1	10	34	35	30	13		TGAGCGTCCGTGTCCACGATCAA		80.1
8	100	13	43	79.2	TTGATCGTAGGCGGTATGCGACGG		1	10	36	18	29	35		CCGTCGCATACCGGCTACGATCAA	4	77.8
9	4	15	45	80.6	TTGACTTGTCCCCAGCACGGCCAT		1	11	28	31	35	15		ATGGCCGTGCTGGGGACAAGTCAA	5	80.4
10	5	17	56	80.4	TTGACGTTGACCAGCCCGTTGCAA		1	12	32	36	12	21		TTGCAACGGGCTGGTCAACGTCAA	6	79.5
11	101	18	64	79.7	TTGACCATGCTTCGAAACGGCGAA		1	15	17	16	35	16		TTCGCCGTTTCGAAGCATGGTCAA		77.8
12	6	20	72	76.4	TTGACGAAGCTTTCCCCCATGATG		1	16	17	28	15	27		CATCATGGGGAAAGCTTCGTCAA		75.6
13	102	21	75	77.4	TTGACGAAATCGGCAACCTATGCG		1	16	24	21	26	29		CGCATAGGTTGCCGATTTCGTCAA	7	76.0
14	7	24	90	81.2	TTGAGCAAGGACGACCGCAAACGG		1	21	34	32	21	35		CCGTTTGCGGTCGTCCTTGCTCAA	8	79.1
15	87	26	94	78.2	TTGAAGTGCAGCCACGAAAGCGAA		1	22	31	30	6	16		TTCGCTTTCGTGGCTGCACTTCAA	9	78.0
16	68	29	112	79.3	TTGAATCGGAGTTGCGCGTTGGAC		1	24	20	29	12	34		GTCCAACGCGCAACTCCGATTCAA	10	77.7
17	8	30	126	82.4	TTGAGATGACGGACGGTGCGGCAA		1	27	35	35	29	21		TTGCCGCACCGTCCGTCATCTCAA	11	80.2
18	103	32	12925	75.7	TTGATCCCGGTATTAGAGCCCACG		1	28	18	3	36	30		CGTGGGCTCTAATACCGGGATCAA		75.9
19	9	33	140	76.3	TTGATCCCATCGAAAGGGACGATG		1	28	24	6	34	27		CATCGTCCCTTTCGATGGGATCAA	12	74.9
20	10	36	150	80.1	TTGATGCGTCTGGGACGTGCCTTG		1	29	8	34	33	11		CAAGGCACGTCCCAGACGCATCAA		78.8
21	104	42	161	80.7	TTGATGCGTGCGCTCAATCGAGGA		1	29	29	13	24	25		TCCTCGATTGAGCGCACGCATCAA		78.8
22	105	43	165	80.1	TTGATGCGAGCCGATGCCATCTTG		1	29	36	27	15	11		CAAGATGGCATCGGCTCGCATCAA		77.6
23	11	44	166	80.3	TTGACACGTCGTCAGCTCCCGTGC		1	30	10	31	28	33		GCACGGGAGCTGACGACGTGTCAA	13	79.9
24	12	48	180	80	TTGACAGCCTGTTGCGGTGCGTCT		1	31	14	29	33	19		AGACGCACCGCAACAGGCTGTCAA	14	80.6
25	106	49	186	79.1	TTGACAGCACGGGACCCCTACACG		1	31	35	32	26	30		CGTGTAGGGGTCCCGTGCTGTCAA	15	79.4
26	13	52	202	76.5	TTGAGTGCGGTACTTGCAGCGATG		1	33	18	11	31	27		CATCGCTGCAAGTACCGCACTCAA	16	77.3
27	107	53	208	80.3	TTGAGTGCTCCCGCAACGTTGTGC		1	33	28	21	12	33		GCACAACGTTGCGGGAGCACTCAA		79.6
28	14	55	222	81.2	TTGAACGGTCTGCACGTCCCAGCC		1	35	8	30	28	36		GGCTGGGACGTGCAGACCGTTCAA		80.2
29	108	58	230	76.8	TTGAACGGCAAACCTCTTGCTTG		1	35	21	23	11	11		CAAGCAAGAGGTTTGCCCGTTCAA		76.6
30	109	59	232	78.7	TTGAACGGGATGGCAAGGACCTCA		1	35	27	21	34	13		TGAGGTCCTTGCCATCCCGTTCAA		78.3

9 of the 16 zip codes in final 4,633 patent order

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	1	3	8	77.6	TTGAAATCCAGCGCAAAATCTGCG		1	4	31	21	4	29		CGCAGATTTTGCGCTGGATTTCAA		75.6
2	2	5	15	77.2	TTGAAAAGCCTACACGACGGCGAA		1	6	26	30	35	16		TTCGCCGTCGTGTAGGCTTTTCAA	1	77.3
3	3	7	22	76.5	TTGATCTGCCATACGGGCTTACGG		1	8	15	35	17	35		CCGTAAGCCCGTATGGCAGATCAA		76.5
4	4	15	45	80.6	TTGACTTGTCCCCAGCACGGCCAT		1	11	28	31	35	15		ATGGCCGTGCTGGGGACAAGTCAA	5	80.4
5	5	17	56	80.4	TTGACGTTGACCAGCCCGTTGCAA		1	12	32	36	12	21		TTGCAACGGGCTGGTCAACGTCAA	6	79.5
6	6	20	72	76.4	TTGACGAAGCTTTCCCCCATGATG		1	16	17	28	15	27		CATCATGGGGAAAGCTTCGTCAA		75.6
7	7	24	90	81.2	TTGAGCAAGGACGACCGCAAACGG		1	21	34	32	21	35		CCGTTTGCGGTCGTCCTTGCTCAA	8	79.1
8	8	30	126	82.4	TTGAGATGACGGACGGTGCGGCAA		1	27	35	35	29	21		TTGCCGCACCGTCCGTCATCTCAA	11	80.2
9	9	33	140	76.3	TTGATCCCATCGAAAGGGACGATG		1	28	24	6	34	27		CATCGTCCCTTTCGATGGGATCAA	12	74.9
10	10	36	150	80.1	TTGATGCGTCTGGGACGTGCCTTG		1	29	8	34	33	11		CAAGGCACGTCCCAGACGCATCAA		78.8
11	11	44	166	80.3	TTGACACGTCGTCAGCTCCCGTGC		1	30	10	31	28	33		GCACGGGAGCTGACGACGTGTCAA	13	79.9
12	12	48	180	80	TTGACAGCCTGTTGCGGTGCGTCT		1	31	14	29	33	19		AGACGCACCGCAACAGGCTGTCAA	14	80.6
13	13	52	202	76.5	TTGAGTGCGGTACTTGCAGCGATG		1	33	18	11	31	27		CATCGCTGCAAGTACCGCACTCAA	16	77.3
14	14	55	222	81.2	TTGAACGGTCTGCACGTCCCAGCC		1	35	8	30	28	36		GGCTGGGACGTGCAGACCGTTCAA		80.2
15	15	67	269	82.2	TGATTCTGGTGCGTGCCAGCCAGC		2	8	33	33	31	31		GCTGGCTGGCACGCACCAGAATCA		80.9
16	16	69	274	76.8	TGATTGTCGCTTTCTGACGGAGCC		2	9	17	8	35	36		GGCTCCGTCAGAAAGCGACAATCA		77.0
17	17	73	294	77.9	TGATCGTTTGCGGGTATCCCTCGT		2	12	29	18	28	10		ACGAGGGATACCCGCAAACGATCA		77.8
18	18	75	310	77.9	TGATCGAAAGGACAGCAGCCTCCC		2	16	25	31	36	28		GGGAGGCTGCTGTCCTTTCGATCA		78.0
19	19	79	328	76.9	TGATGCAAGCAAGAACAGCGTGT		2	21	21	16	30	14		ACAGCGTGTTCGTTGCTTGCATCA		78.2
20	20	92	372	79.5	TGATTGCGAGTGGACCATCGCCAT		2	29	22	32	24	15		ATGGCGATGGTCCACTCGCAATCA		78.7
21	21	94	379	80	TGATCACGCTTGCCATGGACGGAC		2	30	11	15	34	34		GTCCGTCCATGGCAAGCGTGATCA		78.7
22	22	104	418	82	TGATGTGCCTCAACGGGTGCAGCC		2	33	13	35	33	36		GGCTGCACCCGTTGAGGCACATCA		80.7
23	23	111	431	75.6	TGATGGACCGTTAGCCGATGTTGA		2	34	12	36	27	1		TCAACATCGGCTAACGGTCCATCA		76.4
24	24	118	444	79.5	TGATACGGAGGAGGACTGCGTGCG		2	35	25	34	29	29		CGCACGCAGTCCTCCTCCGTATCA		79.0
25	25	129	555	79.9	TTAGGATGAGCCAGCCTGCGAGCC		3	27	36	36	29	36		GGCTCGCAGGCTGGCTCATCCTAA		79.7
26	26	151	690	77.7	AATCTCGTCGTTTCCCCTCATGCG		4	10	12	28	13	29		CGCATGAGGGGAAACGACGAGATT		76.8
27	27	166	805	77.4	AATCGCAACTGTCGTTCACGGTGC		4	21	14	12	30	33		GCACCGTGAACGACAGTTGCGATT		77.9
28	28	167	842	81.4	AATCAGGACACGCAGCGACCTGCG		4	25	30	31	32	29		CGCAGGTCGCTGCGTGTCCTGATT		80.3
29	29	183	926	76.7	AATCGACCCTGTGTCTGCTTTGCG		4	32	14	19	17	29		CGCAAAGCAGACACAGGGTCGATT		77.4
30	30	197	983	76.9	AATCAGCCAAAGCGAAGTGCGATG		4	36	6	16	33	27		CATCGCACTTCGCTTTGGCTGATT		76.6

The 16 zip codes listed in Illumina's US20030096239 are in the same order as the original unpublished 465

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	1		TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTACACGACGGCGAA	77.3	1		5	15	77.2	TTGAAAAGCCTACACGACGGCGAA		1	6	26	30	35	16		TTCGCCGTCGTGTAGGCTTTTCAA
2	2		TTCGAAGCGCACGTCCCTTTTCAA	TTGAAAAGGGACGTGCGCTTCGAA	77.7	2		6	18	79.1	TTGAAAAGGGACGTGCGCTTCGAA		1	6	34	33	17	16		TTCGAAGCGCACGTCCCTTTTCAA
3	3		AACGCGTGGGGAATGGGACATCAA	TTGATGTCCCATTCCCCACGCGTT	78.9	3		9	29	80.3	TTGATGTCCCATTCCCCACGCGTT		1	9	15	28	30	12		AACGCGTGGGGAATGGGACATCAA
4	4		CCGTCGCATACCGGCTACGATCAA	TTGATCGTAGCCGGTATGCGACGG	77.8	4		13	43	79.2	TTGATCGTAGCCGGTATGCGACGG		1	10	36	18	29	35		CCGTCGCATACCGGCTACGATCAA
5	5		ATGGCCGTGCTGGGGACAAGTCAA	TTGACTTGTCCCCAGCACGGCCAT	80.4	5		15	45	80.6	TTGACTTGTCCCCAGCACGGCCAT		1	11	28	31	35	15		ATGGCCGTGCTGGGGACAAGTCAA
6	6		TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA	79.5	6		17	56	80.4	TTGACGTTGACCAGCCCGTTGCAA		1	12	32	36	12	21		TTGCAACGGGCTGGTCAACGTCAA
7	7		CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG	76.0	7		21	75	77.4	TTGACGAAATCGGCAACCTATGCG		1	16	24	21	26	29		CGCATAGGTTGCCGATTTCGTCAA
8	8		CCGTTTGCGGTCGTCCTTGCTCAA	TTGAGCAAGGACGACCGCAAACGG	79.1	8		24	90	81.2	TTGAGCAAGGACGACCGCAAACGG		1	21	34	32	21	35		CCGTTTGCGGTCGTCCTTGCTCAA
9	9		TTCGCTTTCGTGGCTGCACTTCAA	TTGAAGTGCAGCCACGAAAGCGAA	78.0	9		26	94	78.2	TTGAAGTGCAGCCACGAAAGCGAA		1	22	31	30	6	16		TTCGCTTTCGTGGCTGCACTTCAA
10	10		GTCCAACGCGCAACTCCGATTCAA	TTGAATCGGAGTTGCGCGTTGGAC	77.7	10		29	112	79.3	TTGAATCGGAGTTGCGCGTTGGAC		1	24	20	29	12	34		GTCCAACGCGCAACTCCGATTCAA
11	11		TTGCCGCACCGTCCGTCATCTCAA	TTGAGATGACGGACGGTGCGGCAA	80.2	11		30	126	82.4	TTGAGATGACGGACGGTGCGGCAA		1	27	35	35	29	21		TTGCCGCACCGTCCGTCATCTCAA
12	12		CATCGTCCCTTTCGATGGGATCAA	TTGATCCCATCGAAAGGGACGATG	74.9	12		33	140	76.3	TTGATCCCATCGAAAGGGACGATG		1	28	24	6	34	27		CATCGTCCCTTTCGATGGGATCAA
13	13		GCACGGGAGCTGACGACGTGTCAA	TTGACACGTCGTCAGCTCCCGTGC	79.9	13		44	166	80.3	TTGACACGTCGTCAGCTCCCGTGC		1	30	10	31	28	33		GCACGGGAGCTGACGACGTGTCAA
14	14		AGACGCACCGCAACAGGCTGTCAA	TTGACAGCCTGTTGCGGTGCGTCT	80.6	14		48	180	80	TTGACAGCCTGTTGCGGTGCGTCT		1	31	14	29	33	19		AGACGCACCGCAACAGGCTGTCAA
15	15		CGTGTAGGGGTCCCGTGCTGTCAA	TTGACAGCACGGGACCCCTACACG	79.4	15		49	186	79.1	TTGACAGCACGGGACCCCTACACG		1	31	35	32	26	30		CGTGTAGGGGTCCCGTGCTGTCAA
16	16		CATCGCTGCAAGTACCGCACTCAA	TTGAGTGCGGTACTTGCAGCGATG	77.3	16		52	202	76.5	TTGAGTGCGGTACTTGCAGCGATG		1	33	18	11	31	27		CATCGCTGCAAGTACCGCACTCAA
17	17		GGCTGGTTCGGCCCGAAAGCTTAG	CTAAGCTTTCGGGCCGAACCAGCC	78.8															
18	18		GTTCCCAGTGAAGCTGCGATCTGG	CCAGATCGCAGCTTCACTGGGAAC	76.9															
19	19		TACTTGGCATGGAATCCCTTACGC	GCGTAAGGGATTCCATGCCAAGTA	75.2															
20	20		ACTAGCATATTTCAGGGCACCGGC	GCCGGTGCCCTGAAATATGCTAGT	76.9						Tm values calculated for first 110 zip-codes using OligoAnalyzer 3.1 program									
21	21		GAACGGTCAATGAACCCGCTGTGA	TCACAGCGGGTTCATTGACCGTTC	77.2						under standard hybridization conditions									
22	22		GCGGCCTTGGTTCAATATGAATCG	CGATTCATATTGAACCAAGGCCGC	74.6															
23	23		GATCGTTAGAGGGACCTTGCCCGA	TCGGGCAAGGTCCCTCTAACGATC	77.2						Oligo Concentration:	0.25 uM								
24	24		TGGACCTAGTCCGGCAGTGACGAA	TTCGTCACTGCCGGACTAGGTCCA	78.7						Na+ Concentration:	1000 mM								
25	25		ATAAACTACCCAGGACGGGCGGAA	TTCCGCCCGTCCTGGGTAGTTTAT	77.6						Mg++ Concentration:	0 mM								

Chances of first 4 bases of 16 zip code sequences matching TTGA, which is our first tetramer?
One in 3 x 10^{38} or more than a billion times the total number of stars in the Universe! (1 x 10^{29})

An unexplained coincidence?

Barany *et al.*, filed, April 14, 2000

(12) **United States Patent**
Barany et al.

(10) **Patent No.: US 7,455,965 B2**
(45) **Date of Patent: Nov. 25, 2008**

(54) **METHOD OF DESIGNING ADDRESSABLE ARRAY FOR DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING LIGASE DETECTION REACTION**

(75) Inventors: **Francis Barany**, New York, NY (US); **Monib Zirvi**, Willingboro, NJ (US); **Norman P. Gerry**, Boston, MA (US); **Reyna Favis**, Iselin, NJ (US); **Richard Kliman**, Iselin, NJ (US)

FOREIGN PATENT DOCUMENTS

EP 0 357 011 8/1989

(Continued)

OTHER PUBLICATIONS

"Nucleic Acid Hybridization—General Aspects," in *Nonradioactive In Situ Hybridization Application Manual*, Indianapolis, Indiana:

Illumina, filed Aug. 25, 2000 – Never Issues

(19) **United States**
(12) **Patent Application Publication**
Gunderson et al.

(10) **Pub. No.: US 2003/0096239 A1**
(43) **Pub. Date: May 22, 2003**

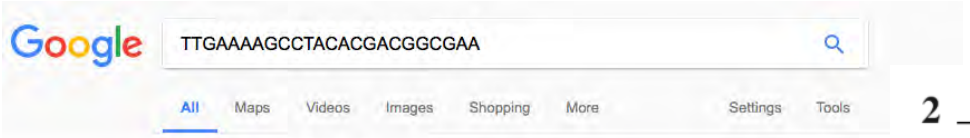
(54) **PROBES AND DECODER OLIGONUCLEOTIDES**

(76) Inventors: **Kevin Gunderson**, Encinitas, CA (US); **Mark Chee**, Del Mar, CA (US)

Related U.S. Application Data

(60) Provisional application No. 60/227,948, filed on Aug. 25, 2000. Provisional application No. 60/228,854, filed on Aug. 29, 2000.

Google search of: TTGAAAAGCCTACACGACGGCGAA



About 5 results (0.45 seconds)

[PDF] Download Sample pages 2 PDF - Springer
www.springer.com/cda/content/document/cda.../97815882939 47-c2.pdf?SGWID... ▾
TTGAAAAGCCTACACGACGGCGAA. TTCGCCGTCGTGTAGGCTTTTCAA.
TTGATCTGCCATACGGGCTTACGG. CCGTAAGCCCGTATGGCAGATCAA.

US20030096239 - Google
www.google.com/patents/US20030096239
1, TTCGCCGTCGTGTAGGCTTTTCAA, **TTGAAAAGCCTACACGACGGCGAA**. 2,
TTCGAAGCGCACGTCCCTTTTCAA, TTGAAAAGGGACGTGCGCTTCGAA.

Notes - Protein Microarrays - ALPF Medical Research
www.alpfmedical.info › Protein Microarrays
We recommend the software program Oligo for LDR primer design. This program is also useful in designing PCR and multiplex PCR primers. This program ...

Microarrays in Clinical Diagnostics - Page 49 - Google Books Result
https://books.google.com/books?isbn=1592599230
Thomas O. Joos, Paolo Fortina - 2005 - Medical
... Zip-Code-Related Oligonucleotides Zip-code TTGAAATCCAGCGCAAAATCTGCG
TTGAAAAGCCTACACGACGGCGAA TTGATCTGCCATACGGGCTTACGG.

Full text of "USPTO Patents Application 09940185" - Internet Archive
https://archive.org/stream/gov.uspto.../09940185-2002-01-31-00002-A.PE_djvu.txt ▾
... **TTGAAAAGCCTACACGACGGCGAA** 2 TTCGAAGCGCACGTCCCTTTTCAA
TTGAAAAGGGACGTGCGCTTCGAA 3 AACGCGTGGGGAATGGGACATCAA TTG ...

Full text of "USPTO Patents Application 09940185" - Internet Archive
https://archive.org/stream/gov.uspto.../09940185-2001-08-27-00001-LET._djvu.txt ▾
... 1 1 OOOOO 1 OO 1 O 1 AAOOO 1 1 1 1 **TTGAAAAGCCTACACGACGGCGAA** 9 Z.
TTPGAAGPGPAPGTPPPTTTTOAA TTGAAAAGGGACGTGCGCTTCGAA i i ...

Probes and decoder oligonucleotides
US 20030096239 A1

ABSTRACT

The present invention is directed to improved methods and compositions for the use of adapter sequences on arrays in a variety of multiplexed nucleic acid reactions, including synthesis reactions, amplification reactions, and genotyping reactions.

Publication number	US20030096239 A1
Publication type	Application
Application number	US 09/940,185
Publication date	May 22, 2003
Filing date	Aug 27, 2001
Priority date	Aug 25, 2000
Inventors	Kevin Gunderson, Mark Chee
Original Assignee	Kevin Gunderson, Mark Chee
Export Citation	BiBTeX, EndNote, RefMan

Referenced by (52), Classifications (9), Legal Events (1)

External Links: USPTO, USPTO Assignment, Espacenet

Seq. ID No.

Decoder Sequence (5'-3')

Probe Sequence (5-3')

1

TTCGCCGTCGTGTAGGCTTTTCAA

TTGAAAAGCCTACACGACGGCGAA

USPTO archives of all the Illumina filings and patent office correspondence may now be found relating to the plagiarized Zirvi-Barany zip codes

2

Applications of the Universal DNA Microarray in Molecular Medicine

Reyna Favis, Norman P. Gerry, Yu-Wei Cheng, and Francis Barany

Summary

Integration of molecular medicine into standard clinical practice will require the availability of diagnostics that are sensitive, rapid, and robust. The backbone technology underlying the diagnostic will likely serve double duty during clinical trials in order to first validate the biomarkers that contribute to both drug response and disease stratification. PCR/LDR/Universal DNA microarray is a promising technology to help drive the transition from the current paradigms of clinical decision making to the new era of personalized medicine. By uncoupling the mutation detection step from array hybridization, this technology becomes fully programmable. It exploits full use of the sensitivity that the ligase detection reaction can provide, while maintaining a rapid read out on a universal microarray. Thus, PCR/LDR/Universal DNA microarray is 50-fold more sensitive and 10-fold more rapid than conventional hybridization-only arrays. The intent of this

Table 1
Sequences of Zip-Code-Related Oligonucleotides

Zip-code	Zip-code complement
TTGAAATCCAGCGCAAAATCTGCG	CGCAGATTTTGCGCTGGATTTCAA
TTGAAAAGCCTACACGACGGCGAA	TTCGCCGTCGTGTAGGCTTTTCAA
TTGATCTGCCATACGGGCTTACGG	CCGTAAGCCCGTATGGCAGATCAA
TTGACTTGTCCCCAGCACGGCCAT	ATGGCCGTGCTGGGGACAAGTCAA
TTGACGTTGACCAGCCCGTTGCAA	TTGCAACGGGCTGGTCAACGTCAA
TTGACGAAGCTTTCCCCCATGATG	CATCATGGGGGAAAGCTTCGTCAA
TTGAGCAAGGACGACCGCAAACGG	CCGTTTGCGGTCGTCCTTGCTCAA
TTGAGATGACGGACGGTGCGGCAA	TTGCCGCACCGTCCGTCATCTCAA
TTGATCCCATCGAAAGGGACGATG	CATCGTCCCTTTCGATGGGATCAA
TTGATGCGTCTGGGACGTGCCTTG	CAAGGCACGTCCCAGACGCATCAA
TTGACACGTCGTCAGCTCCCGTGC	GCACGGGAGCTGACGACGTGTCAA
TTGACAGCCTGTTGCCGTGCGTCT	AGACGCACCGCAACAGGCTGTCAA
TTGAGTGCGGTACTTGCAGCGATG	CATCGCTGCAAGTACCGCACTCAA
TTGAACGGTCTGCACGTCCCAGCC	GGCTGGGACGTGCAGACCGTTCAA

Illumina tries to re-patent Zirvi-Barany's zip code sequences, renames them "IllumaCode ID", and later "Seq. ID No.".

Barany *et al.*, filed, April 14, 2000

(12) **United States Patent**
Barany et al.

(10) **Patent No.: US 7,455,965 B2**
(45) **Date of Patent: Nov. 25, 2008**

(54) **METHOD OF DESIGNING ADDRESSABLE ARRAY FOR DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING LIGASE DETECTION REACTION**

(75) Inventors: **Francis Barany**, New York, NY (US); **Monib Zirvi**, Willingboro, NJ (US); **Norman P. Gerry**, Boston, MA (US); **Reyna Favis**, Iselin, NJ (US); **Richard Kliman**, Iselin, NJ (US)

FOREIGN PATENT DOCUMENTS

EP 0 357 011 8/1989

(Continued)

OTHER PUBLICATIONS

"Nucleic Acid Hybridization—General Aspects," in *Nonradioactive In Situ Hybridization Application Manual*, Indianapolis, Indiana:

Illumina, filed Aug. 25, 2000 – Never Issues

(19) **United States**
(12) **Patent Application Publication**
Gunderson et al.

(10) **Pub. No.: US 2003/0096239 A1**
(43) **Pub. Date: May 22, 2003**

(54) **PROBES AND DECODER OLIGONUCLEOTIDES**

(76) Inventors: **Kevin Gunderson**, Encinitas, CA (US); **Mark Chee**, Del Mar, CA (US)

Related U.S. Application Data

(60) Provisional application No. 60/227,948, filed on Aug. 25, 2000. Provisional application No. 60/228,854, filed on Aug. 29, 2000.

SEQ ID NO:	4,633 ID#	HEX ID#	Tm	ZIPCODE (5'-3')	TETRAMER NUMBERS					
1	3	8	77.6	TTGAAAATCCAGCGCAAAATCTGCG	1	4	31	21	4	29
2	5	15	77.2	TTGAAAAGCCTACACGACGGCGAA	1	6	26	30	35	16
3	7	22	76.5	TTGATCTGCCATACGGGCTTACGG	1	8	15	35	17	35
4	15	45	80.6	TTGACTTGTCCCCAGCACGGCCAT	1	11	28	31	35	15
5	17	56	80.4	TTGACGTTGACCAGCCCGTTGCAA	1	12	32	36	12	21
6	20	72	76.4	TTGACGAAGCTTTCCCCCATGATG	1	16	17	28	15	27
7	24	90	81.2	TTGAGCAAGGACGACCGCAAACGG	1	21	34	32	21	35
8	30	126	82.4	TTGAGATGACGGACGGTGCGGCAA	1	27	35	35	29	21
9	33	140	76.3	TTGATCCCATCGAAAGGGACGATG	1	28	24	6	34	27
10	36	150	80.1	TTGATGCGTCTGGGACGTGCCTTG	1	29	8	34	33	11
11	44	166	80.3	TTGACACGTCGTCAGCTCCCGTGC	1	30	10	31	28	33
12	48	180	80	TTGACAGCCTGTTGCGGTGCGTCT	1	31	14	29	33	19
13	52	202	76.5	TTGAGTGCGGTACTTGCAGCGATG	1	33	18	11	31	27
14	55	222	81.2	TTGAACGGTCTGCACGTCCCAGCC	1	35	8	30	28	36
15	67	269	82.2	TGATTCTGGTGCGTGCCAGCCAGC	2	8	33	33	31	31
16	69	274	76.8	TGATTGTCGCTTTCTGACGGAGCC	2	9	17	8	35	36

TABLE 2

IllumaCode ID	Decoder Sequence (5'-3')	Probe Sequence (5'-3')
1	TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTACACGACGGCGAA
2	TTCGAAGCGCACGTCCCTTTTCAA	TTGAAAAGGGACGTGCGCTTCGAA
3	AACGCGTGGGGAATGGGACATCAA	TTGATGTCCCATTCCCCACGCGTT
4	CCGTCGCATACCGGCTACGATCAA	TTGATCGTAGCCGGTATGCGACGG
5	ATGGCCGTGCTGGGGACAAGTCAA	TTGACTTGTCCCCAGCACGGCCAT
6	TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA
7	CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG
8	CCGTTTGCGGTCGTCCTTGCTCAA	TTGAGCAAGGACGACCGCAAACGG
9	TTCGCTTTCGTGGCTGCACTTCAA	TTGAAGTGCAGCCACGAAAGCGAA
10	GTCCAACGCGCAACTCCGATTCAA	TTGAATCGGAGTTGCGCGTTGGAC
11	TTGCCGCACCGTCCGTCATCTCAA	TTGAGATGACGGACGGTGCGGCAA
12	CATCGTCCCTTTCGATGGGATCAA	TTGATCCCATCGAAAGGGACGATG
13	GCACGGGAGCTGACGACGTGTCAA	TTGACACGTCGTCAGCTCCCGTGC
14	AGACGCACCGCAACAGGCTGTCAA	TTGACAGCCTGTTGCGGTGCGTCT
15	CGTGTAGGGGTCCCGTGCTGTCAA	TTGACAGCACGGGACCCCTACACG
16	CATCGCTGCAAGTACCGCACTCAA	TTGAGTGCGGTACTTGCAGCGATG
17	GGCTGGTTCGGCCCGAAAGCTTAG	CTAAGCTTTCGGGCCGAACCAGCC
18	GTTCCCAGTGAAGCTGCGATCTGG	CCAGATCGCAGCTTCACTGGGAAC
19	TACTTGGCATGGAATCCCTTACGC	GCGTAAGGGATTCCATGCCAAGTA
20	ACTAGCATATTTCAGGGCACCGGC	GCCGGTGCCCTGAAATATGCTAGT

Illumina – enabled by PE Biosystems providing them Zirvi-Barany's unpublished trade secret sequences sometime in 1999-2000 – attempted to re-patent Zirvi-Barany's ***exact*** zip code sequences. This is part of a deliberate strategy by Illumina and PE Biosystems – as evidenced by the un-redacted "First amendment to the Joint development agreement" to systematically and fraudulently purloin the true inventors intellectual property.

Gunderson & Chee plagiarized Zirvi-Barany's zip codes, which were used to achieve Illumina bead decoding.

	A	B	C	D	E	F
1	1		TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTACACGACGGCGAA	77.3	1
2	2		TTCGAAGCGCACGTCCCTTTTCAA	TTGAAAAGGGACGTGCGCTTCGAA	77.7	2
3	3		AACGCGTGGGGAATGGGACATCAA	TTGATGTCCCATTCCCCACGCGTT	78.9	3
4	4		CCGTCGCATACCGGCTACGATCAA	TTGATCGTAGCCGGTATGCGACGG	77.8	4
5	5		ATGGCCGTGCTGGGGACAAGTCAA	TTGACTTGTCCCCAGCACGGCCAT	80.4	5
6	6		TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA	79.5	6
7	7		CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG	76.0	7
8	8		CCGTTTGCGGTCGTCCTTGCTCAA	TTGAGCAAGGACGACCGCAAACGG	79.1	8
9	9		TTCGCTTTCGTGGCTGCACTTCAA	TTGAAGTGCAGCCACGAAAGCGAA	78.0	9
10	10		GTCCAACGCGCAACTCCGATTCAA	TTGAATCGGAGTTGCGCGTTGGAC	77.7	10
11	11		TTGCCGCACCGTCCGTCATCTCAA	TTGAGATGACGGACGGTGCGGCAA	80.2	11
12	12		CATCGTCCCTTTCGATGGGATCAA	TTGATCCCATCGAAAGGGACGATG	74.9	12
13	13		GCACGGGAGCTGACGACGTGTCAA	TTGACACGTCGTCAGCTCCCGTGC	79.9	13
14	14		AGACGCACCGCAACAGGCTGTCAA	TTGACAGCCTGTTGCGGTGCGTCT	80.6	14
15	15		CGTGTAGGGGTCCCGTGCTGTCAA	TTGACAGCACGGGACCCCTACACG	79.4	15
16	16		CATCGCTGCAAGTACCGCACTCAA	TTGAGTGCGGTACTTGCAGCGATG	77.3	16
17	17		GGCTGGTTCGGCCCGAAAGCTTAG	CTAAGCTTTCGGGCCGAACCAGCC	78.8	
18	18		GTTCCCAGTGAAGCTGCGATCTGG	CCAGATCGCAGCTTCACTGGGAAC	76.9	
19	19		TACTTGGCATGGAATCCCTTACGC	GCGTAAGGGATTCCATGCCAAGTA	75.2	
20	20		ACTAGCATATTTCAGGGCACCGGC	GCCGGTGCCCTGAAATATGCTAGT	76.9	
21	21		GAACGGTCAATGAACCCGCTGTGA	TCACAGCGGGTTCATTGACCGTTC	77.2	
22	22		GCGGCCTTGGTTCAATATGAATCG	CGATTCATATTGAACCAAGGCCGC	74.6	
23	23		GATCGTTAGAGGGACCTTGCCCGA	TCGGGCAAGGTCCCTCTAACGATC	77.2	
24	24		TGGACCTAGTCCGGCAGTGACGAA	TTCGTCACTGCCGGACTAGGTCCA	78.7	
25	25		ATAAACTACCCAGGACGGGCGGAA	TTCCGCCCGTCCTGGGTAGTTTAT	77.6	
26	26		CATCGGTTCGCGCCAATCCAGATA	TATCTGGATTGGCGCGAACCGATG	76.9	
27	27		GTCGGGCATAGAGCCGACCACCCT	AGGGTGGTCGGCTCTATGCCCGAC	80.6	
28	28		CTTGGGTCATGATTCACCGTGCTA	TAGCACGGTGAATCATGACCCAAG	75.1	
29	29		TGCCTAACGTGCTAATCAGCAGCG	CGCTGCTGATTAGCACGTTAGGCA	77.2	
30	30		CGCATGTTGGAGCATATGCCCTGA	TCAGGGCATATGCTCCAACATGCG	77.3	
31	31		AGCCACTGCATCAGTGCTGTTCAA	TTGAACAGCACTGATGCAGTGGCT	77.6	
32	32		GGTTGTTTTGAGGCGTCCCACACT	AGTGTGGGACGCCTCAAAACAACC	77.5	
33	33		TCGACCAAGAGCAAGGGCGGACCA	TGGTCCGCCCTTGCTCTTGGTCGA	81.2	
34	34		GACATCGCTATTGCGCATGGATCA	TGATCCATGCGCAATAGCGATGTC	75.8	
35	35		GAAATACGAAGTCTGCGGGAGTCG	CGACTCCCGCAGACTTCGTATTTC	74.8	
36	36		TGTCATGAATGATTGATCGCGCGA	TCGCGCGATCAATCATTCATGACA	75.8	
37	37		ATATCGGGATTCGTTCCCGGTGAA	TTCACCGGGAACGAATCCCGATAT	76.1	
38	38		GCGAGCGTACCGAAGGGCCTAGAA	TTCTAGGCCCTTCGGTACGCTCGC	79.1	
39	39		TTACCGGCAGCGGACTTCCGAATT	AATTCGGAAGTCCGCTGCCGGTAA	78.5	
40	40		GTAATCGAGAGCTGCGCGCGGTCT	AGACGGCGCGCAGCTCTCGATTAC	79.7	
41	41		CCTGTTAGCGTAGGCGAGTCGATC	GATCGACTCGCGTACGCTAACAGG	76.0	
42	42		TAGCGGACCGGCAGAATGAGTTCC	GGAACTCATTCTGCCGGTCCGCTA	77.8	
43	43		GGTACATGCACTACGCGCACTCGG	CCGAGTGCGCGTAGTGCATGTACC	78.4	
44	44		AATTCATCTCGGACTCCCGCGGTA	TACCGCGGGAGTCCGAGATGAATT	77.4	
45	45		GCCAAATCTGGATTGGCAGGAATG	CATTCCTGCCAATCCAGATTTGGC	74.9	
46	46		[illegible]	[illegible]	[illegible]	

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		5	15	77.2	TTGAAAAGCCTACACGACGGCGAA		1	6	26	30	35	16		TTCGCCGTCGTGTAGGCTTTTCAA
2		6	18	79.1	TTGAAAAGGGACGTGCGCTTCGAA		1	6	34	33	17	16		TTCGAAGCGCACGTCCCTTTTCAA
3		9	29	80.3	TTGATGTCCCATTCCCCACGCGTT		1	9	15	28	30	12		AACGCGTGGGGAATGGGACATCAA
4		13	43	79.2	TTGATCGTAGCCGGTATGCGACGG		1	10	36	18	29	35		CCGTCGCATACCGGCTACGATCAA
5		15	45	80.6	TTGACTTGTCCCCAGCACGGCCAT		1	11	28	31	35	15		ATGGCCGTGCTGGGGACAAGTCAA
6		17	56	80.4	TTGACGTTGACCAGCCCGTTGCAA		1	12	32	36	12	21		TTGCAACGGGCTGGTCAACGTCAA
7		21	75	77.4	TTGACGAAATCGGCAACCTATGCG		1	16	24	21	26	29		CGCATAGGTTGCCGATTTCGTCAA
8		24	90	81.2	TTGAGCAAGGACGACCGCAAACGG		1	21	34	32	21	35		CCGTTTGCGGTCGTCCTTGCTCAA
9		26	94	78.2	TTGAAGTGCAGCCACGAAAGCGAA		1	22	31	30	6	16		TTCGCTTTCGTGGCTGCACTTCAA
10		29	112	79.3	TTGAATCGGAGTTGCGCGTTGGAC		1	24	20	29	12	34		GTCCAACGCGCAACTCCGATTCAA
11		30	126	82.4	TTGAGATGACGGACGGTGCGGCAA		1	27	35	35	29	21		TTGCCGCACCGTCCGTCATCTCAA
12		33	140	76.3	TTGATCCCATCGAAAGGGACGATG		1	28	24	6	34	27		CATCGTCCCTTTCGATGGGATCAA
13		44	166	80.3	TTGACACGTCGTCAGCTCCCGTGC		1	30	10	31	28	33		GCACGGGAGCTGACGACGTGTCAA
14		48	180	80	TTGACAGCCTGTTGCGGTGCGTCT		1	31	14	29	33	19		AGACGCACCGCAACAGGCTGTCAA
15		49	186	79.1	TTGACAGCACGGGACCCCTACACG		1	31	35	32	26	30		CGTGTAGGGGTCCCGTGCTGTCAA
16		52	202	76.5	TTGAGTGCGGTACTTGCAGCGATG		1	33	18	11	31	27		CATCGCTGCAAGTACCGCACTCAA

Tm values calculated for first 110 zip-codes using OligoAnalyzer 3.1 program
under standard hybridization conditions

Oligo Concentration:	0.25 uM
Na+ Concentration:	1000 mM
Mg++ Concentration:	0 mM

Illumina – enabled by PE Biosystems providing them Zirvi-Barany's unpublished trade secret sequences sometime in 1999-2000 – attempted to re-patent Zirvi-Barany's _exact_ zip code sequences. This is part of a deliberate strategy by Illumina and PE Biosystems – as evidenced by the un-redacted "First amendment to the Joint development agreement" to systematically and fraudulently purloin the true inventors intellectual property.

Zirvi-Barany describe properties of zip code addresses;

Illumina plagiarizes:

Zip code length of 24-mer.
"IllumaCodes" 24-mer Illumina Adapter sequences, "randomly picked by computer". In actuality, they seeded the program with 16 Zirvi-Barany zip codes.

Zirvi-Barany describe uniform hybridization. >85% of Tm values within the range of 75 – 80 C inclusive.
Illumina claims random 24-mers with Tm centered around 72° C and with a spread of 5 degrees which would be 67-77° C. In actuality, "IllumaCode" sequences also had Tm range of mostly 75-80° C inclusive, and almost all were 74-81° C, which is centered on the exact range as the Zirvi-Barany zip codes.

Barany Lab describe 25% or greater differences between two zip codes. For 24-mers, that equals probe-decoder complementarity score <= 18.
Illumina claims "Probe-Decoder complementarity Score < 14". In actuality; there are:
1,461 pairs with complementarity score of 14;
338 pairs with complementarity score of 15;
70 pairs with complementarity score of 16;
22 pairs with complementarity score of 17; and
1 pair with complementarity score of 18;
In other words 18/24 = 25% or greater differences - exactly matches Barany Lab teaching.

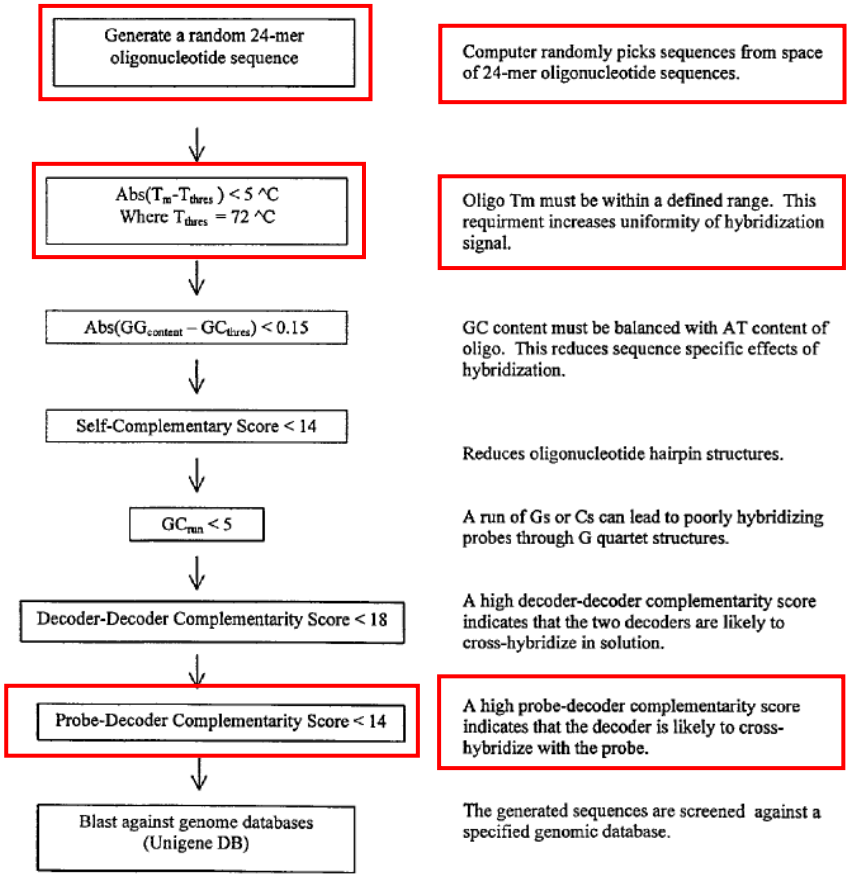
Illumina Application: Gunderson & Chee, Fig. 1

Description of algorithm to select "best" oligonucleotide adapter sequences.

Requirements for good sequences:

- Generates adequate hybridization signal intensity when employed in an experiment.
- Exhibits minimal cross-reactivity with other adapter sequences.
- Unique within the human genome sequence. This requirement can be extended to the genomic sequence of other organisms such as the fruit fly, the mouse, etc.

One method of generating sequences that meet the above requirements is to randomly generate sequences of given lengths and than pass these filters through a set of heuristic acceptance filters. In particular, the 24-mer Illumina Adapter sequences (IllumaCodes) were chosen as follows.



Illumina submits full application 09/940,185 on August 27, 2001:

Attorney Docket No.: A-69605-1/RMS/DCF [469249-00110]

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As the below named inventor, I hereby declare that:

My residence, post office address and citizenship is as stated below next to my name;

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

PROBES AND DECODER OLIGONUCLEOTIDES

the specification of which:
- ☐ is attached hereto.
- ☒ was filed on: August 27, 2001
  as Application No.: 09/940,185
  and was amended on: _____ (if applicable).

I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability as defined in 37 C.F.R. § 1.56.

Illumina provides **false statements** to the USPTO, claiming to have invented "**Probes and decoder oligonucleotides**" when in fact they plagiarized the first 16 zipcodes directly from an unpublished Zirvi-Barany document they obtained through Illumina's collaboration with PE-Biosystems. They did this to avoid paying the true inventors their rightful royalties.

Attorney Docket No.: A-69605-1/RMS/DCF [469249-00110]

Signature: Kevin Gunderson Date: 2/4/2003

Full Name of First Inventor: GUNDERSON (Family Name) Kevin (First Given Name) L. (Second Given Name)

Citizenship: United States of America

Residence: 1543 Juniper Hill Drive, Encinitas, California 92024

Post Office Address: Same as above

Signature: Mark Chee Date: 2/4/2003

Full Name of Second Inventor: CHEE (Family Name) Mark (First Given Name) S. (Second Given Name)

Citizenship: Australia

Residence: 155 15th Street, No. 22, Del Mar, California 92014

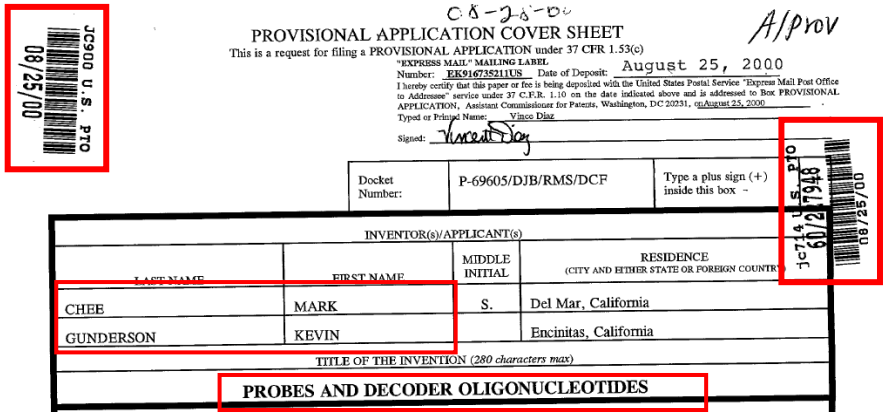
Post Office Address: Same as above

Kevin Gunderson and Mark Chee sign their name on February 4, 2003 to declare:

"I believe that I am the original, first, and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled: **Probes and decoder oligonucleotides**."

"I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability as defined in 37 C.F.R. § 1.56."

Illumina submits provisional application 60/227,948 (M. Chee and K. Gunderson) on August 25, 2000:



PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

"EXPRESS MAIL" MAILING LABEL Number: EK916735211US Date of Deposit: August 25, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box PROVISIONAL APPLICATION, Assistant Commissioner for Patents, Washington, DC 20231, on August 25, 2000

Typed or Printed Name: Vince Diaz

Signed:

Docket Number:	P-69605/DJB/RMS/DCF	Type a plus sign (+) inside this box →

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
CHEE	MARK	S.	Del Mar, California
GUNDERSON	KEVIN		Encinitas, California

TITLE OF THE INVENTION (280 characters max)

PROBES AND DECODER OLIGONUCLEOTIDES

TABLE 1

IllumaCode ID	Decoder (5'-3')
17	GGCTGGTTCGGCCCGAAAGCTTAG
18	GTTCCCAGTGAAGCTGCGATCTGG
19	TACTTGGCATGGAATCCCTTACGC
20	ACTAGCATATTTCAGGGCACCGGC

TABLE 2

IllumaCode ID	Decoder Sequence (5'-3')	Probe Sequence (5'-3')
1	TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTACACGACGGCGAA
2	TTCGAAGCGCACGTCCCTTTTCAA	TTGAAAAGGGACGTGCGCTTCGAA
3	AACGCGTGGGGAATGGGACATCAA	TTGATGTCCCATTCCCCACGCGTT
4	CCGTCGCATACCGGCTACGATCAA	TTGATCGTAGCCGGTATGCGACGG
5	ATGGCCGTGCTGGGGACAAGTCAA	TTGACTTGTCCCCAGCACGGCCAT
6	TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA
7	CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG
8	CCGTTTGCGGTCGTCCTTGCTCAA	TTGAGCAAGGACGACCGCAAACGG
9	TTCGCTTTCGTGGCTGCACTTCAA	TTGAAGTGCAGCCACGAAAGCGAA
10	GTCCAACGCGCAACTCCGATTCAA	TTGAATCGGAGTTGCGCGTTGGAC
11	TTGCCGCACCGTCCGTCATCTCAA	TTGAGATGACGGACGGTGCGGCAA
12	CATCGTCCCTTTCGATGGGATCAA	TTGATCCCATCGAAAGGGACGATG
13	GCACGGGAGCTGACGACGTGTCAA	TTGACACGTCGTCAGCTCCCGTGC
14	AGACGCACCGCAACAGGCTGTCAA	TTGACAGCCTGTTGCGGTGCGTCT
15	CGTGTAGGGGTCCCGTGCTGTCAA	TTGACAGCACGGGACCCCTACACG
16	CATCGCTGCAAGTACCGCACTCAA	TTGAGTGCGGTACTTGCAGCGATG
17	GGCTGGTTCGGCCCGAAAGCTTAG	CTAAGCTTTCGGGCCGAACCAGCC
18	GTTCCCAGTGAAGCTGCGATCTGG	CCAGATCGCAGCTTCACTGGGAAC
19	TACTTGGCATGGAATCCCTTACGC	GCGTAAGGGATTCCATGCCAAGTA
20	ACTAGCATATTTCAGGGCACCGGC	GCCGGTGCCCTGAAATATGCTAGT

TABLE 3

IllumaCode ID	Decoder Sequence (5'-3')	Probe Sequence (5'-3')
17	TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTACACGACGGCGAA
18	GTTCCCAGTGAAGCTGCGATCTGG	CCAGATCGCAGCTTCACTGGGAAC
19	TACTTGGCATGGAATCCCTTACGC	GCGTAAGGGATTCCATGCCAAGTA
20	ACTAGCATATTTCAGGGCACCGGC	GCCGGTGCCCTGAAATATGCTAGT
301	CCCGATATTCGCCCGGCCTATTCG	CGAATAGGCCGGGCGAATATCGGG
302	CGAGAAGATGCCTCACGCAACCAA	TTGGTTGCGTGAGGCATCTTCTCG
303	AACCTTGACCCGTGGATGACGCTA	TAGCGTCATCCACGGGTCAAGGTT
6	TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA
7	CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG
306	GCTTCCGGATGAACGGGATGGTTG	CAACCATCCCGTTCATCCGGAAGC
307	CCCTCCATGTTCTTCGAACGGTTT	AAACCGTTCGAAGAACATGGAGGG
308	TTGATGGGCGGCAATGCTCTTGCT	AGCAAGAGCATTGCCGCCCATCAA

TABLE 4

IllumaCode ID	Decoder Sequence (5'-3') + 5' T	Probe Sequence (5'-3') + 5' T
17	TTTCGCCGTCGTGTAGGCTTTTCAA	TTTGAAAAGCCTACACGACGGCGAA
18	TGTTCCCAGTGAAGCTGCGATCTGG	TCCAGATCGCAGCTTCACTGGGAAC
19	TTACTTGGCATGGAATCCCTTACGC	TGCGTAAGGGATTCCATGCCAAGTA
20	TACTAGCATATTTCAGGGCACCGGC	TGCCGGTGCCCTGAAATATGCTAGT
301	TCCCGATATTCGCCCGGCCTATTCG	TCGAATAGGCCGGGCGAATATCGGG
302	TCGAGAAGATGCCTCACGCAACCAA	TTTGGTTGCGTGAGGCATCTTCTCG
303	TAACCTTGACCCGTGGATGACGCTA	TTAGCGTCATCCACGGGTCAAGGTT
6	TTTGCAACGGGCTGGTCAACGTCAA	TTTGACGTTGACCAGCCCGTTGCAA
7	TCGCATAGGTTGCCGATTTCGTCAA	TTTGACGAAATCGGCAACCTATGCG
306	TGCTTCCGGATGAACGGGATGGTTG	TCAACCATCCCGTTCATCCGGAAGC
307	TCCCTCCATGTTCTTCGAACGGTTT	TAAACCGTTCGAAGAACATGGAGGG
308	TTTGATGGGCGGCAATGCTCTTGCT	TAGCAAGAGCATTGCCGCCCATCAA

Illumina attempt to re-patent Zirvi-Barany's *exact* zip code sequences; accidentally leaves one with wrong ID in Table 1, includes all in Table 2, and has them in a bizarre order in tables 3 & 4.

Illumina submits provisional application 60/227,948 (M. Chee and K. Gunderson) on August 25, 2000:

08/25/00 JC900 U.S. PTO

C.A-28-00 A/Prov

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

"EXPRESS MAIL" MAILING LABEL Number: EK916735211US Date of Deposit: August 25, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box PROVISIONAL APPLICATION, Assistant Commissioner for Patents, Washington, DC 20231. on August 25, 2000

Typed or Printed Name: Vince Diaz

Signed:

Docket Number:	P-69605/DJB/RMS/DCF	Type a plus sign (+) inside this box →

JC714 U.S. PTO 60/227948 08/25/00

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
CHEE	MARK	S.	Del Mar, California
GUNDERSON	KEVIN		Encinitas, California

TITLE OF THE INVENTION (280 characters max)

PROBES AND DECODER OLIGONUCLEOTIDES

TABLE 2

IllumaCode ID	Decoder Sequence (5'-3')	Probe Sequence (5'-3')
1	TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTACACGACGGCGAA
2	TTCGAAGCGCACGTCCCTTTTCAA	TTGAAAAGGGACGTGCGCTTCGAA
3	AACGCGTGGGGAATGGGACATCAA	TTGATGTCCCATTCCCCACGCGTT
4	CCGTCGCATACCGGCTACGATCAA	TTGATCGTAGCCGGTATGCGACGG
5	ATGGCCGTGCTGGGGACAAGTCAA	TTGACTTGTCCCCAGCACGGCCAT
6	TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA
7	CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG
8	CCGTTTGCGGTCGTCCTTGCTCAA	TTGAGCAAGGACGACCGCAAACGG
9	TTCGCTTTCGTGGCTGCACTTCAA	TTGAAGTGCAGCCACGAAAGCGAA
10	GTCCAACGCGCAACTCCGATTCAA	TTGAATCGGAGTTGCGCGTTGGAC
11	TTGCCGCACCGTCCGTCATCTCAA	TTGAGATGACGGACGGTGCGGCAA
12	CATCGTCCCTTTCGATGGGATCAA	TTGATCCCATCGAAAGGGACGATG
13	GCACGGGAGCTGACGACGTGTCAA	TTGACACGTCGTCAGCTCCCGTGC
14	AGACGCACCGCAACAGGCTGTCAA	TTGACAGCCTGTTGCGGTGCGTCT
15	CGTGTAGGGGTCCCGTGCTGTCAA	TTGACAGCACGGGACCCCTACACG
16	CATCGCTGCAAGTACCGCACTCAA	TTGAGTGCGGTACTTGCAGCGATG
17	GGCTGGTTCGGCCCGAAAGCTTAG	CTAAGCTTTCGGGCCGAACCAGCC

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number
WO 02/16649 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/26519

(22) International Filing Date: 27 August 2001 (27.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/227,948 25 August 2000 (25.08.2000) US
60/228,854 29 August 2000 (29.08.2000) US

(71) Applicant: ILLUMINA, INC. [US/US]; Suite 200, 9390 Towne Centre Drive, San Diego, CA 92121 (US).

(72) Inventor: GUNDERSON, Kevin; 1543 Juniper Hill Drive, Encinitas, CA 92024 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

TABLE 2

Seq. ID No.	Decoder Sequence (5'-3')	Probe Sequence (5'-3')
1	TTCGCCGTCGTGTAGGCTTTTCAA	TTGAAAAGCCTACACGACGGCGAA
2	TTCGAAGCGCACGTCCCTTTTCAA	TTGAAAAGGGACGTGCGCTTCGAA
3	AACGCGTGGGGAATGGGACATCAA	TTGATGTCCCATTCCCCACGCGTT
4	CCGTCGCATACCGGCTACGATCAA	TTGATCGTAGCCGGTATGCGACGG
5	ATGGCCGTGCTGGGGACAAGTCAA	TTGACTTGTCCCCAGCACGGCCAT
6	TTGCAACGGGCTGGTCAACGTCAA	TTGACGTTGACCAGCCCGTTGCAA
7	CGCATAGGTTGCCGATTTCGTCAA	TTGACGAAATCGGCAACCTATGCG
8	CCGTTTGCGGTCGTCCTTGCTCAA	TTGAGCAAGGACGACCGCAAACGG
9	TTCGCTTTCGTGGCTGCACTTCAA	TTGAAGTGCAGCCACGAAAGCGAA
10	GTCCAACGCGCAACTCCGATTCAA	TTGAATCGGAGTTGCGCGTTGGAC
11	TTGCCGCACCGTCCGTCATCTCAA	TTGAGATGACGGACGGTGCGGCAA
12	CATCGTCCCTTTCGATGGGATCAA	TTGATCCCATCGAAAGGGACGATG
13	GCACGGGAGCTGACGACGTGTCAA	TTGACACGTCGTCAGCTCCCGTGC
14	AGACGCACCGCAACAGGCTGTCAA	TTGACAGCCTGTTGCGGTGCGTCT
15	CGTGTAGGGGTCCCGTGCTGTCAA	TTGACAGCACGGGACCCCTACACG
16	CATCGCTGCAAGTACCGCACTCAA	TTGAGTGCGGTACTTGCAGCGATG
17	GGCTGGTTCGGCCCGAAAGCTTAG	CTAAGCTTTCGGGCCGAACCAGCC

Illumina attempt to re-patent Zirvi-Barany's *exact* zip code sequences; changes "IllumaCode ID" to "Seq. ID No.", and drops Mark Chee as inventor when moving from provisional to full application. What are they trying to hide?

Illumina submits full application 09/940,185 = WO 2002016649 on August 27, 2001.

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number
WO 02/16649 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/26519

(22) International Filing Date: 27 August 2001 (27.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/227,948 25 August 2000 (25.08.2000) US
60/228,854 29 August 2000 (29.08.2000) US

(71) Applicant: ILLUMINA, INC. [US/US]; Suite 200, 9390 Towne Centre Drive, San Diego, CA 92121 (US).

(72) Inventor: GUNDERSON, Kevin; 1543 Juniper Hill Drive, Encinitas, CA 92024 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

WO 02/16649 A2

(54) Title: PROBES AND DECODER OLIGONUCLEOTIDES

(57) Abstract: The present invention is directed to improved methods and compositions for the use of adapter sequences on arrays in a variety of multiplexed nucleic acid reactions, including synthesis reactions, amplification reactions, and genotyping reactions.

Illumina submits full application of "Probes and Decoder Oligonucleotides" on August 27, 2001, which deletes the words "**universal arrays**" in the abstract to the patent office.

1. **Adapter molecules and universal arrays and their use in detection of targets**
Quick View
By Gunderson, Kevin
From PCT Int. Appl. (2002), WO 2002016649 A2 20020228. | Language: English, Database: CAPLUS

Adapter molecules and universal arrays and their use in detection of targets

By: Gunderson, Kevin
Assignee: Illumina, Inc., USA

The present invention is directed to improved methods and compns. for the use of adapter sequences on arrays in a variety of multiplexed nucleic acid reactions, including synthesis reactions, amplification reactions, and genotyping reactions. Thus, a method for immobilizing target nucleic acids comprises attaching an adapter nucleic acid to a target nucleic acid to form a modified target nucleic acid, then contacting this modified nucleic acid with an array of at least 25 different addresses, each address contg. a different adapter probe. Further, a method for detecting target nucleic acids makes use of these immobilized target nucleic acids by providing a means of detecting the immobilized target mols. The use of adapter nucleic acids allows one to prep. "universal arrays", the capture probes of which can be modified by different adapter nucleic acids depending upon the use to which it is to be applied.

Illumina attempt to re-patent Zirvi-Barany's *exact* zip code sequences; deletes the words "**universal arrays**" in the abstract to the USPTO, in an effort to hide that Illumina's application is really the intellectual property of WO97/31256, i.e. the "Zip Code Chemistry" invented by the Barany lab.

Illumina own provisional patent application shows that Universal Bead Arrays are Zip Code Arrays.

Illumina 60/180810 Provisional, Filed Feb. 7, 2000

02-09-00

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

"EXPRESS MAIL" MAILING LABEL Number EL406567991US, Date of Deposit February 7, 1999

I hereby certify that this paper or fee and listed enclosures is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box Provisional Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231, on February 7, 1999

Typed or Printed Name: HAMMID SANCHEZ

Signed:

Docket Number:	P-68929/DJB/RMS/DCF	Type a plus sign (+) inside this box →

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
FAN	Jian-Bing		San Diego, California

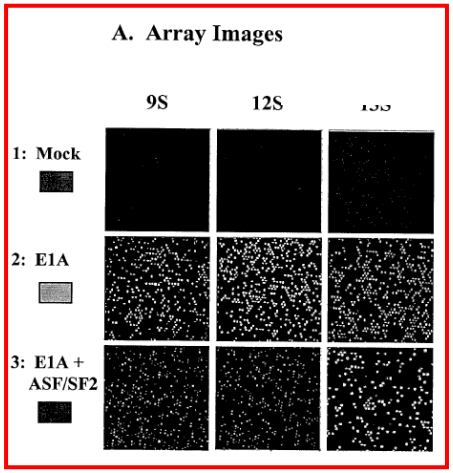
TITLE OF THE INVENTION (280 characters max)

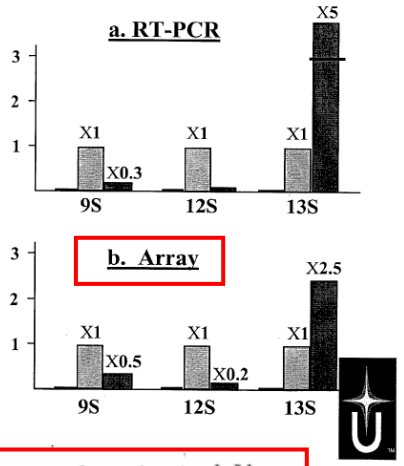
GENE EXPRESSION PROFILING

02/07/00 jc688 U.S. PTO

JC553 U.S. PTO 60/180810 02/07/00

Detection of Alternative Splicing on Zip-Code Arrays and Comparison with RT-PCR





Currently, Illumina has developed universal BeadArrays™ with 128 unique addresses that hybridize efficiently and with high specificity. Development of arrays with up to 2,000 addresses is in progress. In

Illumina admits – 4 years after Barany Lab's zip code patent was filed – that "Illumina has developed **Universal** BeadArrays with 128 unique addresses that hybridize efficiently and with high specificity". Such "**Universal** BeadArrays" are called Zip-Code Arrays and shown to work in their last figure, which tellingly is deleted from all subsequent patent applications from this patent family.

Illumina attempts to hastily rename probes containing zip codes as "adapters".

Illumina Provisional 60/311,194, Filed Aug. 9, 2001

08-10-01

A/Prov

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

"EXPRESS MAIL" MAILING LABEL Number EL 758644122 US

Date of Deposit August 9, 2001

I hereby certify that this paper or fee and listed enclosures is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box Provisional Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

Typed or Printed Name: Darryl Kriner

Signed:

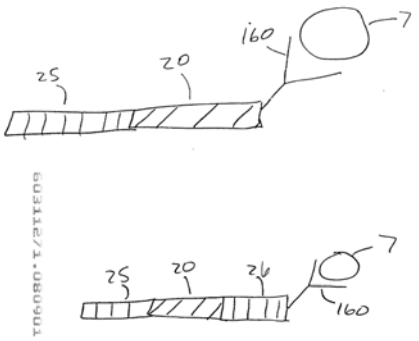
Docket Number:	P-68929-7/DJB/RMS/DCF	Type a plus sign (+) inside this box →

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Fan	Jian-Bing		San Diego, California

TITLE OF THE INVENTION (280 characters max)

MULTIPLEXED DETECTION METHODS



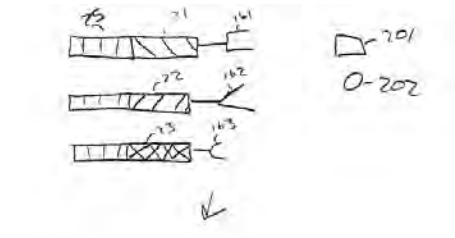


Figure 11

Figure 11 depicts two configurations of probes for multiplex detection of analytes. A depicts a probe containing an adapter **20**, an upstream priming site **25** and a target-specific portion, i.e. bioactive agent **160** bound to a target analyte **7**. B depicts a probe containing an adapter **20**, an upstream universal priming site **25**, a downstream universal priming site **26** and a target-specific portion, i.e. bioactive agent **160** bound to a target analyte **7.**

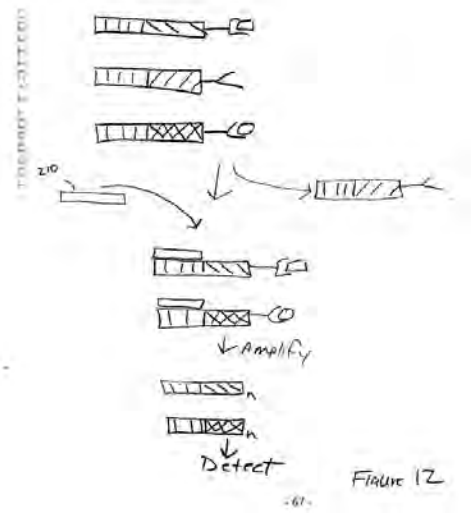


Figure 12

Figure 12 depicts a preferred method for multiplex detection of analytes. Probes containing universal priming sequence **25** and adapters that identify the target analyte to be detected **21, 22 and 23**, and target specific portions, i.e. bioactive agents **161, 162 and 163** are contacted with target analytes **201 and 202**. Probes to which target analytes bind are contacted with universal primers **210** and amplification reaction mixture. Amplicons are detected and serve as an indication of the presence of the target analyte.

Illumina hastily submits provisional 60/311,194 on Aug. 9, 2001, tries to rename probes containing zip codes, universal priming sequences, and target sequences as "adapters"

Illumina Makes Further Use of Barany Lab's Intellectual Property and Exact Zip Codes, and Uses Seventeen Different Alternative Names, to Obfuscate the Origin of Zip Codes to the USPTO and the Court

The term "zip code" enters Illumina patent applications only after JBA signed with PEB on November 9th, 1999

60090473 as originally filed on June 24, 1998
- All figures relate to using enzymes or antibodies on the surface of beads, none were retained in final application.

09189543 as originally filed on Nov. 10, 1998
- This version has no examples and no figures, and no explanation of decoding. No mention or thought of zip codes.

09344526 as originally filed on June 24, 1999 = US7060431
- This version has no examples and no figures, and no explanation of decoding. No mention or thought of zip codes.

60172106 as originally filed on Dec. 23, 1999
- Illumina officially begins collaboration with PE-Biosystems on 11-9-1999 – which had access to confidential Zirvi-Barany 465 set Zip codes and Zirvi-Barany Intellectual Property including trade secrets
- Example 1 and Figure 1 now show for the first time; decoding arrays, clearly using zip code idea, but deliberately concealed the sequence of oligonucleotides used and the actual hybridization and wash conditions they used.

60235531 as originally filed on Sept. 26, 2000
- Similar to above, highlight success of error correction, which is solely dependent on using zip code sequences.

09748706 as originally filed on Dec. 22, 2000 = US7033754
- Propose primer-extension approach to decode arrays, no evidence this ever worked. They accidentally use the words "zip code loci" in example 6. Imagine if Samsung had used the word "iPhone" in their patents with no explanation.

60302213 as originally filed on June 28, 2001
- No new examples provided, but some hastily thrown together figures speculating use of reversible terminators for decoding, in combination with zip codes. An exonuclease approach for decoding is described, but it will not work. Illumina plagiarizes the word "zipcode" zip code" or "cZip" 16 times, without ever defining the words.

10187321 as originally filed on June 28, 2002 = US7226734
- No new examples, but one new figure is provided with no data. The figure reveals that approximately 1520 beads were correctly decoded – this matches the 1536 – 16 = 1520 functional bead types in the Gunderson et. al 2004 paper wherein the same data was derived by serial hybridization of pools to zip code oligonucleotides onto the addressable arrays.

The term "zip code" enters Illumina patent applications only after a BA signed with PEB on November 9th, 1999

terminator colors. After the reaction is over the array tip is imaged to capture the color of the beads at this stage. Beads are then immediately stripped of color or label by denaturing and washing off the first extended primer at each zip code loci. This process can be repeated through multiple stages. In subsequent stages, the primer extension reactions contain the same reagents as the previous extension reaction except

09748706 as originally filed on Dec. 22, 2000: Illumina lawyers probably instructed inventors to remove the word "**zipcode**" from the patent application, but in their rush, they missed "**zip code loci**" in example 6, because it has a space between "**zip**" and "**code**". No explanation given for the term.

60302213 as originally filed on June 28, 2001: Illumina lawyers probably figured the collaboration with PE-Biosystems is going well, so left in "zipcode", "zipcodes", "zip code loci" "zip codes" and "cZip".

Figure 13 Depicts construction of probes on bead containing encoding sequences, zipcodes, and a gene-specific sequence. Two different encoding cassettes are employed to facilitate the primer extension reaction using primers with universal or degenerate bases. Using four color sequencing and a single base code, 6 bases generates 4^6 = 4096 codes, likewise four color hybridization using single hybridization colors also generates 4096 codes. The grand total number of codes is 4096*4096 = >16 million. If only a single color/two state scheme is employed for hybridization, than 2^6 = 64 codes are generated. The grand total is 4096*64 = 262,144 codes. The zipcode sequences can also be constructed so as to be overlapping to reduce the length of the overall sequence (i.e. cZip#1 vs. cZip#2).

09748706 as originally filed on Dec. 22, 2000 = US7033754

- Propose primer-extension approach to decode arrays, no evidence this ever worked. They accidentally use the words "zip code loci" in example 6. Imagine if **Samsung** had used the word "**iPhone**" in their **patents with no explanation**.

60302213 as originally filed on June 28, 2001

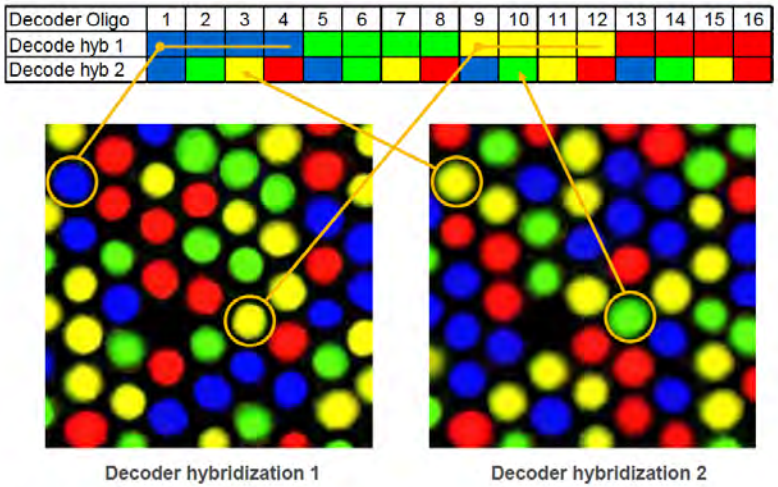
- No new examples provided, but some hastily thrown together figures speculating use of reversible terminators for decoding, in combination with zip codes. An exonuclease approach for decoding is described, but it will not work. Illumina plagiarizes the word "zipcode" zip code" or "cZip" 16 times, without ever defining the words.

Illumina's Proprietary Software Retains Fingerprints of Zip Code Inventors Intellectual Property

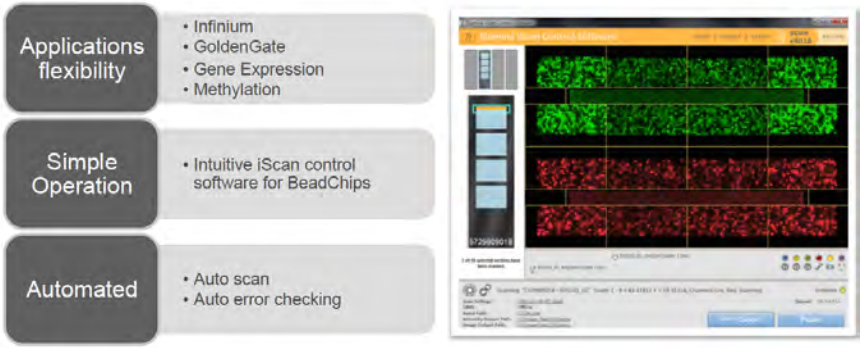
DMAP files are collections of Zip codes/Capture Oligonucleotides 1

Bead Decoding

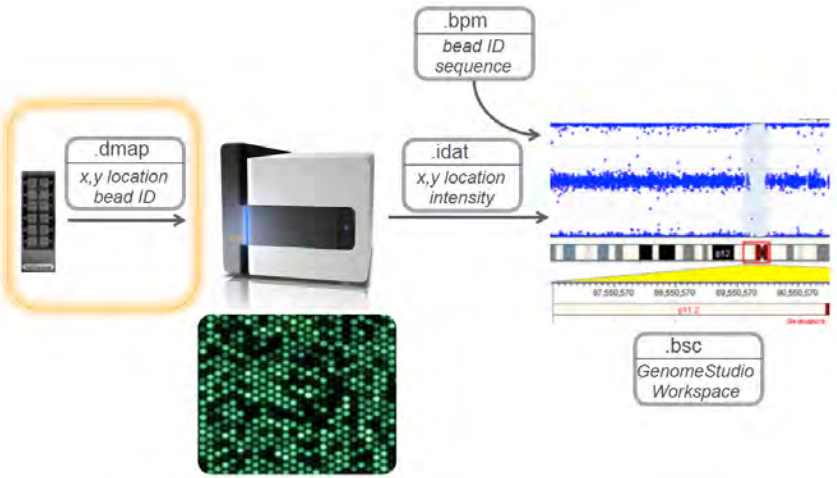
Example: 16 Bead Types



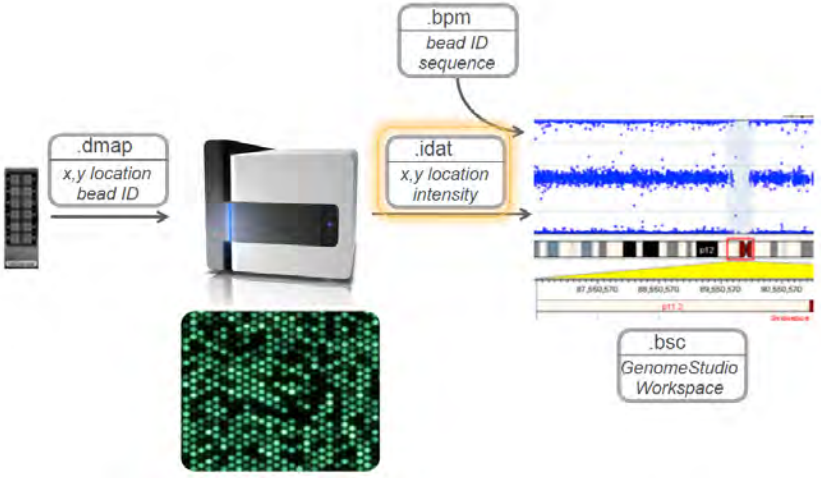
iScan Control Software



Collecting Data from a BeadChip: DMAPs



Collecting Data from a BeadChip: .idat



David Walt testimony on March 6, 2013 at Syntrix trial:

A. It's two-dimensional, because you can specify the location of any bead just by defining two dimensions: an X-dimension and a Y-dimension. So that will tell you exactly where each bead is. And the only way that one can do that is with something that is planar.

Q. It may have been clear to some, but not all. You describe here the arrangement of the beads on the top of a fiberoptic bundle as two-dimensional. If those beads were laid out on a silicon or a glass or a plastic slide, one of those substrates as described in your patent, would that be two-dimensional or a three-dimensional configuration?

A. That would be two-dimensional.

MR. ROSENBAUM: No further questions.

David Walt admits that one needs to specify the location of each bead just by defining two dimensions: an X-dimension and a Y-dimension. In other words the location of each bead is defined before use.

Why is the term "ZipCode" in Illumina's software? Part 2

The term "ZipCode" is used as a variable in multiple areas for software to identify/decode beads with XY addresses, just as in Universal Zip Code arrays as described by Barany Lab IP. It means that Illumina programmers thought of the sequences as "ZipCode" sequences in DMAP files used by Illumina.

There is no "area code" or "phone number" in any of the .dll files. There is also a parameter in the ZipCode data structure titled "ExcludeFromObfuscationAttribute" indicating they were trying to keep this data secret.

In File IlmnDataFiles.dll:

Int16 DESCryptoServiceProvider SymmetricAlgorithm set_Mode PaddingMode set_Padding set_IV set_Key ICryptoTransform CreateEncryptor CryptoStream CryptoStreamMode FlushFinalBlock WriteByte FileNotFoundException ReadByte CreateDecryptor get_Position FileLoadException NotSupportedException SetLength **ExcludeFromObfuscationAttribute** RuntimeTypeHandle GetTypeFromHandle System.Xml.Serialization XmlSerializer Serialize get_BaseStream IsNaN Deserialize ReadChars GetFileName ReadInt16 Read

Int64 BackgroundMean BackgroundStdDev **ZipCode** StdDev Mean Median NumNonOutlierBeads TrimmedMean Read

In File CommonVeraScan.dll:

LARGE_INTEGER **ZipCode** CoreData CoreDataArray CoreStatus Sorting EventRecursionBlocker EventBarrier RegistrationFormat RegistrationModeEnum RegistrationScoreStruct RegistrationParameters Polynomial Illumina.Common.LinearAlgebra IMatrix ICholeskyDecomposition ILuDecomposition IQrDecomposition ISingularValueDecomposition IEigenvalueDecomposition Matrix CholeskyDecomposition LuDecomposition QrDecomposition SingularValueDecomposition EigenvalueDecomposition MathHelper ILog Illumina DriveType DiskSpaceInfo VolumeInfo SystemUtils MathSupportFunctions ImageChangedEventArgs ImageSizeChangedEventArgs ImageSizeChangedEventHandler VisibilityChangedEventHandler IImageBase XYZIntPoint ProfileEventHandler Profiler ClonedProfilerData ProfileContextValue ProfileContextData ProfileContext Filename AdditionalContextException MultiException TransformationType Transformation TransformAttributes ProfileEvent EventType ProfileEventArgs EventsHelper AsyncFire mscorlib System ValueType Object Enum ICloneable EventArgs TEventArgs MulticastDelegate IDisposable Attribute ApplicationException X Y .ctor LoadLocations SaveLocations Z System.Drawing PointF GetHashCode Equals op_Equality op_Inequality String IsActualVersionA_AtLeast_TargetVersionB _Value value__ Synchronous Asynchronous GreenCY3 RedCY5 BeadEnd DistalEnd ArrayMatrix

Drawing2D ToAltString CalculateOrthogonality Invert TransformPoints TransformPoint TransformVectors TransformVector Transform InverseTransform Invert3x3 ConvertArray invert3x3 mult3x3WithVec mult3x3 Multiply3x3Precise CreateMatrix CreateTransformation CreateTransformationRobust Elements XOffset YOffset Rotation XScaling YScaling Shear Timestamp Type Context ActivityName ActivityFullName get_Event _event AsyncFireCleanUp Delegate InvokeDelegateBlocking InvokeDelegateHybrid FireEventBlocking FireEventAsync x y FilePath Locs OutputPath z orig obj left right actualVersionA featureVersionB numberOfSignificantSegments okIfUndefinedVersionA

resultIfUndefinedVersionA **zipcode** Xvalues Yvalues slope intercept System.Runtime.InteropServices OutAttribute list mad sortedList stddev sortedValues p lowerBound upperBound xyCoords zCoords.

Why is the term "ZipCode" in Illumina's software? Part 1A

"DMAP files identify bead locations on your BeadChip and quantify the signal associated with each bead."

DMAP Files

Identify bead locations on your BeadChip and quantify the signal associated with each bead.

Effective date: October 2012*

Concurrent with the upgrade of the Decode File Client to version 3.0, the retention policy for the DMAP files has been expanded to incorporate Illumina's continuing goal of world-class support and responsiveness to customer feedback.

DMAPs will be available a minimum of 12 months from the manufacturing date**, which corresponds to the maximum time from manufacturing to expiration for our current BeadChips. During this time, customers can download the DMAPs as many times as they wish and the files will not be deleted.

We are aware that some customers may wish to have the DMAPs available even after the arrays have expired. Although it is not advisable to run Illumina BeadChips after their expiration, the DMAPs that were not downloaded during the initial 12-month period will remain available for up to an additional 12-months (2-years from manufacturing). The only exception will be if the DMAPs are successfully dow
time frame. In this case, the DMAPs will be removed and unavailable for dow
successful download.

*Files created before this date are only available for 13 months from the date
**Manufacturing dates can be found on BeadChip packages

DMAP Decode File Download Utility v3.0.2

Instructions for installing and using the DMAP Download Client Utility to download DMAP files.

- Decode File Client Setup 64 3.0.2 - Unzip and launch to install. Do not uninstall any previous versions of the client.
- DMAP DLL Files - For the error "Error validating \xxx .dmap gz," refer to the installation instructions.

Files

FILE NAME
Decode File Client Setup 64 v3.0.2
DMAP DLL Files
Decode File Client v3.02 Software Release Notes
DMAP Decode File Client User Guide (11337856 C)
System Requirements for DMAP Decode File Download Utility v3.0.2

Why is the term "ZipCode" in Illumina's software? Part 1B

Each BeadChip requires the user to download a DMAP file for that array.

Finding BeadChip DMAP Files in Access by Account Mode

Access by Account mode enables you to download any DMAP files that you have purchased within the last 12 months through your MyIllumina account.

Access by Account Mode Tabs

In Access by Account mode, you will be able to see and use the Main, Download Status and Log, Alerts, SMTP Test, and Help tabs:

- **Main tab:** Enables you to find and download BeadChip DMAP files.
- **Download Status and Logs tab:** Enables you to view download progress and status, abort the download, and save a download log to a file.
- **Alerts:** When AutoPilot is selected on the Main tab, enables you to enter contact names and email addresses to which messages will be sent based on the parameters you select. You can have Decode File Client send you email notifications for the following conditions: job start, job finish, and errors. Each line in the table enables you to send out a message to one email for one type of condition. If you want to send emails to multiple email addresses, you can add multiple email addresses. If you want to send emails for different conditions to the same email address, you can enter the email address several times and select one condition for each instance.
- **SMTP Test:** Enables you to set SMTP parameters for alerts messaging.
- **Help tab:** Enables you to read the user help for the Decode File Client.

Selecting and Downloading BeadChip DMAP Files

Before downloading BeadChip DMAP files, ensure that you have enough free space on your computer's download destination.

To download the DMAP files from the list of found BeadChips, do the following:

View Available DMAP Files

1 On the **Main** tab, select one of the following options:

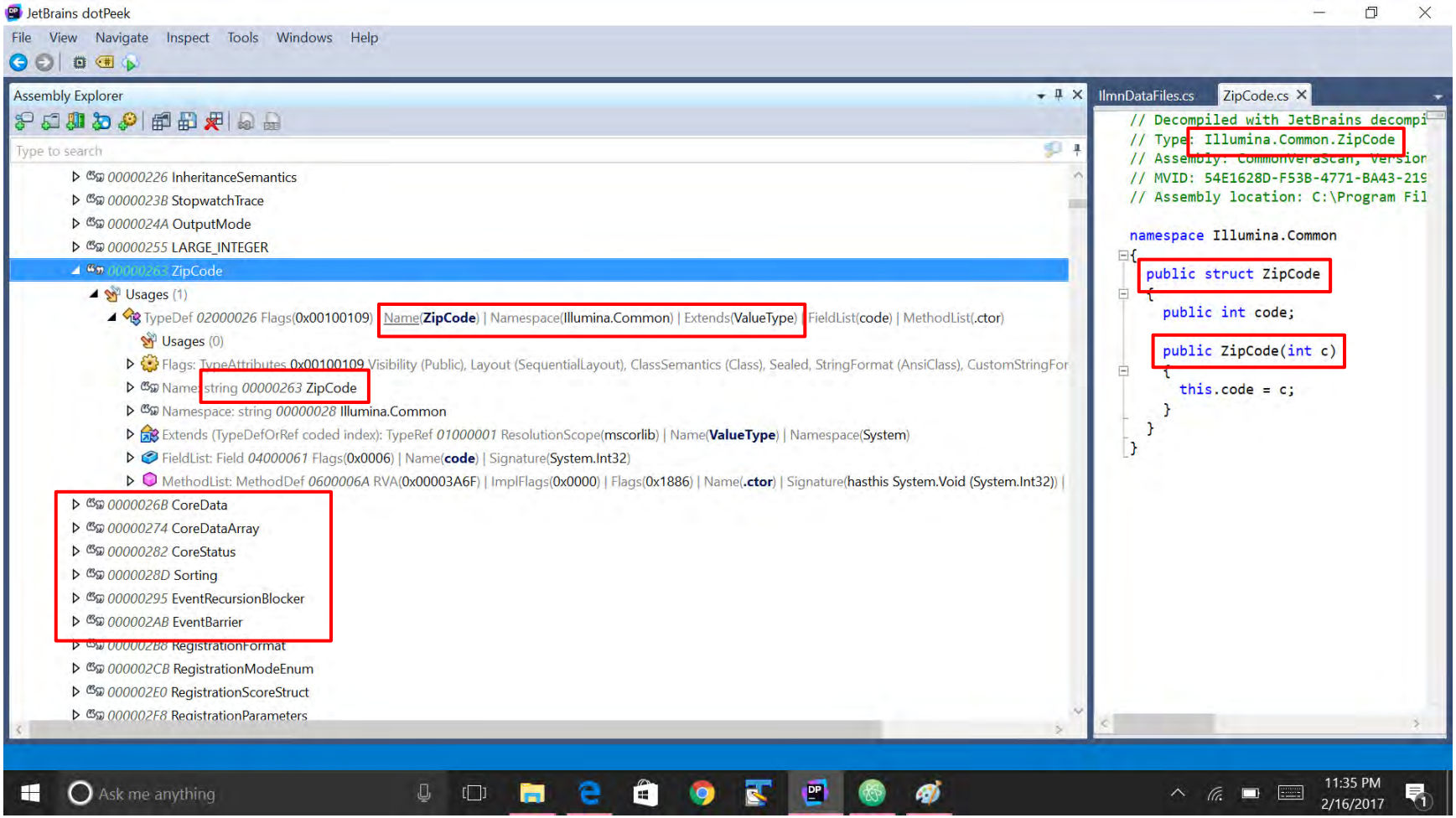
Option	Use
AutoPilot	Automatically checks for new DMAP files and downloads them as they become available. AutoPilot is set to check for new BeadChip decode files every 24 hours. To use AutoPilot, you will need to specify your SMTP server (using the SMTP Test tab) and set up an alert (using the Alerts tab).
All my BeadChips that have NOT been downloaded	Displays all serial numbers that have not yet been downloaded
All my BeadChips	Displays a list of all BeadChip serial numbers that you have ordered and that are available for download
BeadChips by Purchase Order	Displays all of the serial numbers that are associated with a specific PO# and are available for download
BeadChips by barcodes	Displays only the serial numbers for barcodes entered in the dialog. You may copy and paste or scan barcodes directly into the dialog box.

2 If you have selected BeadChips by Purchase Order or BeadChips by barcode, enter one or more purchase order numbers or barcodes respectively. If you have multiple purchase order numbers, enter them in the text box, separated by commas. If you have multiple barcodes, enter one per line.

3 Click **Find**. The Decode File Client displays a list of available BeadChip barcodes.

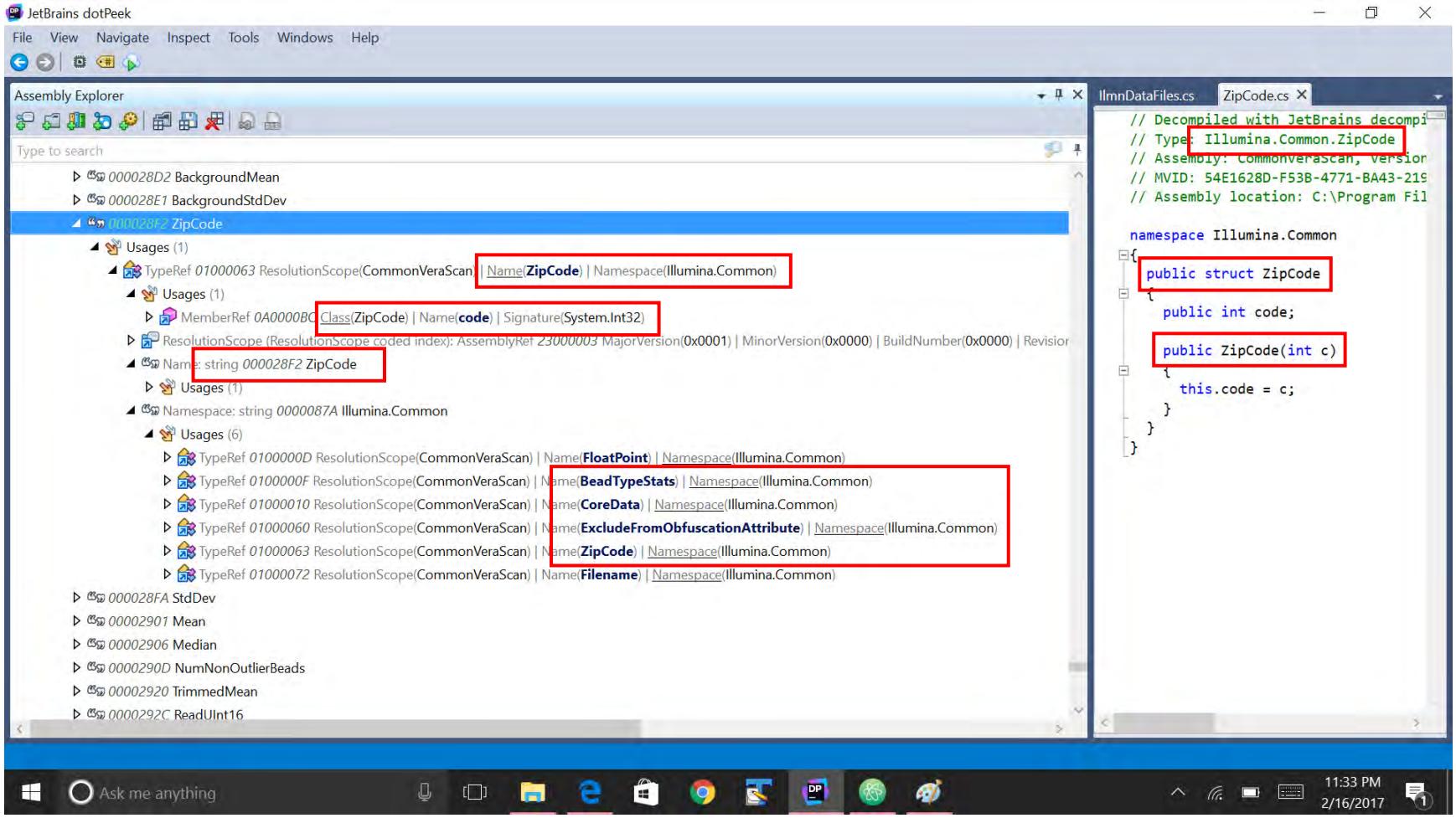
Why is the term "ZipCode" in Illumina's software? Part 4

Screenshot showing data structures labeled ZipCode in multiple places in these two important Dynamically Linked Libraries (.dll files) used by the DMAP software critical for determining the location of beads on a BeadChip and other Illumina arrays.



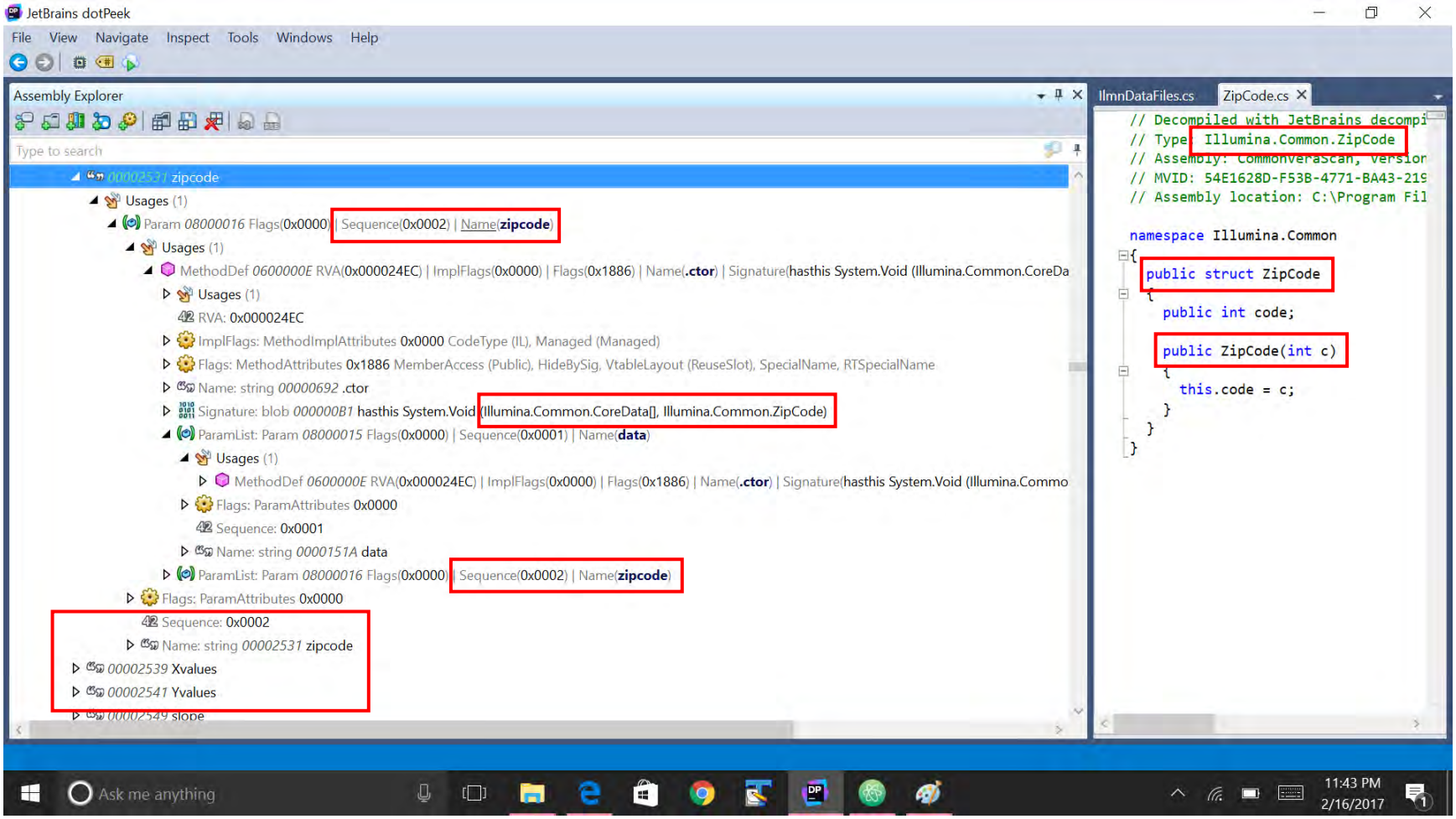
Why is the term "ZipCode" in Illumina's software? Part 3

Screenshot showing data structures labeled ZipCode in multiple places in these two important Dynamically Linked Libraries (.dll files) used by the DMAP software critical for determining the location of beads on a BeadChip and other Illumina arrays.



Why is the term "zipcode" in Illumina's software? Part 5

Screenshot showing data structures labeled "ZipCode" and "zipcode" in multiple places in these two important Dynamically Linked Libraries (.dll files) used by the DMAP software critical for determining the location of beads on a BeadChip and other Illumina arrays. This even lists "X values" and "Y values" right after the word "zipcode".



Illumina's Own FOIA Request to the NIH on Illumina's Grant Applications Triggers the True Inventors to Follow Illumina's Trail

Illumina unusual January 05, 2015 FOIA request for copies of its own grants:

Bartok, Lauren (NIH/OD) [C]

From: Noon, Will <wnoon@illumina.com>
Sent: Monday, January 05, 2015 7:16 PM
To: NIH FOIA
Subject: Freedom of Information Act Request

Hello,

I would like to request a copies of some funded grant applications under the Freedom of Information Act (FOIA). The grant information is below:

Project Number: 1R21HG001911-01 (Former Number 1R01HG001911-01)
Title: RANDOMLY ORDERED DNA ARRAYS FOR SNP DISCOVERY AND TYPING
Project Leader: Chee, Mark S.
Awardee Organization: Illumina, Inc.

Project Number: 1R44HG002003-01
Title: RANDOMLY ORDERED DNA ARRAYS FOR SNP GENOTYPING
Project Leader: Chee, Mark S.
Awardee Organization: Illumina, Inc.

Project Number: 1R43CA081952-01
Title: GENE EXPRESSION ANALYSIS ON RANDOMLY ORDERED DNA ARRAYS
Project Leader: Chee, Mark S.
Awardee Organization: Illumina, Inc.

Project Number: 1R43CA083398-01
Title: PARALLEL ARRAY PROCESSOR
Project Leader: Chee, Mark S.
Awardee Organization: Illumina, Inc.

I am an in-house attorney at Illumina, Inc., the awardee organization for all of these grants, and therefore am requesting this document on behalf of the awardee. Illumina, Inc. approves of the release of the grant application, solely to *Illumina, Inc., without any redactions.* **If a FOIA request is not the correct procedure for an awardee to request a copy of their own grant application, please let me know what the appropriate procedure would be.**

My contact information is:

William Noon
Illumina, Inc.
5200 Illumina Way
San Diego, CA 92122
Phone: 858-202-4780
Email: wnoon@illumina.com

I am willing to accept electronic copies of the documents via email.

If you have any questions or require further information, please let me know. Thank you in advance for your assistance.

Illumina: Fan SBIR 1R43CA097851-01 Grant, submitted November 30, 2001

C.4 High Throughput SNP Genotyping on Fiber Optic Arrays

We have developed a highly multiplexed method for SNP genotyping by combining an oligonucleotide ligation-based assay (OLA) with read-out on random arrays of universal capture probes (Gerry et al., 1999) (Fig. 6). There are a number of advantages of this assay system, outlined below.

Universal Array. By designing the assay system to use arrays of universal capture probes, many different sets of SNPs or methylation target sequences can be analyzed using a standard array. This provides a great deal of flexibility, and also reduces costs. The universal array contains probes that are sufficiently long to hybridize with high specificity. Similar approaches have been used in parallel analyses of yeast deletion strains (Shoemaker et al., 1996; Winzeler et al., 1999) and SNP genotyping (Fan et al., 2000; Gerry et al., 1999).

Sequence Specificity. The oligo ligation assay requires two different oligonucleotides to hybridize contiguously before ligation can occur. In addition, the ligase enzyme discriminates against even a single base mismatch in the vicinity of the ligation site. This built-in specificity improves the ability to analyze complex genomic samples accurately. Complex samples, such as human genomic DNA, contain many similar sequences and hence provide more opportunities for incorrect hybridization.

Multiplexing. The design of the assay is such that the oligo ligation step occurs first, on genomic DNA, and is then followed by PCR using universal primers incorporated in the ligation oligos (Fig. 6). This is in direct contrast to most other genotyping approaches, in which PCR amplification occurs first. Because the approach shown in Fig. 6 uses universal primers, and the PCR templates are relatively short, the robustness of multiplex PCR amplification is increased. We have been able routinely to carry out high levels of multiplexing. This is an important factor in reducing assay costs and increasing the scalability of the system.

Applicants are aware that reviewers penalize NIH grant applications that are not properly referenced. Thus, Zirvi's FOIA request reveals for the first time that Illumina's J.B. Fan admits that the highly multiplexed method for SNP genotyping is based on direct ligation of two oligonucleotides that "hybridize contiguously" with readout on "random arrays of universal capture probes", and properly cites Gerry-Barany 1999. Other than Illumina's provisional patent application 60/180810 (submitted by J.B. Fan, with an embedded grant section written by Fu, where the word "Zip code Array" was accidentally left in a Figure), Illumina consistently tried to obfuscate that their bead arrays were literally zip code arrays as described by Barany Lab in the WO97/31256 patent.

Illumina: Fan SBIR 1R43CA097851-01 Grant, submitted November 30, 2001

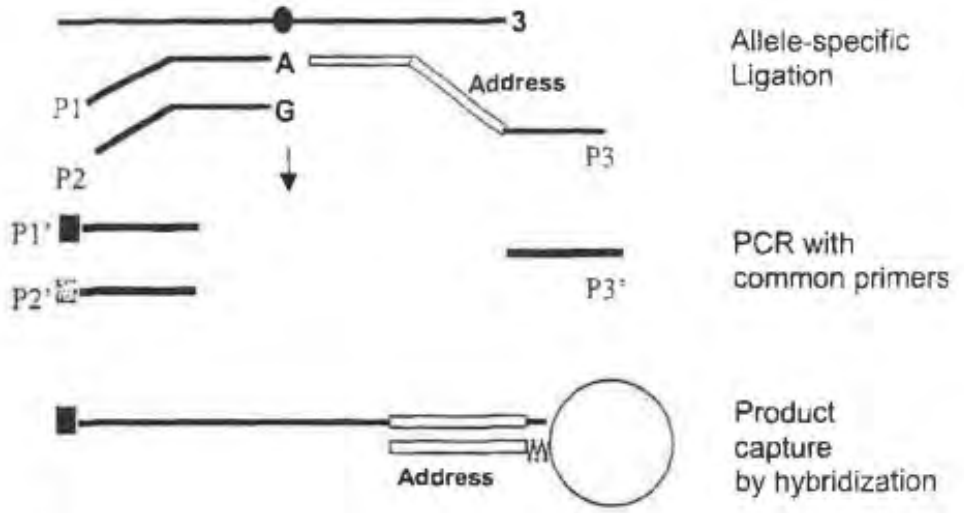


Figure 6. Schmatic of the OLA-PCR SNP genotyping assay. For each SNP, three OLA probes are designed: two allele-specific oligos, each corresponding to one allele, and one locus-specific oligo. The ASOs consist of two parts: the SNP-specific sequence and a universal PCR primer sequence (P1 or P2) at their 5'-end. The LSOs consist of three parts: the SNP-specific sequence, a unique address sequence which is complementary to a capture sequence immobilized on the array, and a universal PCR primer sequence (P3) at their 3'-end. A ligation reaction joins the ASO and LSO oligos to create a PCR template that can be amplified with universal primers (P1', P2', and P3). The ligation reaction provides allele selectivity: only if the 3' end nucleotide on the ASO matches the SNP sequence in the template is ligation carried out efficiently. The PCR products, wich are fluorescently labeled by incorporation of fluorophors at the 5' ends of P1' and P2', are hybridized to capture probes on the beads in the array. The ratio of the fluorescent signals from two allele specific ligation products indicates the genotype.

Through a FOIA request, Zirvi learns that Illumina's J.B. Fan grant application to the NIH on November 30, 2001, literally infringes on Barany Lab's LDR-PCR technology as covered in Barany Lab's '917 patent filed Feb 9, 1996, Barany Lab's '470 patent filed May 29, 1996, and Barany Lab's '293 patent filed Jan 6, 1999. It is a direct ligation (i.e. LDR) followed by PCR amplification with universal PCR primers, followed by zip-code capture on a solid support. These are outside the scope of the original Joint Development Agreement between Illumina and PE Biosystems, which only covered the IP in WO97/31256 patent ('917 series). PE Biosystems, which was responsible for overseeing the collaboration through a joint steering committee, deliberately withheld knowledge of this infringement to defraud the true inventors.

Illumina: Chee SBIR 1U54HG002753-01 Grant, submitted May 28, 2002

D.1.1 Assay Format is Designed for Multiplexing

Proprietary Info

Proprietary Info

The first step in conventional SNP genotyping is to amplify the SNP of interest from genomic DNA. In contrast, we perform an allelic discrimination step directly on genomic DNA. This is done using two allele-specific oligos, each 5' tailed with a different universal PCR priming sequence (Fig. 6). The product of this enzymatic procedure is an allele-specific PCR template. (In the case of a heterozygous DNA, both allele-specific PCR templates are produced at this step). PCR is then carried out, using only three primers (P1', P2' and P3' in Fig. 6). We have found that this procedure is amenable to high levels of multiplexing. Currently, when sufficient SNPs are available, we multiplex up to 1,152 genotyping reactions

Proprietary Info

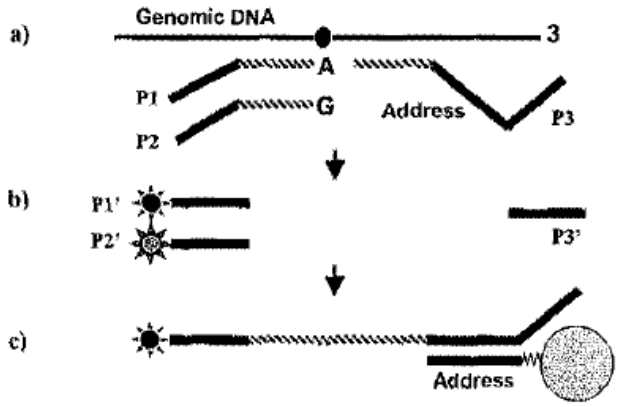


Figure 6. Assay format. a) For each SNP of interest, two allele-specific oligos and a locus-specific common oligo are annealed to genomic DNA (e.g. in a 1,152-plex reaction, a total of 3,456 oligos are annealed simultaneously, in the same reaction well in a microtiter plate). If an allele-specific oligonucleotide is complementary to the genomic DNA, a ligation product is formed. This product has universal priming sites at the 5' and 3' ends (i.e. P1 or P2, and P3). If the genomic DNA is heterozygous, then two products are formed: P1-P3 and P2-P3. b) Universal primers are added and PCR is carried out. The two allele-specific universal

Through a very recent FOIA request, Zirvi learns that Illumina's Mark Chee submitted a grant application to the NIH on May 28th, 2002, and Figure 6 reveals that this grant is completely based on Barany Lab's LDR-PCR technology as covered in Barany LAb's '917, '470, and '293 patent families. Illumina forgot to redact the figure legend which states "If an allele-specific oligonucleotide is complementary to the genomic DNA, a ligation product is formed" literally as described in the Barany Lab patents.

Illumina: Chee SBIR 1U54HG002753-01 Grant, submitted May 28, 2002

primers, P1' and P2' are fluorescently labeled, each with a different dye. Each amplicon contains an address that is complementary to a probe in the array, so that the genotype of each SNP can be read out on a different bead type in the array. c) The PCR amplicons are hybridized to an array of beads. The ratio of the two fluorescent signals indicates the genotype.

Another key aspect of the assay design is the incorporation of an address sequence, so that the assay products can be read out on a universal array (Chen et al. 2000; Fan et al. 2000; Gerry et al. 1999; Iannone et al. 2000). This provides flexibility. The probes on the array are random, artificial sequences that are not SNP-specific. Any set of SNPs can be analyzed simply by building the address sequences into the SNP-specific assay oligonucleotides (Section D.1.1).

Proprietary Info

Proprietary Info

genome. The use of a universal array simplifies manufacturing and reduces costs. The universal array is implemented on our BeadArray™ platform, detailed below.

D.1.2 Array Matrix Platform

The randomly ordered BeadArray technology, invented at Tufts University (Michael et al. 1998; Walt 2000), has been developed at Illumina as a platform for SNP genotyping and other high-throughput assays. Each array is assembled on an optical imaging fiber bundle consisting of about 50,000 individual fibers fused together into a hexagonally packed matrix. The ends of the bundle are polished, and one end is etched to produce a well in each fiber. This process takes advantage of the intrinsic structure of the optical fibers in the bundle (Fig. 7).

Through a very recent FOIA request, Zirvi learns that Illumina's Mark Chee submitted a grant application to the NIH on May 28th, 2002, and Figure 6 legend continued reveals that this grant is completely based on Barany Lab's LDR-PCR technology as covered in Barany Lab's '917, '470, and '293 patent families. Zirvi's FOIA request reveals for the first time that Illumina's M. Chee admits that the "assay products can be readout on a universal array" and properly cites Gerry-Barany 1999. Again, Illumina chose to redact a key sentence about the Barany Lab's Universal arrays. Since the material is over 15 years old, it is difficult to understand what would be proprietary, or suitable for a patent submission. Discovery of an un-redacted version of this application would reveal if Illumina deliberately hid information from the NIH to fraudulently obtain US government funding.

Illumina and PE Biosystems both have defrauded the True Inventors of Rightful Royalties: Appendix

Gerry-Barany JMB 1999 publication: Zip code, cZip, cZ1, Zip1, Z1.

Gerry-Barany Publication , Sept. 17, 1999

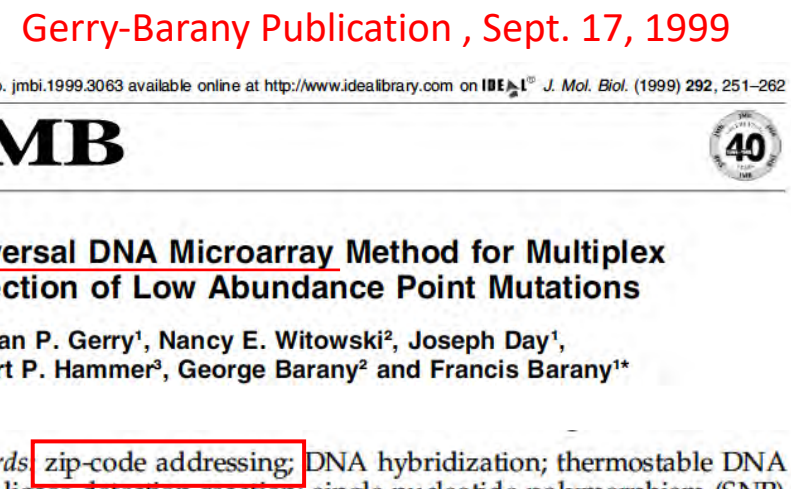
Article No. jmbi.1999.3063 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 292, 251–262

JMB 40

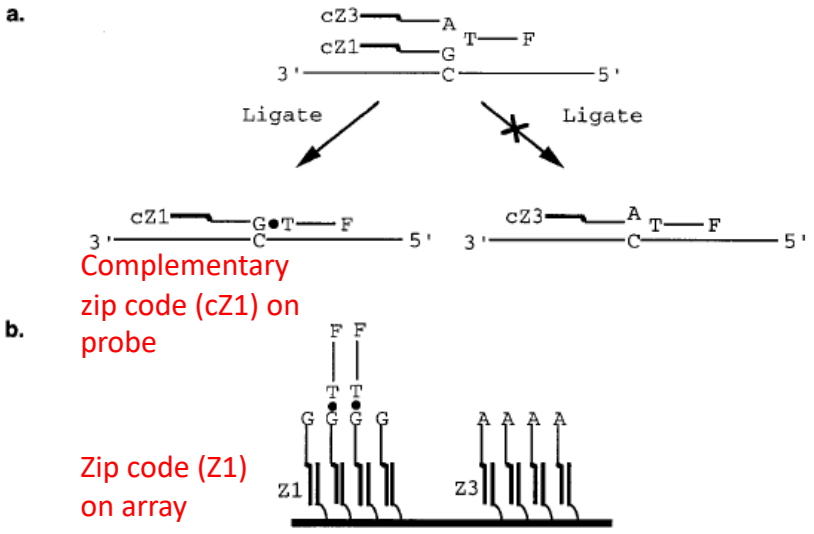
Universal DNA Microarray Method for Multiplex Detection of Low Abundance Point Mutations

Norman P. Gerry[1], Nancy E. Witowski[2], Joseph Day[1], Robert P. Hammer[3], George Barany[2] and Francis Barany[1]*

Keywords: zip-code addressing; DNA hybridization; thermostable DNA ligase; ligase detection reaction; single nucleotide polymorphism (SNP)

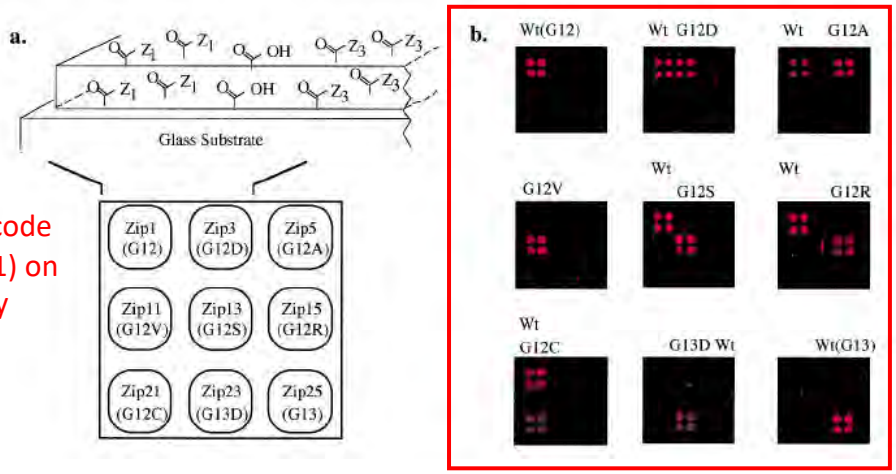


Zip code (Zip1) on probe



Complementary zip code (cZ1) on probe

Zip code (Z1) on array



Zip code (Zip1) on array

Barany *et al.*, Filed Feb 9, 1996

(12) **United States Patent**
Barany et al.

(10) **Patent No.:** **US 7,083,917 B2**
(45) **Date of Patent:** ***Aug. 1, 2006**

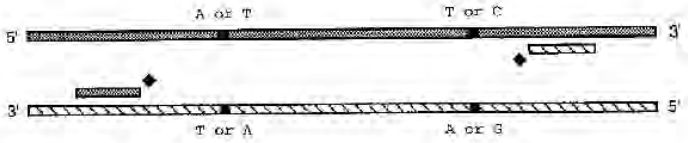
(54) **DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING THE LIGASE DETECTION REACTION WITH ADDRESSABLE ARRAYS**

(75) Inventors: **Francis Barany**, New York, NY (US); **George Barany**, Falcon Heights, MN (US); **Robert P. Hammer**, Baton Rouge, LA (US); **Maria Kempe**, Minneapolis, MN (US); **Herman Blok**, Wemeldinge (NL); **Monib Zirvi**, New York, NY (US)

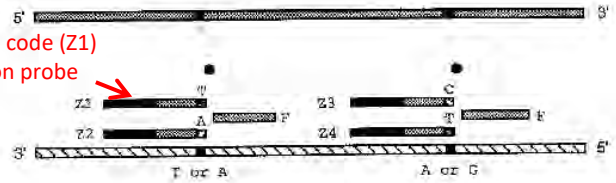
5,143,854	A	9/1992	Pirrung et al.
5,202,231	A	4/1993	Drmanac et al.
5,258,506	A	11/1993	Urdea et al.
5,278,298	A	1/1994	Chakraborty et al.
5,288,468	A	2/1994	Church et al.
5,290,925	A	3/1994	Fino
5,314,809	A	5/1994	Erlich et al.
5,324,633	A	6/1994	Fodor et al.
5,352,582	A	10/1994	Lichtenwalter et al.
5,371,241	A	12/1994	Brush et al.
5,391,480	A	2/1995	Davis et al.
5,405,783	A	4/1995	Pirrung et al.
5,407,798	A	4/1995	Martinelli et al.

PCR/ LDR

1. PCR amplify region(s) containing mutations using primers, dNTPs and *Taq* polymerase.



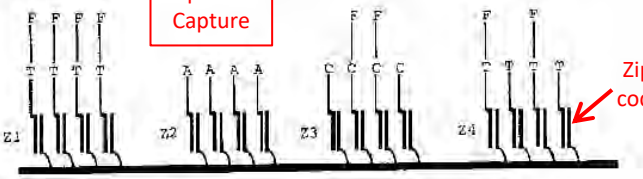
2. Perform LDR using allele-specific LDR primers and thermostable ligase. Allele specific oligonucleotides ligate to common oligonucleotides only when there is perfect complementarity at the junction.



Zip code (Z1) on probe

3. Capture fluorescent products on addressable array and quantify each allele.

Zip code Capture



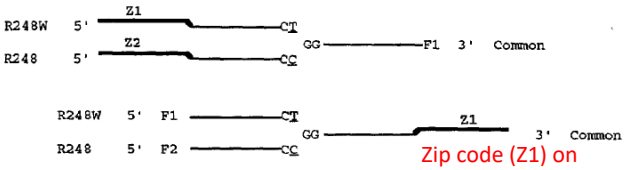
Zip code (Z1) on Array

Zip code

Project 5 will develop solid-phase approaches for the simultaneous detection of LDR products. Products will be captured on a spatially addressable array, so that the position of a signal identifies a mutation. Each LDR product will have a "zip code" tail, which will be captured selectively by a "complementary zip code" on the solid support. Multiple reuse of a universal "complementary zip code" array is envisaged to allow detection of a wide range of cancers and genetic diseases.

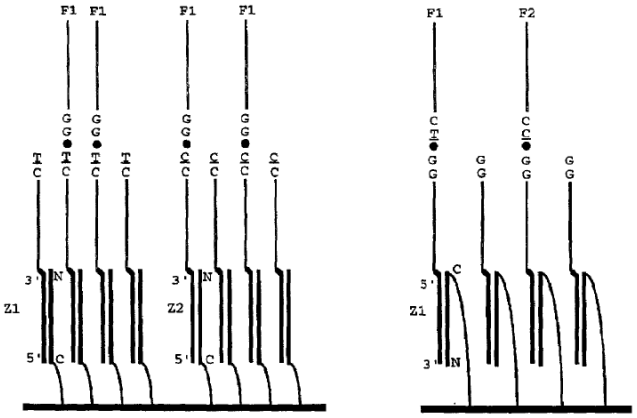
1994 NCI grant application had "zip code" tail on LDR product, and universal "complementary zip code" on the array.

Zip code (Z1) on upstream probe



Zip code (Z1) on downstream probe

Zip code (Z1) on Array



Zip code (Z11) on Array

TABLE 2

List of all 5 DNA/PNA oligonucleotide address sequences.

Zip #	Zip code	Sequence (5'→3' or NH_2 → COOH)	G + C
Zip11	1-4-3-6-6-1	TGCG-GGTA-CAGC-ACCT-ACCT-TGCG (SEQ. ID. No. 2)	15
Zip12	2-4-4-6-1-1	ATCG-GGTA-GGTA-ACCT-TGCG-TGCG (SEQ. ID. No. 3)	14
Zip13	3-4-5-6-2-1	CAGC-GGTA-GACC-ACCT-ATCG-TGCG (SEQ. ID. No. 4)	15
Zip14	4-4-6-6-3-1	GGTA-GGTA-ACCT-ACCT-CAGC-TGCG (SEQ. ID. No. 5)	14
Zip15	5-4-1-6-4-1	GACC-GGTA-TGCG-ACCT-GGTA-TGCG (SEQ. ID. No. 6)	15

Zip code (12, 14) on Array

TABLE 3

	Oligonucleotides used (5' to 3')
12	Aminolink- spacer 18- ATC GGG TAG GTA ACC TTG CGT GCG (SEQ. ID. No. 13)
14	Aminolink- spacer 18- GGT AGG TAA CCT ACC TCA GCT GCG (SEQ. ID. No. 14)
comp 12	CGC ACG CAA GGT TAC CTA CCC GAT (SEQ. ID. No. 15)
comp 14	CGC AGC TGA GGT AGG TTA CCT ACC (SEQ. ID. No. 16)

Complementary Zip code (comp 12, and comp 14) on Probe

Joint Development Agreement, November 9th, 1999

1.3. "Pre-Collaboration Illumina Intellectual Property" means all Intellectual Property Rights that are owned by, either partially or wholly, or licensed to, or otherwise controlled by, Illumina as of the Effective Date.

1.4. "Collaboration Illumina Intellectual Property" means all Intellectual Property Rights arising out of work performed under this Agreement that are conceived solely by one or more employees or agents of Illumina or its Affiliates.

6. **Intellectual Property; Patent Prosecution and Litigation; Licenses; Trademarks**

6.1. Ownership of Intellectual Property.

6.1.1. Pre-Collaboration Illumina Intellectual Property and Collaboration Illumina Intellectual Property. All rights and title to Pre-Collaboration Illumina Intellectual Property and Collaboration Illumina Intellectual Property, whether patentable or copyrightable or not, will belong to Illumina and will be subject to the terms and

6.2. Filing of Patent Applications.

6.2.1. Collaboration Illumina Intellectual Property. Illumina will have the first right, using in-house or outside legal counsel selected by Illumina's sole discretion, to prepare, file, prosecute, maintain and extend patent applications for Collaboration Illumina Intellectual Property in countries of Illumina's choosing. Illumina will bear all costs relating to such activities. Illumina will solicit PEB's advice and review of the patent applications, and Illumina will take into consideration PEB's advice thereon. If Illumina elects not to

Illumina's intellectual property should have been reviewed by PE Biosystems. Thus, PE Biosystems was an "enabler" for Illumina, and did not notify the true inventors that their IP was being used or copied.

Zip code, cZip, Decoders, IllumaCode, IlluniCode, Capture Probe, Bead identifier, Adapter Sequence, Universal Tag Sequence

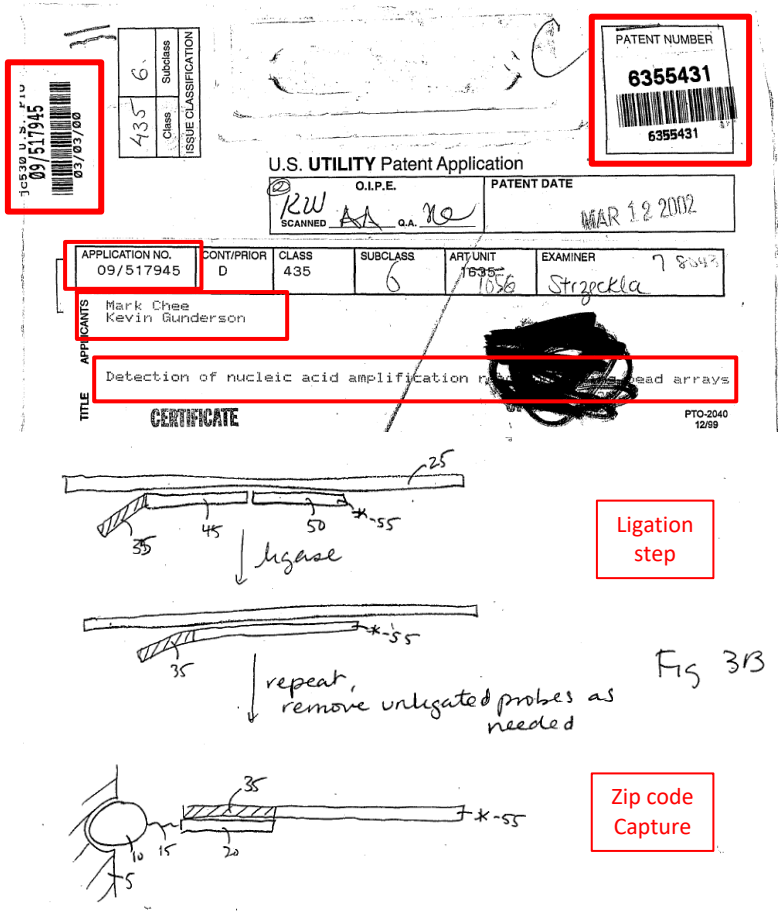


Fig 3B

Figures 3A and 3B depict two preferred embodiments of OLA amplification. Figure 3A depicts a first ligation probe **45** and a second ligation probe **50** with a label **55**. Upon addition of the ligase, the probes are ligated. The reaction can be repeated and then the ligated primer is added to the array as above. Figure 3B depicts the same reaction but using adapter sequences.

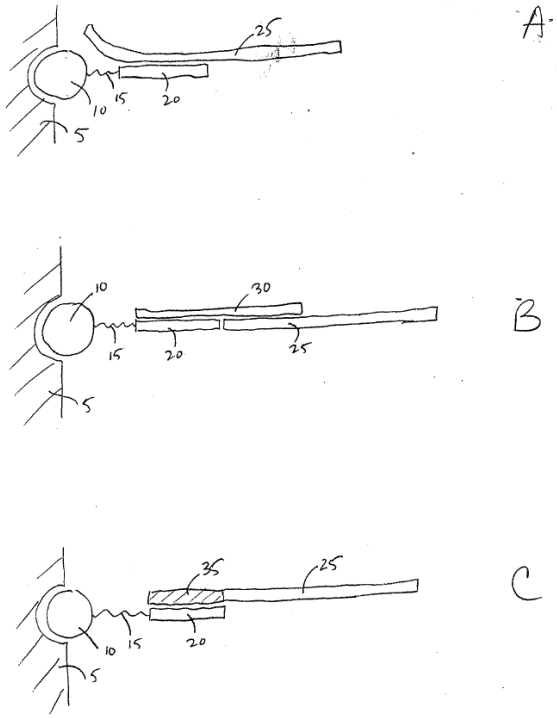


Fig 1

Figures 1A, 1B and 1C depict three different embodiments for attaching a target sequence to an array. The solid support **5** has microsphere **10** with capture probe **20** linked via a linker **15**. Figure 1A depicts direct attachment; the capture probe **20** hybridizes to a first portion of the target sequence **25**. Figure 1B depicts the use of a capture extender probe **30** that has a first portion that hybridizes to the capture probe **20** and a second portion that hybridizes to a first domain of the target sequence **25**. Figure 1C shows the use of an adapter sequence **35**, that has been added to the target sequence, for example during an amplification reaction as outlined herein.

In these hastily drawn Illumina patent application figures, Illumina uses the terms “adapter sequence” and “capture probe” to describe Barany Lab’s zip code sequences.

Illumina submits provisional application 60/135,123 (K. Gunderson) on May 20, 1999:

A/PROV

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

"EXPRESS MAIL" MAILING LABEL Number EM193415915US. Date of Deposit May 20, 1999. I hereby certify that this paper or fee and listed enclosures is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box Provisional Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231, on May 20, 1999

Typed or Printed Name: APRIL RICH

Signed:

05/20/99 JC571 U.S. PTO

Docket Number:	P-68088/DJB/RMS	Type a plus sign (+) inside this box → +

INVENTOR(S)/APPLICANT(S)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
GUNDERSON	Kevin		Encinitas, California

jc542 U.S. PTO 60/135123 05/20/99

TITLE OF THE INVENTION (280 characters max)

ADDRESSING ARRAYS USING SEQUENCE SPECIFIC ADAPTERS

ADDRESSING ARRAYS USING SEQUENCE SPECIFIC ADAPTERS

The present invention is directed to the use of "universal bead arrays" and methods thereof. A type of addressable array has been previously described, see WO 97/31256, hereby incorporated by reference in its entirety.

Barany Lab Patent

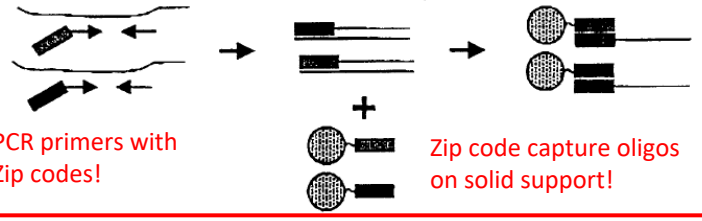
INVENTION DISCLOSURE FORM

1. **Name:** Kevin Gunderson, Mark Chee, John Stuelpnagel
2. **Date:** 5/10/1999
3. **Title of the Invention:** **Addressing an Array using Sequence-Specific Adapters**
4. **Describe the invention:** Use additional sheets if necessary. Attach descriptive materials such as drawings, sketches, photographs, etc. which may help illustrate the invention. Delineate new and important features. Make sure to include both the preferred embodiment as presently identified, and alternative constructions, procedures or equivalent components which can accomplish the same result as the preferred embodiment.

This invention disclosure describes the use of DNA adapters in conjunction with arrays comprising nucleic acids.

By attaching a specific hybridizing sequence, or "adapter" (e.g. DNA oligonucleotide) to a molecule of interest, and providing a complementary hybridizing sequence attached to a solid support (e.g. in a bead array format), the molecule of interest can be targeted specifically to the solid support. This provides a means of addressing a molecule of interest to a specific location. In the case of a randomly ordered array, the location, although specific, does not need to be known at the time of addressing. It can be determined after the fact, by decoding (as described previously in an Illumina patent application).

The adapter can be attached to the molecule of interest in a number of ways. If the molecule of interest is a nucleic acid sequence, the adapter can be incorporated by ligation, by chemical attachment, by synthesis, or other methods for joining nucleic acids. An example of incorporation by means of a PCR reaction is shown in Figure 1:



PCR primers with Zip codes!

Zip code capture oligos on solid support!

Figure 1. Incorporation of a nucleic acid adapter into a PCR product and subsequent targeting to a specific location. The adapter is chemically synthesized as part of a PCR primer. PCR reactions can be carried out singly or in multiplex format. The left panel shows 2 PCR reactions being carried out together, with each reaction labeled with a different adapter sequence (shown as grey and black rectangles). The center panel shows how the labeled PCR products can be mixed with solid supports carrying sequences complementary to the adapters. The right hand panel shows how each adapter binds to its complement, effectively targeting each PCR product to a separate location.

5. **State the primary purpose of the invention, including the need satisfied or problem solved by the invention:**

The invention allows a specified set of nucleic acid sequences to be used as "adapters" for many different assays, so that the assays, regardless of their composition, can be directed to specific locations and resolved from one another on an array comprising sequences complementary to the adapters.

6. **Prior art.** Include references, articles, talks, abstracts, patents, etc. which are relevant to either the state of the prior art or to the invention. Please include dates and provide copies whenever possible:

Chee, M. S. (1991) "Enzymatic multiplex DNA sequencing" Nucleic Acids Research **19**, 3301-3305 and refs therein.

Shoemaker, D. D. et al. (1998) Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nature Genetics **14**, 450-456.

Illumina attempts to re-patent Barany Lab's universal arrays and zip code primers, by trying to rename them as "adapter".

Illumina's provisional application 60/135,123 (K. Gunderson) tries to re-patent Barany Lab's ideas from 3 years earlier.

Barany *et al.*, Filed Feb 9, 1996

(12) **United States Patent**
Barany et al.

(10) **Patent No.: US 7,083,917 B2**
(45) **Date of Patent: *Aug. 1, 2006**

(54) **DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING THE LIGASE DETECTION REACTION WITH ADDRESSABLE ARRAYS**

(75) Inventors: **Francis Barany**, New York, NY (US); **George Barany**, Falcon Heights, MN (US); **Robert P. Hammer**, Baton Rouge, LA (US); **Maria Kempe**, Minneapolis, MN (US); **Herman Blok**, Wemeldinge (NL); **Monib Zirvi**, New York, NY (US)

5,143,854	A	9/1992	Pirrung et al.
5,202,231	A	4/1993	Drmanac et al.
5,258,506	A	11/1993	Urdea et al.
5,278,298	A	1/1994	Chakraborty et al.
5,288,468	A	2/1994	Church et al.
5,290,925	A	3/1994	Fino
5,314,809	A	5/1994	Erlich et al.
5,324,633	A	6/1994	Fodor et al.
5,352,582	A	10/1994	Lichtenwalter et al.
5,371,241	A	12/1994	Brush et al.
5,391,480	A	2/1995	Davis et al.
5,405,783	A	4/1995	Pirrung et al.
5,407,798	A	4/1995	Martinelli et al.
5,412,087	A	5/1995	McGall et al.

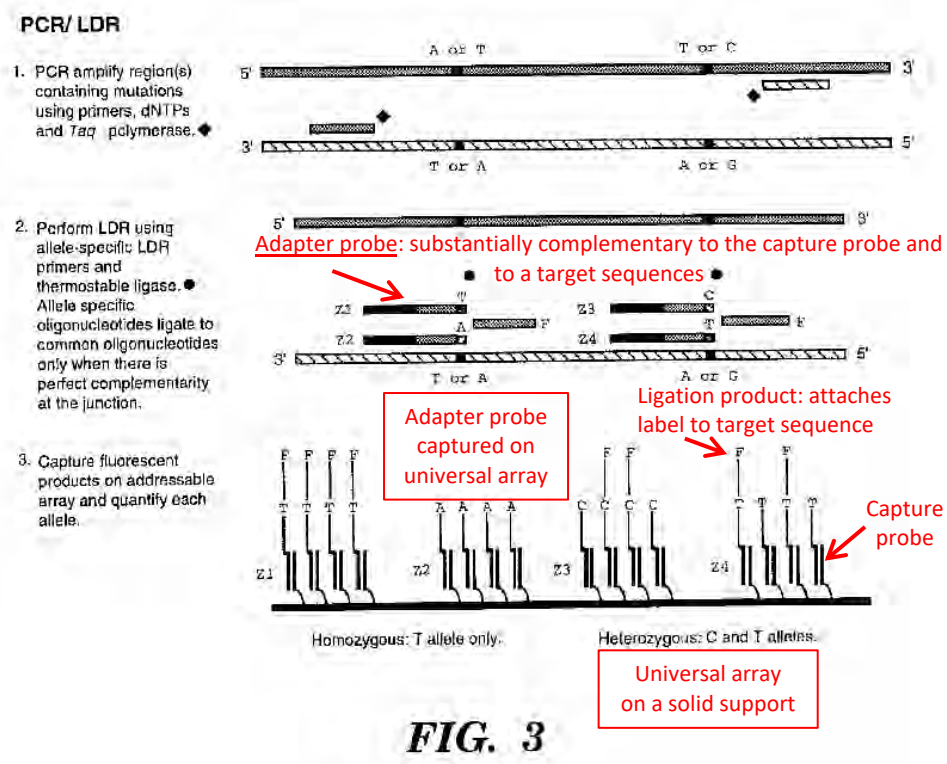
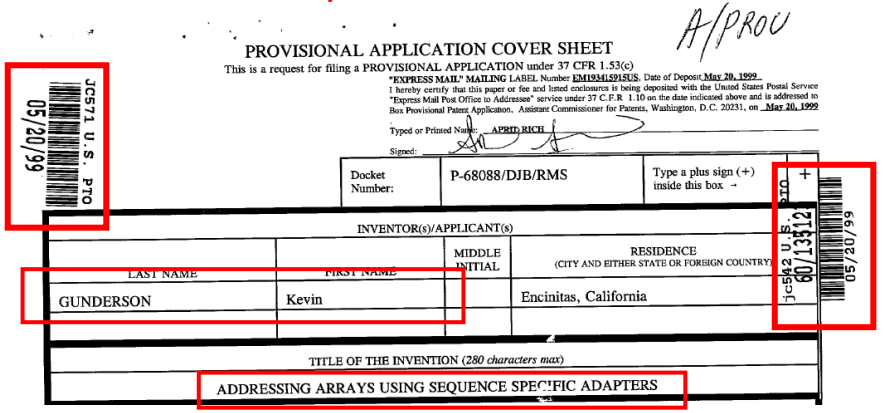


FIG. 3

Illumina, Filed May 20, 1999



The present invention is directed to the use of "universal bead arrays" and methods thereof. A type of addressable array has been previously described, see WO 97/31256, hereby incorporated by reference in its entirety.

Barany Lab Patent

Generally, as is outlined in the attached invention disclosure, adapters can be made for bead arrays such are generally described in U.S.S.N.s 09/189,543; 08/944,850; 09/033,462; 09/287,573; 09/151,877; 09/187,289 and 09/256,943; and PCT applications US98/09163; US98/21193; US99/04473 and US98/05025, all of which are hereby incorporated by reference.

The microsphere array comprises subpopulations of microspheres that comprise capture probes that will hybridize to the adapter probes. The adapter probes generally comprise at least two parts; a first part that is substantially complementary to the capture probe on the bead, and a second part that is substantially complementary to a target sequence (although sandwich assays may also be done). Samples comprising target sequences are then added to the bead array, and detection proceeds via detection of a label directly or indirectly attached to the target sequence as an indication of the presence, absence or amount of the target sequence.

The methods of the invention find particular use in genotyping assays, i.e. the detection of particular nucleotides at specific positions.

Illumina substitutes the terms "adapter sequence" and "capture probe" instead of Barany Lab's term "zip code".

Illumina submits provisional application 60/135,123 (K. Gunderson) on May 20, 1999: Claims to be first to describe "Universal adapters", which is just another term for Barany LAb's "zip codes".

A/PROV

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

"EXPRESS MAIL" MAILING LABEL Number EM193415915US. Date of Deposit May 20, 1999

I hereby certify that this paper or fee and listed enclosures is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box Provisional Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231, on May 20, 1999

Typed or Printed Name: APRIL RICE

Signed:

JC571 U.S. PTO 05/20/99

Docket Number:	P-68088/DJB/RMS	Type a plus sign (+) inside this box → +

JC542 U.S. PTO 60/135123 05/20/99

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
GUNDERSON	Kevin		Encinitas, California

TITLE OF THE INVENTION (280 characters max)

ADDRESSING ARRAYS USING SEQUENCE SPECIFIC ADAPTERS

11. **What are the immediate research plans or steps to be taken:**

Research is ongoing. A set of 16 adapters is in use; set of 100 adapters will be prepared & tested in near future (Kevin Gunderson).

13. **Earliest date and place invention was conceived, and substance of conception (identify people and records to support date and place, such as notebook numbers and pages):**

Ask Kevin Gunderson re adapters & applications. "Universal adapters" first described here.

14. **Name, title, signature, and address of each person who made an intellectual contribution to the invention described in this disclosure:**

Kevin Gunderson 5/12/99

Mark Chee 5/12/99
155 15th Street, #24
Del Mar, CA 92014

John Stuelpnagel

7. **Are there any publications, abstracts, submitted manuscripts, talks, etc. on this work (either already done or in the works)? Please provide details and dates:**

None. But see related invention disclosure by Kevin Gunderson on use of oligonucleotide ligation with arrays.

8. **Compare new and important features of the invention with the prior art, explaining why and how the invention is better:**

With this invention, a master set of beads and decoder oligonucleotides can be reused many times. In the prior art, each assay would require the design and synthesis of a new set of capture sequences and decoders.

9. **Please list known competitors or alternate technologies which solve the same problem:**

None known.

10. **Are there commercial products you envision? Please describe:**

Libraries of beads comprising nucleic acid sequences; cognate libraries of labeled decoder nucleic acids; adapter sequences in electronic form; software for operating on adapter sequences (e.g. to design oligonucleotides incorporating adapter sequences); adapter sequences in physical form (e.g. "universal" adapters that can be attached to molecules of interest.

15. **Name and signature of two witnesses who are not inventors who understand the technical aspects of this invention:**

Name Anthony W. Czarnik
Title CSO
Address 1101 Camino Arroyo
San Diego, CA

Todd A. Dickinson 5/12/99
3435 Lebon Dr. 1133
San Diego CA
92122

Illumina attempts to re-patent Barany Lab's universal arrays and zip code primers, shamelessly claiming ""Universal adapters" first described here".

4 years later, Illumina attempt to re-patent Barany Lab's '917 family patents.

60/180810 Provisional application filed by Jian-Bing Fan of Illumina, February 7, 2000

In order to make use of the array, the identity of the beads at each location must be determined. Illumina, Inc. has developed proprietary methods of rapidly and efficiently decoding an entire array which may contain up to 2,000 unique probe sequences. In order to make the technology available to researchers worldwide, Illumina has recently partnered with PE Biosystems to commercialize the BeadArray™ technology in combination with PE Biosystems' ZipCode™ chemistry. As shown in Figure 2, hybridization to the bead arrays is straightforward and has been demonstrated in Dr. Walt's laboratory at Tufts University and at Illumina.

The addressable array strategy:

To avoid the disadvantage of designing new arrays for each new set of targets, an alternative approach is to provide an array that is "universal" and can be used for any set of biological targets. This allows an investigator to use the same array for different target sequences, which removes the need for costly specialized designs. A universal array of this type has been described (Gerry et al., 1999). Such an array consists of a set of artificially generated probes that are sufficiently long and unique to hybridize with high specificity. These probes act as addresses or "zip-codes" on an array. In order to make use of the array, target sequences to be analyzed are linked with specific zip-code sequences (after PCR amplification, the complementary zip-code sequence will be used to hybridize to the zip-code probes on the array). Thus, any set of targets can be analyzed using the same set of zip-codes by attaching them to appropriate targets. This approach has been used for point mutation and SNP analysis (Gerry et al., 1999; Fan et al., 1999).

Fan, J-B. et al., (1999). Scaleable parallel genotyping using high-density oligonucleotide tag arrays. Manuscript in preparation.

Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G., & Barany, F. (1999). Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol. Biol.* **292**:251-262.

Illumina admits – 4 years after Barany Lab's zip code patent was filed – that "in order to make use of the array, the identity of the bead at each location must be determined". Illumina admits that they use "PE Biosystems' ZipCode™ chemistry." Fu, who wrote the grant application admits the zip-code arrays were invented by Gerry-Barany, 1999, while Fan puts in a fake reference to claim the same year.

Illumina: Fan SBIR 1R43CA097851-01 Grant, submitted November 30, 2001

Form Approved Through 05/2004

790807 Services

PI: FAN, JIAN-BING
1 R44 CA097851-01
Dual: HG,RR
IRG: ZRG1 SSS-Y(10) B

Council: 05/2002

Received: 12/03/2001

Follow instructions carefully
Do not exceed 56-character length restrictions, including spaces

1. TITLE OF PROJECT
High-Throughput Methylation Profiling System

2. RESPONSE TO SPECIFIC REQUEST FOR APPLICATIONS OR PROGRAM ANNOUNCEMENT OR SOLICITATION ☐ NO ☒ YES
(If "Yes," state number and title)
Number: PHS 2001-2 Title: Phase I FAST-TRACK

3. PRINCIPAL INVESTIGATOR/PROGRAM DIRECTOR	New Investigator ☐ No ☒ Yes
3a. NAME (Last, first, middle) Fan, Jian-Bing	3b. DEGREE(S) Ph.D
3c. POSITION TITLE Director, Genetic Analysis	3d. MAILING ADDRESS (Street, city, state, zip code) Illumina, Inc. 9885 Towne Centre Drive San Diego, CA 92121-1975
3e. DEPARTMENT, SERVICE, LABORATORY, OR EQUIVALENT Genomics	
3f. MAJOR SUBDIVISION Molecular Biology	

RESEARCH PLAN

A. SPECIFIC AIMS

This proposal aims to develop a robust and ultra high-throughput technology for simultaneously measuring methylation at many specific sites in a genome. The technology will be based on a bead-based array platform for genetic analysis. We have already used this platform to create a large scale single nucleotide polymorphism (SNP) genotyping system capable of generating more than a million genotypes per instrument per day. By adapting this system to detect and analyze methylation, we aim to provide a tool that will enable genome-wide methylation profiling in large populations. Methylation is involved in gene regulation and altered methylation patterns have been associated with various types of cancers (Baylin et al., 2001; Momparler and Bovenzi, 2000; Warnecke and Bestor, 2000). Therefore, the technology developed in this proposal will provide a powerful tool not only for fundamental genomic research, but also for cancer biology studies, with potential application to cancer classification and diagnosis, and anti-cancer drug target identification and drug screening.

Through a Zirvi FOIA request, the inventors learn that Illumina's J.B. Fan submitted a grant application to the NIH on November 30, 2001. In this funded application, Illumina boasts the ability to generate more than a million genotypes per instrument per day, but fraudulently conceal that they use the Barany Lab IP and trade secrets.

Illumina: Fan SBIR 1R43CA097851-01 Grant, submitted November 30, 2001

C.4.1 Genotyping Results

We are currently routinely multiplexing 96 SNPs per reaction (Fig. 7) and have also obtained good results with 384 multiplexes. DNA consumption is ~ 1 ng per genotype (100 ng in a 96-plex reaction), which is very efficient.

Principal Investigator Fan, Jian-Bing

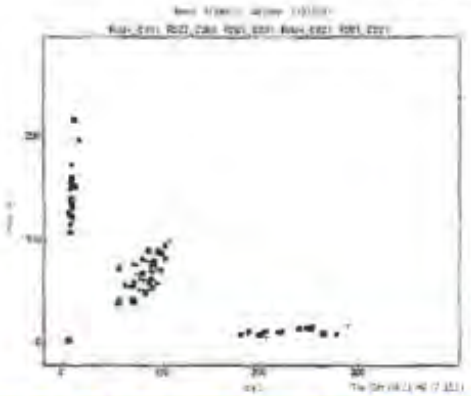


Figure 7. Example of a SNP assay from a 96-plex analysis on optical fiber bundle arrays in a 96-array matrix format. An additional 95 SNPs (not shown) were genotyped in the same reaction. The genotypes shown are for 96 samples processed in parallel in an array matrix. The intensity of allele "a" (labeled with Cy3) is plotted on the x-axis, and allele "b" (labeled with Cy5) is plotted on the y-axis. Each data point represents data from multiple beads in the array. As can be seen from the figure, all three genotypes are represented. The low intensity point near the origin is from a bad DNA sample. These results were generated using a high-throughput automated system in which robotic pipettors are used to carry out the assay procedures.

The system is accurate (~99%) and reproducible (~99%) at 96-plex, and currently generates more than half a million genotypes per day using a small number of array matrices.

Through a Zirvi FOIA request, the true inventors learn that Illumina's J.B. Fan grant application to the NIH on November 30, 2001, which literally infringes upon Barany Lab's '917, '470, and '293 patents using direct LDR-PCR with zip code array capture, shows "accurate" and "reproducible" results with multiplexing 96 SNPs per reaction, and claims good results with a 384 multiplex. This newly revealed information shows that contrary to Illumina's assertions, LDR-PCR works just fine without the need to do a "gap-LCR" step, and that the two are functionally equivalent.

Illumina: Fan SBIR 1R43CA097851-01 Grant, submitted November 30, 2001

Reading out methylation status. We will implement a methylation detection method using the SNP genotyping strategy shown above (Fig. 6). Non-methylated cytosines (C) will be converted to uracil (U) when treated with bisulphate. Uracil's hybridization behavior is similar to that of thymine (T). The detection of the methylation status of a particular cytosine can thus be carried out using a genotyping assay for a C/T polymorphism. Oligo-ligation assay probes can be designed such that one "ASO" is targeted to the "C" allele, i.e. the methylated one, while the other "ASO" is targeted to the "T", i.e. the non-methylated one (Fig. 6). The OLA-PCR assay procedures shown in Fig. 6 will be applied thereafter to determine the methylation status at the targeted sites. Each specific methylation site is interrogated with three oligos: C allele-specific, T allele-specific, and a downstream locus-specific oligo with an address sequence that uniquely corresponds to one specific bead type on the array. Standard protocols developed for SNP genotyping, such as the protocols for oligo annealing and ligation, PCR amplification, and array hybridization, can all be adapted for the methylation detection.

The experimental scheme described in this proposal is advantageous over other methylation detection methods in additional respects. (1) It allows oligos to hybridize directly with their genomic target regions, thereby omitting individual target-specific amplification. (2) The use of universal primers in PCR reaction reduces biased signal amplification, increasing the ability to multiplex the assay robustly and to provide quantitative measurements. The ability to amplify robustly also provides the potential to detect methylation with small numbers of cells, although that is not an aspect we plan to investigate in Phase I.

In addition, Illumina has partnered with Applied Biosystems to commercialize bead array systems for genotyping. Applied Biosystems will be responsible for worldwide sales, marketing and support of the collaboration systems, which are expected to be available in 2002.

Through a Zirvi FOIA request, the true inventors learn that Illumina's J.B. Fan grant application to the NIH on November 30, 2001, which literally infringes upon Barany Lab's '917, '470, and '293 patents using direct LDR-PCR with zip code array capture, was funded to do work invented by the Barany Lab. Illlumina understood that Barany Lab's technology was "advantageous over other methylation detection methods" and cites the exact advantages that were articulated in Barany Lab's IP and trade secrets. Further, the grant highlights the Illumina and Applied Biosystems partnership, implying that Applied Biosystems was fully aware of this work, yet deliberately withheld Illumina's infringement to defraud the true inventors.

Illumina: Chee SBIR 1U54HG002753-01 Grant, submitted May 28, 2002

Form Approved Through 05/2004

PI: CHEE, MARK
1 U01 HG002753-01
Dual:
IRG: ZHG1 SRC(99)

Council: 08/2002

Received: 05/29/2002

Department of Health and Human Services
P 1 5 0 6 4 alth Services
Application

Do not exceed 56-character length restrictions, including spaces.

1. TITLE OF PROJECT
Highly Parallel SNP Genotyping for a Haplotype Map

2. RESPONSE TO SPECIFIC REQUEST FOR APPLICATIONS OR PROGRAM ANNOUNCEMENT OR SOLICITATION ☐ NO ☒ YES
(If "Yes," state number and title)
Number: HG-02-005 Title: Large-Scale Genotyping for the Haplotype Map of the Human Genome

3. PRINCIPAL INVESTIGATOR/PROGRAM DIRECTOR		New Investigator ☒ No ☐ Yes
3a. NAME (Last, first, middle) Chee, Mark Stephen		3b. DEGREE(S) PhD
3c. POSITION TITLE Vice President, Genomics		3d. MAILING ADDRESS (Street, city, state, zip code) 9885 Towne Centre Drive San Diego CA 92121-1975
3e. DEPARTMENT, SERVICE, LABORATORY, OR EQUIVALENT		
3f. MAJOR SUBDIVISION Molecular Biology		
3g. TELEPHONE AND FAX (Area code, number and extension) TEL: (858) 202-4503	FAX: (858) 202-4680	E-MAIL ADDRESS: mchee@illumina.com

4. HUMAN SUBJECTS RESEARCH	4a. Research Exempt ☐ No ☒ Yes If "Yes," Exemption No. 4		5. VERTEBRATE ANIMALS ☒ No ☐ Yes	
☐ No ☒ Yes	4b. Human Subjects Assurance No.	4c. NIH-defined Phase III Clinical Trial ☐ No ☐ Yes	5a. If "Yes," IACUC approval Date	5b. Animal welfare assurance no

6. DATES OF PROPOSED PERIOD OF SUPPORT (month, day, year—MM/DD/YY)		7. COSTS REQUESTED FOR INITIAL BUDGET PERIOD		8. COSTS REQUESTED FOR PROPOSED PERIOD OF SUPPORT	
From	Through	7a. Direct Costs ($)	7b. Total Costs ($)	8a. Direct Costs ($)	8b. Total Costs ($)
9/20/2002	9/19/2004	$7,072,747	$7,310,685	$14,879,437	$15,355,660

Through a Zirvi FOIA request, the true inventors learn that Illumina's Mark Chee fraudulently submitted a grant application to the NIH on May 28th, 2002, which was subsequently funded for $15,355,660. Thus, the true inventors learn for the first time that Illumina was paid over $15 million by the NIH to develop highly parallel SNP genotyping based on Barany Lab's LDR-PCR technology as covered in Barany Lab's '917 patent filed Feb 9, 1996, Barany Lab's '470 patent filed May 29, 1996, and Barany Lab's '293 patent filed Jan 6, 1999.

Illumina: Chee SBIR 1U54HG002753-01 Grant, submitted May 28, 2002

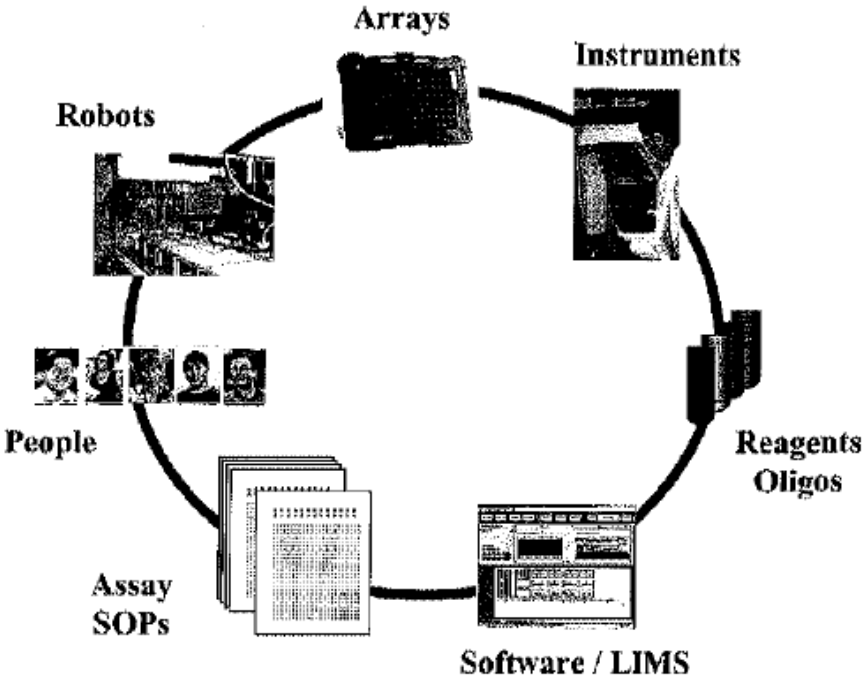


Figure 1. Informatically integrated high-throughput genotyping system based on BeadArray technology. The system is modular, and integrated using barcode reading and LIMS supervision. Miniaturization is achieved by using a fiber-optic bundle as a substrate for the highest-density microarray available today. Ninety-six of these arrays are held together in a matrix (Array of Arrays™ matrix) that matches the spacing of a standard 96-well microplate. In order to use efficiently the high capacity of the platform, the SNP assays developed at Illumina have been designed for a high level of multiplexing. Currently, Illumina's production genotyping system routinely multiplexes 288 SNP assays in each well of a microplate, and we have recently achieved excellent results with 1,152-plex assays.

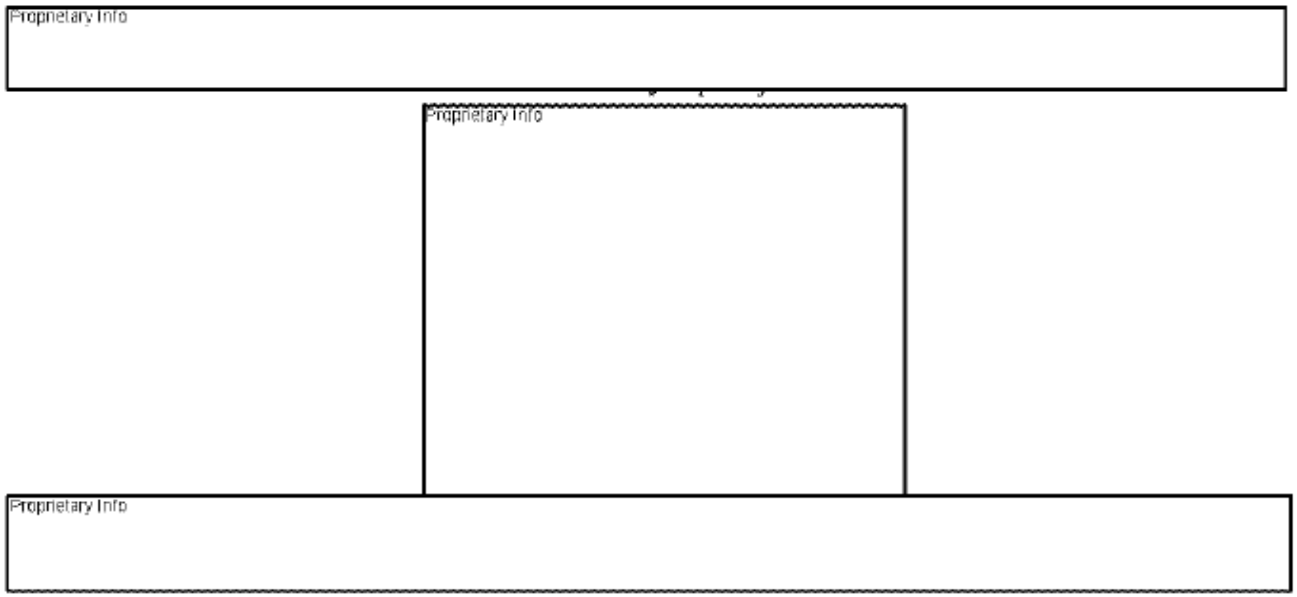
Through a Zirvi FOIA request, the true inventors learn that Illumina's Mark Chee fraudulently submitted a grant application to the NIH on May 28th, 2002, that is completely based on Barany Lab's LDR-PCR technology as covered in Barany Lab's '917 patent filed Feb 9, 1996, Barany Lab's '470 patent filed May 29, 1996, and Barany Lab's '293 patent filed Jan 6, 1999. The entire integrated system is completely dependent on Barany Lab's zip code arrays and LDR-PCR reactions, without these critical components, the system would not exist.

Illumina: Chee SBIR 1U54HG002753-01 Grant, submitted May 28, 2002

C.4 Multiplexing

By using objective measures of quality described above, we have optimized assay procedures to increase multiplexing levels from 48-plex to 1,152-plex within a period of about 12 months (Fig. 3). Since

Principal Investigator **Chee, Mark S.**



Through a Zirvi FOIA request, the true inventors learn that Illumina's Mark Chee fraudulently submitted a grant application to the NIH on May 28th, 2002, that is completely based on Barany Lab's LDR-PCR technology as covered in Barany Lab's '917 patent filed Feb 9, 1996, Barany Lab's '470 patent filed May 29, 1996, and Barany Lab's '293 patent filed Jan 6, 1999. The grant boasts 1,152-plex LDR-PCR with zip code capture reactions, however Illumina chose to redact Figure 3 and supporting information. Since the material is over 15 years old, it is difficult to understand what would be proprietary, or suitable for a patent submission. Discovery of an un-redacted version of this application would reveal if Illumina is hiding information to deprive the true inventors of rightful royalties.

EXHIBIT 9

The Affidavit of Dr. Francis Barany

1. My name is Dr. Francis Barany. I make this affidavit based on my own personal knowledge.

2. I am a Professor of Microbiology at Weill Cornell Medicine, I am best known for inventing the ligase chain reaction (LCR), ligase detection reaction (LDR), and Universal DNA arrays currently used worldwide to diagnose genetic diseases, detect infectious pathogens, and identify cancer mutations and other diseases using DNA microarrays and targeted next-generation sequencing (NGS). As of this writing, I hold 74 (US) and over 100 (international) patents widely used in molecular diagnostic and sequencing.

3. I am a National Academy of Inventors fellow (2016). I was also named to the 2004 "Scientific American 50" — Scientific American magazine's prestigious annual list recognizing leadership in science and technology from the past year. I was included as "Research Leader in Medical Diagnostics," for my leadership role in the development of a universal array genomic chip that allows for the rapid and accurate detection of cancers and other diseases — especially breast and colon cancer. The Universal Zip Code DNA array, unlike previous mutation-detection methods of its kind, is the first programmable DNA array. It does not require a redesign each time a new gene is discovered. The universal array's unique design may use ligation primers that guide a fluorescence-labeled signals to array "addresses" that indicate disease. This Universal Zip Code DNA array and other technologies developed in my laboratory were licensed to Applied Biosystems.

4. Dr. Monib Zirvi, MD/PhD, graduated with a PhD from my laboratory at Weill Cornell Medicine in January of 1999. He joined my laboratory in August of 1994. Dr. Zirvi is a co-inventor of the “Universal Zip Code DNA array” and is a co-inventor of 20 issued US patents with me. Furthermore, at my request after he had graduated from my lab, Dr. Zirvi designed sets of Zip Code DNA sequences for use in creating Universal DNA Arrays. Using his expertise in Computer Engineering and programming, Dr. Zirvi designed and selected both a 4633 Zip Code set and a 465 Zip Code set for creation of large Universal Zip Code Arrays. He did this on his time and using his own equipment and I have retained copies of that work since 1999 which he helped me label on my computer. As such he is the primary co-author of those sets and a co-inventor of Universal Zip Code DNA arrays as described in international patent application WO97/31256.

5. These Zip Code sets in electronic files were shared in confidence with Applied Biosystems and somehow wound up in the hands of Illumina. In August 2015, Dr. Zirvi and I were shocked to discover that Illumina had submitted a patent application to the USPTO falsely declaring that they had invented Illumacodes (one of many pseudonyms they use for Zip Code sequences to obfuscate the true origin of this intellectual property). In fact, Illumacodes 1 through 16 were verbatim copies of 16 Zip Code sequences from one of Dr. Zirvi’s 465 Zip Code *unpublished* files. Further, Dr. Zirvi found that Illumina used the word “ZipCode” multiple times in their DMAP software to decode all Illumina arrays.

6. Dr. Zirvi diligently attempted to help Cornell and ThermoFisher (which now owns Applied Biosystems) in *Cornell v Illumina* (1:10-cv-00433-LPS) case to obtain rightful royalties and recognition as inventors of the Zip Code technology from Illumina for this infringement of our intellectual property. Throughout this litigation, the academic inventors were told that attorneys Matthew Pearson of Akin Gump and Roger Chin and Douglas Lumish of Latham and Watkins represented the inventors’ interests, and that all interests were aligned.

7. However, while the *Cornell v Illumina* case was taking place, Illumina and ThermoFisher were secretly working together on a product called “Ampliseq for Illumina” while pretending to be adversaries in the case. On information and belief this collaboration was initiated at least as early as January 2015, prior even to the Markman hearing, and did not become publicly known until January of 2018. This fraud was kept hidden from Cornell, which filed a Rule 60(b) motion against both Illumina and ThermoFisher.

8. Illumina has used Zip Code technology to obtain a dominant position in genomics and DNA sequencing. Illumina has employed a strategy of "efficient infringement" against academic inventors by refusing to pay royalties and filing IPRs and ex partes to overwhelm intellectual property right holders with legal costs. This strategy is used as a sword and shield to avoid "third party intellectual property rights and other encumbrances," as Illumina stated in the unredacted First Amendment Agreement obtained from the SEC.

9. Tellingly, during and after the time that Illumina and ThermoFisher were colluding to defraud Cornell in the *Cornell v Illumina* case, both companies were working with security forces of the Chinese Communist government to differentiate Uyghurs from Han Chinese, as part of their apartheid efforts to subjugate and detain the Muslim Uyghur minorities in modern day concentration camps. (See: https://www.justsecurity.org/64605/scientists-are-aiding-apartheid-in-china/). As a son of two survivors of Nazi concentration camps, this begs the question: Can these companies be trusted with huge databases of confidential DNA information from US Citizens?

10. Now, Illumina/Grail is attempting to use their monopoly position in NGS and lobbying efforts to influence the US Congress and Senate to allow them to obtain control of the nascent, and yet unproven, Multi-Cancer Early Detection liquid biopsy technology (MCED). These tests claim to identify early cancers from a blood sample, yet there is currently no evidence of clinical utility, in other words, no evidence that they will save any lives. Briefly, the reasons are that the Illumina/Grail MCED assay:

i. Misses most of the truly early cancers,
ii. Some cancers would be detected anyway due to symptoms, and
iii. Most cancers they do detect are late-stage cancers where the patient eventually dies, so the patient doesn't benefit from the "early" detection.

11. Illumina is trying to force its acquisition of Grail, despite a standstill order by antitrust regulators. Illumina is attempting to extract ~$60 billion annually from US taxpayers, or about 7% of the current Medicare budget. This is despite the risk of adverse outcomes from unnecessary invasive procedures from the up to 60% false positive results from Illumina/Grail tests. In the interest of the future of academic research and, more importantly, the economic viability and future of the national healthcare system, I am compelled to file this affidavit in support of Dr. Zirvi's legal action.

Respectfully submitted on February 6, 2023:

Dr. Francis Barany

Professor Francis Barany
National Academy of Inventors Fellow
Dept. of Microbiology & Immunology
Weill Cornell Medicine

Tel 212.746.6509 or 6507 or 6524
barany@med.cornell.edu

Mobile 917.957.3976

State of Texas

County of Bell

This instrument was acknowledged before me by means of an interactive two-way audio and video communication on 02/06/2023 by Dr. Francis Barany.

NOTARY PUBLIC STATE OF TEXAS

Jessica Hammett

ID NUMBER
13354991-7
COMMISSION EXPIRES
January 24, 2026

Jessica Hammett

Notarized online using audio-video communication

EXHIBIT 10

02-09-00

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c)

"EXPRESS MAIL" MAILING LABEL Number EL406567991US, Date of Deposit February 7, 1999
I hereby certify that this paper or fee and listed enclosures is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box Provisional Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231, on February 7, 1999.
Typed or Printed Name: HAMMID SANCHEZ
Signed:

Docket Number:	P-68929/DJB/RMS/DCF	Type a plus sign (+) inside this box →

JC553 U.S. PTO 60/170810 02/07/00

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
FAN	Jian-Bing		San Diego, California

TITLE OF THE INVENTION (*280 characters max*)

GENE EXPRESSION PROFILING

CORRESPONDENCE ADDRESS

ROBIN M. SILVA
FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP, 4 Embarcadero Center, Suite 3400, San Francisco

STATE	CA	ZIP CODE	94111	COUNTRY	US

ENCLOSED APPLICATION PARTS (*check all that apply*)

X	Specification	*Number of Pages*	44		Small Entity Statement
X	Drawings	*Number of Sheets*	5		Other (*specify*):

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (*check one*)

X	A check (No. 28307) or money order is enclosed to cover the Provisional filing fees		PROVISIONAL FILING FEE AMOUNT $150.00	$150
	The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number:	06-1300 (Order No.P-68929/DJB/RMS)		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No.

☐ Yes, the name of the U.S. Government Agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE: Robin M. Silva Date February 7, 2000

TYPED or PRINTED NAME ROBIN M. SILVA REGISTRATION NO. (*if appropriate*) 38,304

☐ Additional inventors are being named on separately numbered sheets attached hereto.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Burden Hour Statement: This form is estimated to take .2 hours to complete. Times will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Assistant Commissioner for Patents, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

INVENTION DISCLOSURE FORM

This form is provided to permit evaluation of the patent potential of company inventions, and to facilitate preparation of patent applications when warranted. Please fill in each space as completely as possible, and use additional sheets when necessary.

1. Name: **Jian-Bing Fan**

2. Date:

3. State the Title of the Invention:

Genome-wide gene expression profiling, alternative splicing monitoring and genotyping with direct template annealing, oligo-ligation and general amplification.

4. Describe the invention: Use additional sheets if necessary. Attach descriptive materials such as drawings, sketches, photographs, etc. which may help illustrate the invention. Delineate new and important features. Make sure to include both the preferred embodiment as presently identified, and alternative constructions, procedures or equivalent components which can accomplish the same result as the preferred embodiment.

We describe here a very flexible approach for gene expression profiling, alternative splicing monitoring and genotyping, without any prior amplification of specific targets (please see attached experimental schemes).

5. State the primary purpose of the invention, including the need satisfied or problem solved by the invention:

The invention provides a powerful approach for genome-wide gene expression profiling, alternative splicing monitoring and genotyping.

6. Please list what you feel is the prior art: please include references, articles, talks, abstracts, patents, etc. which are relevant to either the state of the prior art or to the invention. Please include dates and provide copies whenever possible:

PCR.
OLA genotyping.

7. Are there any publications, abstracts, submitted manuscripts, talks, etc. on this work (either already done or in the works)? Please provide details and dates:

None.

8. Compare new and important features of the invention with the prior art, explaining why and how the invention is better:

The novelty of this invention combines the flexibility derived from the bar-coding approach, the specificity offered by the oligo-ligation assay, and the intrinsic high sensitivity coming from the general amplification procedure. It provides a powerful approach for genome-wide gene expression profiling, alternative splicing monitoring and genotyping, without any specific prior amplification of targeted genomic regions.

9. Please list known competitors or alternate technologies which solve the same problem:

None.

10. Are there commercial products you envision? Please describe:

No.

11. What are the immediate research plans or steps to be taken:

We are testing the sensitivity, accuracy and the quantitative performance of this method, in collaboration with our collaborators at UCSD. We have submitted a grant proposal to National Cancer Institute in

12. What are the longer term research plans or steps to be taken:

We will implement this procedure in our gene expression profiling project, cancer molecular classification project, and genotyping projects.

13. Earliest date and place invention was conceived, and substance of conception (identify people and records to support date and place, such as notebook numbers and pages):

Invented in at Illumina and UCSD (grant proposal preparation).

Please also see related Illumina invention disclosures filed previously.

14. Name, title, signature, and address of each person who made an intellectual contribution to the invention described in this disclosure:

Introduction:

The strategy is designed to provide a very sensitive and accurate approach for genome-wide gene expression profiling, alternative splicing monitoring and genotyping, without any target-specific amplification.

Part I:

1. As shown in Figure 1, oligos are synthesized such that each will consist of 4 parts (from 5′ to 3′; please note the direction drawn is from right to left):
   - An upstream universal priming site (U),
   - An unique zip-code sequence (Zip) to target the oligo to a specific address on the array,
   - A sequence complementary to a gene-specific exon sequence (EX, in Figure 1a), or a specific splice junction (SJ, in Figure 1b), and
   - A downstream universal priming site (D).

2. Many different oligos, corresponding to many different genes or splicing junctions can be pooled for specific applications. Pooled oligos will be annealed to total RNA ~~or~~ poly(A)$^+$ mRNA from cells or tissues.

3. Oligos annealed to mRNAs will be PCR-amplified using the two universal primers.
   - One of the primers can be chemically or fluorescently labeled, for example, biotinylated.
   - The PCR products can be labeled after PCR amplification.
   - The number of PCR cycles will be optimized to obtain sufficient products and at the same time maintain the original ratio of individual oligos.

4. The PCR products will be hybridized to an array and the hybridization signals will be detected and quantified.
   - The array can be a universal array corresponding to the Zip-code used in the assay.
   - The array can also be a customized array corresponding to the specific targets tested in the assay (i.e. the specific genes or gene splicing junctions).

5. Individual gene expression level or splicing events will then be detected and quantified by analyzing the array data.

- Samples from different cells or cell stages can be labeled differently and differences in gene expression or alternative splicing will be revealed by co-hybridization.

6. The experimental scheme described here is advantageous over standard array-based gene expression protocols in several respects: (1) This method has the potential to detect gene expression or alternative splicing events from a single cell or a few cells because of signal amplification of annealed oligos. (2) It allows oligos to hybridize directly with their RNA targets, thereby omitting a cDNA conversion step. (3) The hybridization reaction takes place in solution instead of on a surface so that DNA and RNA hybridize more predictably according to their thermodynamic properties. (4) The poly $(A)^+$ RNA selection step eliminates excess oligos as well as those that are nonspecifically annealed to RNA (in the presence of rRNAs and tRNAs as competitors) or stick to the container surface. As a result, isolated oligos will reflect the level of individual gene expression level or splicing events in cells and the background signal should be significantly reduced. (5) The use of universal primers avoids biased signal amplification in PCR. The major potential pitfall of this approach is nonspecific annealing of oligos. As a result, those oligos will be amplified, which may surpass specific signals resulting from specific annealing. This may be especially problematic with rare mRNA isoforms.

Part II:

Therefore, an alternative approach is proposed in the following:

1. As shown in Figure 2, two oligos as a set are synthesized for each assay, instead of a single long oligo.

- The upstream oligo will consist of three parts (from 5′ to 3′): the upstream universal priming site, a zip-code sequence, and a sequence complementary to a gene-specific exon sequence (Figure 2a), a specific splicing junction sequence (Figure 2b), or a specific genomic region containing a polymorphic site need to be tested (Figure 2c).
- The downstream oligo will also have three parts (from 5′ to 3′): a sequence complementary to the downstream exonic sequence (Figure 2a and 2b) or genomic sequence (Figure 2c) that is immediately adjacent to the region covered by the upstream oligo, a different zip-code sequence, and the downstream universal priming site.
- The relative positions of the target-specific sequences in the upstream and the downstream oligos can be adjusted.
- At least two-allele specific upstream (with the variable base at the 3′-end) or downstream (with the variable base at the 5′-end) oligos will

be designed for the genotyping application. The allele-specific oligos will be specified with specific zip-code. The discriminative ligation with the allele-specific oligos will be used to determine the genotypes in the samples analyzed.

- The downstream oligo will be kinased so that it contains a 5'-phosphate prior to ligation.
- The second zip-code sequence in the downstream oligo can be omitted.
- To further increase the oligo annealing and ligation specificity, thereby increasing the accuracy of the assay, a third oligo can be designed (Figure 2d). In this way, correct oligo annealing at three adjacent postions and two ligation events are required to generate an PCR-amplifiable template. The upstream, downstream and the third oligos can be designed in which the third oligo is placed in between the upstream and the downstream oligos (Figure 2d).
- The third oligo can be as short as an 8-mer or 9-mer, so that a library of all possible third oligos can be pre-generated.

2. The target-specific sets of oligos will be annealed to total RNA or poly (A)$^+$ RNA, or genomic DNA. A ligation reaction is carried out as to link the upstream (and the middle) and the downstream oligos that are bridged together through spliced mRNA or genomic DNA.
   - This reaction can be carried out with T4 DNA ligase or thermostable DNA ligase at higher ligation temperatures.

3. After the ligation step, oligos will be PCR-amplified using two universal primers. All the subsequent procedures will be carried out as described in Part I (2-5) (see above).

4. The oligo-ligation procedure will prevent oligos that are nonspecifically stuck from being amplified, thereby increasing the accuracy of the assay.

5. Furthermore, since each ligated and amplified target will have two zip-code sequences so that two separate probes on the array will be measured for one particular assay target (gene, splicing event or SNP), which will reduce hybridization variation associated with individual zip-code sequences.

Part III:

In conclusion, the experimental procedure described above combines efficiency and accuracy to conduct a parallel, genome-wide analysis of gene expression, alternative splicing and genotyping. The technology is novel in multiple aspects

as described above and will have a broad impact on both basic and clinical research.

Figure 1a: A flow chart for array-based detection of gene expression

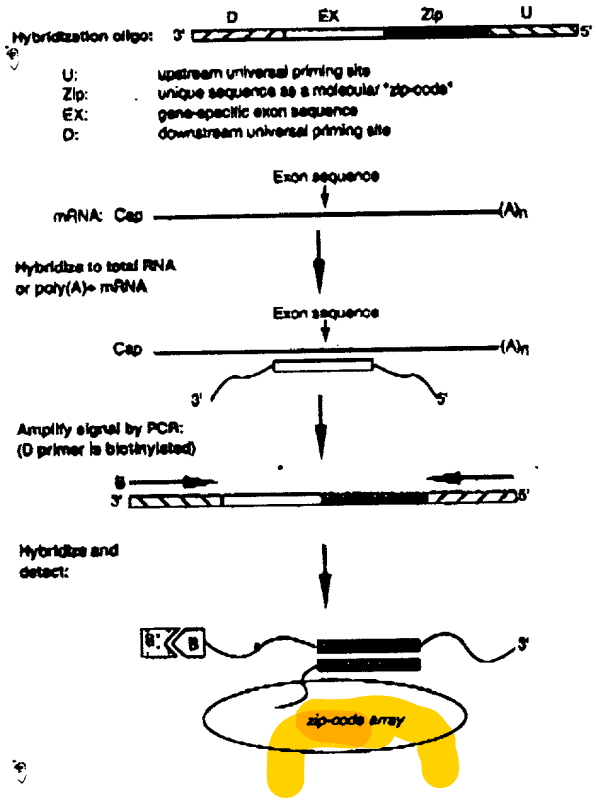


Figure 1b: A flow chart for array-based detection of RNA alternative splicing

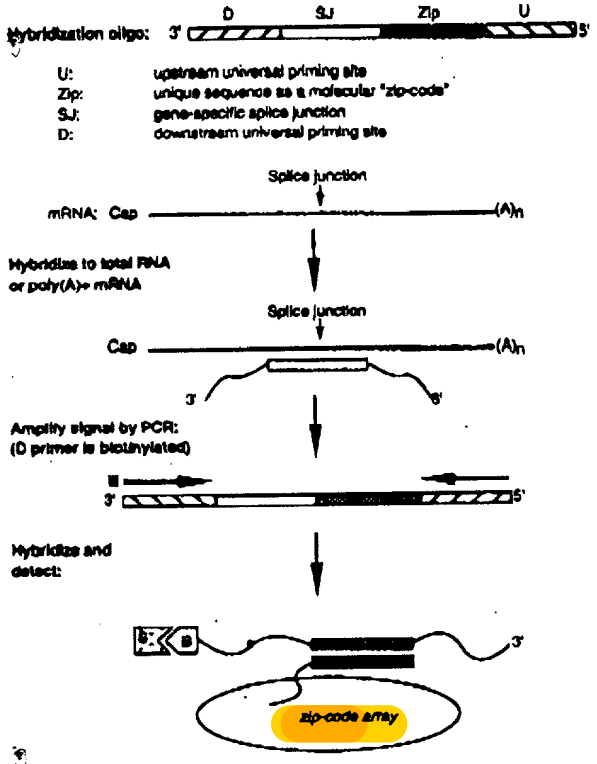


Figure 2a: Genome-wide gene expression profiling using oligo-ligation strategy

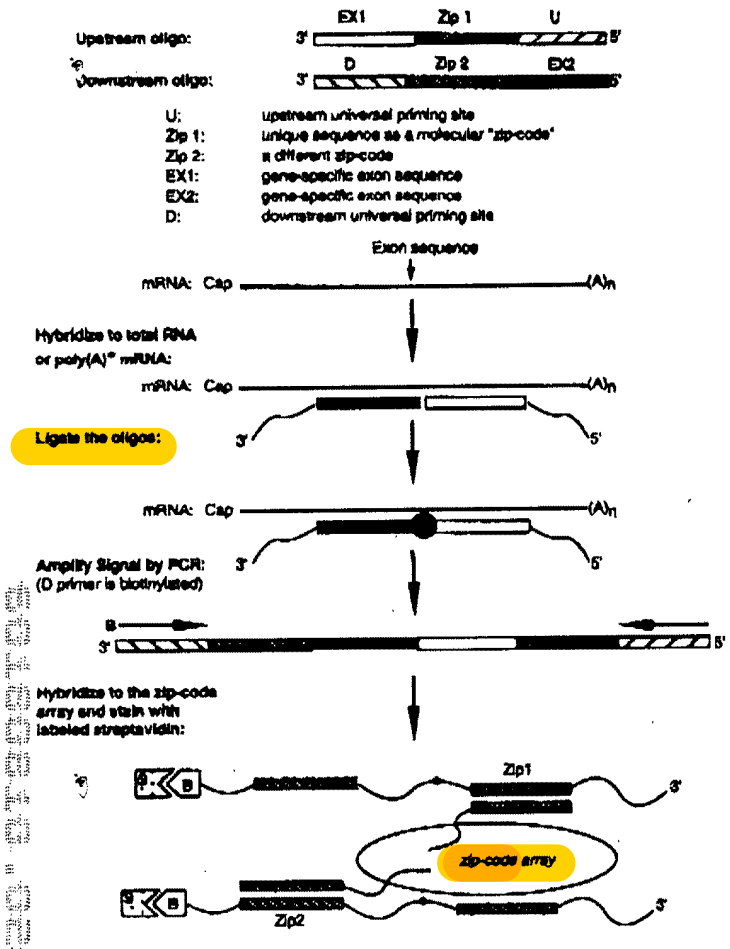


Figure 2b: Genome-wide RNA alternative splicing monitoring using oligo-ligation strategy

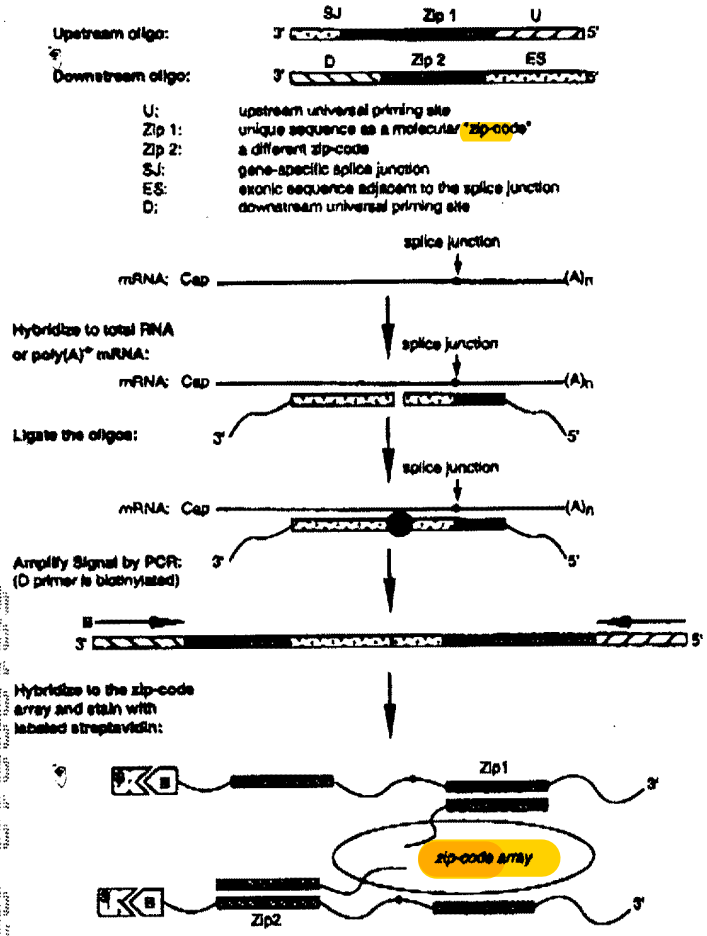


Figure 2c: Direct genotyping using a whole-genome oligo-ligation strategy

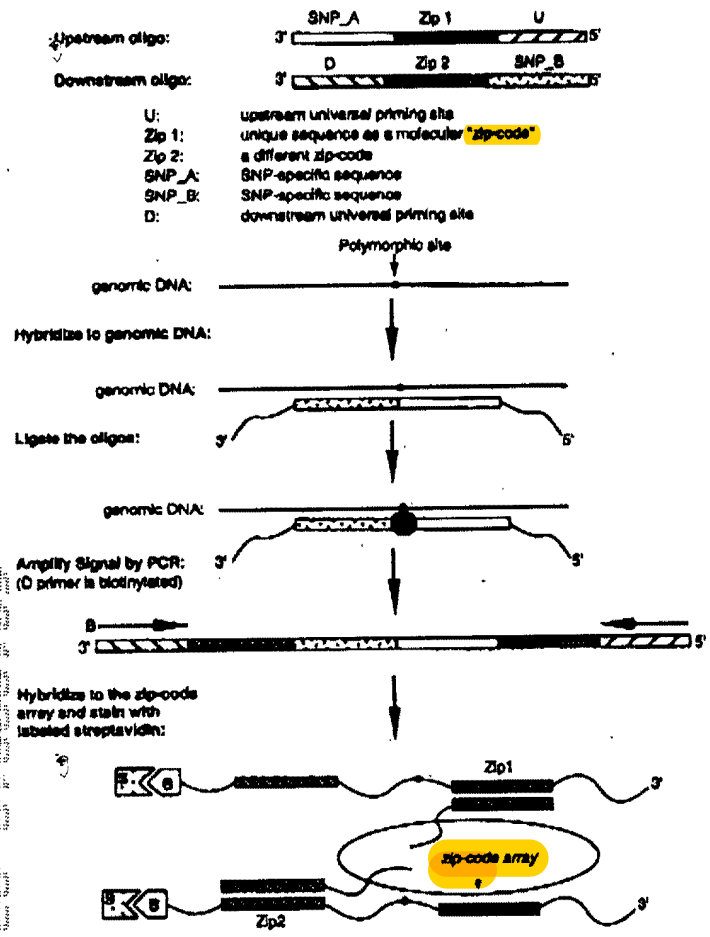
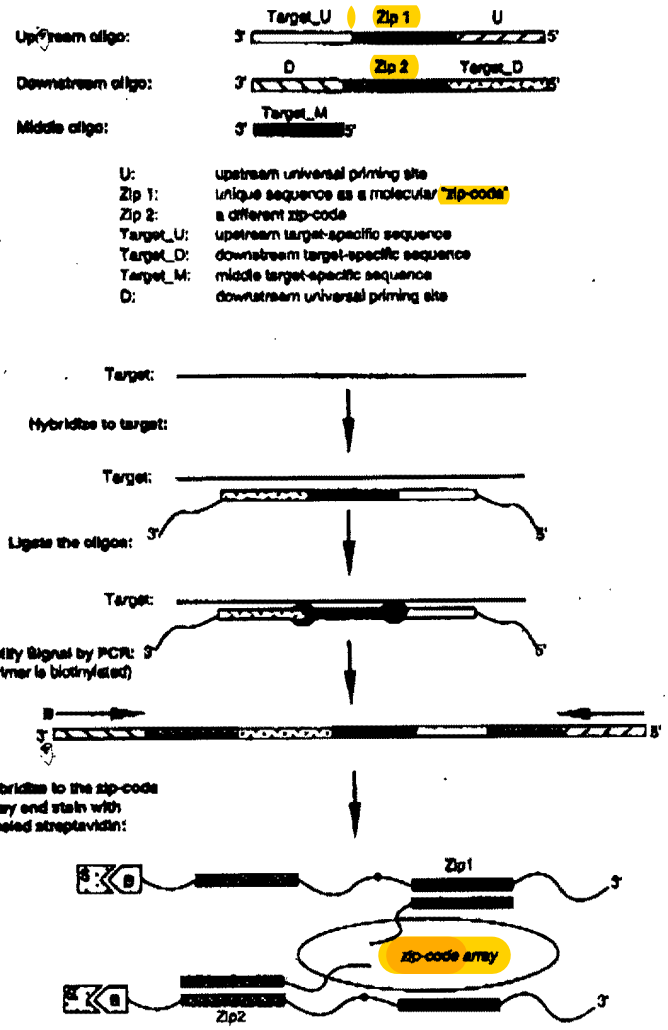


Figure 2d: Whole-genome oligo-ligation strategy



The R21 phase studies

a. Specific Aims

Alternative RNA splicing is widespread in higher eukaryotic cells and plays a vital role in gene expression. However, detection and analysis of alternative splicing currently rely on RNase protection and RT-PCR assays, which are labor intensive, inefficient, and low scale, especially in the era of functional genomics. Here, we propose to develop a parallel assay system for alternative splicing using the latest DNA array technology. Briefly, we plan to use a sequence addressing strategy that will allow multiplexed assays to be analyzed in parallel on an oligonucleotide probe array. We will synthesize DNA oligos so that each contains a 20 nt target sequence complementary to a specific splice junction and a unique 20 nt "zip-code" sequence. Using this zip-code system, large numbers of alternative splicing events can be simultaneously detected with pooled oligos, and individual alternative splicing events can be resolved spatially by hybridization of the assay oligonucleotides to different zip-code probes on a "universal" array. This project will be carried out in close collaboration between my laboratory at UCSD and Illumina, Inc., an array technology company based in San Diego. Our assay explores molecular alterations at the RNA processing level, which is complementary to those based on monitoring gene expression. Because specific alternative splicing events are known to be associated with various types of cancer, the technology developed in this proposal will provide a powerful tool in cancer classification, diagnosis, disease target identification, high throughput drug screening, and mechanistic studies. Our specific aims in this exploratory phase are to:

Aim 1: Establish a set of criteria for selecting splice junction targets.

We plan to develop computational tools based on a series of selection criteria to help the selection of zip-code sequences and couple them with alternative splicing target sequences. These tools will aid in our experimental design, and importantly, they will become integrated components of a systematic large-scale parallel analysis of alternative splicing in the future.

Aim 2: Prove the concept with alternative splicing reporters in transiently transfected cells.

We intend to express two well-characterized alternative splicing reporters, β-thalassemia and E1A, in HeLa and COS cells and assay their splicing patterns by standard RT-PCR techniques and by the proposed DNA array-based assay. This experimental system will allow us to systematically develop optimal experimental conditions and establish appropriate controls for the array-based assay.

Aim 3: Validate the technology on endogenous transcripts in model systems.

To prepare for its application in cancer classification, it is essential to validate the experimental strategy on endogenously expressed genes because the complexity of the system will be greater than that of transiently transfected splicing reporters. We plan to select a group of well-documented alternative splicing events, zip-code their splice junctions, and test their splicing patterns in a few model cell lines. Some known cell-specific alternative splicing events will be included as internal controls in these experiments. In these studies, we will systematically improve the accuracy and efficiency of the assay system.

b. Background and Significance

Genes in higher eukaryotic cells contain introns, which are removed during RNA processing to generate mature functional mRNAs. In most cases, the removal of introns is efficient, and thus these splicing events are constitutive. However, many transcripts are alternatively processed to generate multiple mRNAs from a single mRNA precursor (pre-mRNA) through the use of different 5' or 3' splice sites, exon inclusion or exclusion, and intron retention. The complexity of gene expression is further increased in many cases by coupling alternative splicing with alternative promoters and the use of alternative polyadenylation sites. Based on comparison among expressed sequence tags (ESTs) in databases, it is estimated that as many

as 30% of genes in humans exhibit alternative splicing (Gelfand et al., 1999). Considering that one transcript often gives rise to more than two isoforms, the number of alternatively spliced mRNAs may surpass the total number of genes that are expressed in a higher eukaryotic organism. Because alternatively spliced transcripts may encode protein isoforms that have distinct functions, it becomes a major challenge in functional genomics to relate a biological function not only to the expression of specific genes but also to their isoforms resulting from post-transcriptional processing. This is particularly pertinent to cancer research as molecular alterations during malignancy may result from changes not only in gene expression but also in RNA processing

Importance of Alternative Splicing in Biology and Medicine: Alternative splicing plays a vital role in biology and medicine. Here I describe a few examples where the functional consequences of alternative splicing are well documented.

FGFR2/KGFR: Epithelial cells secrete acidic Fibroblast Growth Factor (aFGF), which binds and activates its receptor FGFR2 on the cell surface of fibroblasts. Conversely, fibroblasts secrete Keratinocyte Growth Factor (KGF), which binds and activates KGFR on epithelial cells. Interestingly, FGFR2 and KGFR are generated from the same pre-mRNA by alternative splicing (Miki et al., 1992). Such cell-specific alternative splicing must be tightly regulated because cells expressing both a growth factor and its specific receptor will be transformed to uncontrolled growth.

Apoptotic regulators: A number of apoptotic regulators such as Bcl-x, Ced-4, and Caspase-2 (Ich-1) have two isoforms generated by alternative splicing (reviewed by Jiang and Wu, 1999). In each case, one form promotes programmed cell death and the other prevents cell death. Thus, alternative splicing provides a life or death choice in determining and regulating the ratio of these isoforms.

CD44: This important cell surface molecule is involved in tissue-specific targeting of T cells, B cells, and macrophages in the immune system as well as in cell adhesion and signal transduction. The transcript has 10 alternative exons, which are included/excluded in combination to generate numerous isoforms. Alterations in CD44 splicing are among the best tumor markers (reviewed by Goodison and Tarin, 1998). CD44 alternative splicing appears to be regulated by cytokines and by oncogenic activation, and the inclusion of a specific exon (v6) was shown to cause tumor metastasis in a model system (Gunthert et al., 1992).

AML1: This transcription factor is required for granulocyte differentiation. The protein contains an N-terminal DNA binding and protein dimerization domain, and a C-terminal transcriptional activation domain. In 20% of acute myelogenous leukemia (AML) patients, the N-terminal sequence of AML1 is fused to sequences from other chromosomes via chromosome translocation. However, in many AML cases, no chromosome translocation is detected., but a change in alternative splicing of AML1 pre-mRNA appears instead. Alternative splicing results in a truncated version of AML1, which was shown to suppress granulocyte differentiation (Tanaka, et al., 1995). Thus, some fraction of AML cases may be triggered by a malfunction in splicing control and regulation.

In conclusion, alternative splicing is associated with important biological events, and in many cases, the pattern or alteration of alternative splicing may be markers for specific diseases and/or targets for disease prevention and intervention.

The splicing problem in the post-genome era: The above examples demonstrate that alternative splicing is widespread and plays a critical role in the regulation of gene expression in development, differentiation and disease. However, the functional significance, if any, of reported alternative splicing in the vast majority of cases remains to be determined. Furthermore, evidence for numerous alternative splicing events is emerging from comparison between cDNA (mostly ESTs) and genomic sequences. However, this large number of

mRNA isoforms may arise in several ways: (1) True isoforms resulting from alternative splicing; (2) Aberrant transcripts due to mutations in splicing signals (e.g. mutations in splice sites). (3) Contamination with intron-containing nuclear pre-mRNA, and (4) Experimental errors in library construction and sequencing. The first two classes should be further investigated, but the latter two classes are artifacts and should be eliminated. The problem is that, while the number of potential mRNA isoforms is overwhelming, we are still relying on conventional methods, such as RNase protection and RT-PCR, to detect and analyze individual splicing events. Therefore, we are faced with the challenge of developing an assay that will allow parallel analysis of a large number of alternative splicing events, and to make biological sense of them.

Current Approaches: In the past decade, microarray technology has revolutionized biomedical research and the utility of the technology has been demonstrated in monitoring gene expression (Lockhart et al., 1996), resequencing (Chee et al., 1996), and genotyping single nucleotide polymorphisms (SNPs) (Wang et al., 1998). Here, I briefly describe the major varieties of microarray technology and discuss how they might be adapted for large scale analysis of alternative splicing.

cDNA microarrays: This technique was pioneered by Pat Brown and co-workers, and basically involves spotting PCR products of individual genes onto solid supports (membrane or glass) to form a microarray (Schena, et al., 1995). Total RNA isolated from cells or tissues is converted to cDNAs flanked by a T7 promoter for in vitro transcription. Fluorescently labeled transcripts are hybridized to the microarray. This simple and powerful technique may be adapted to alternative splicing analysis, by spotting cDNA fragments derived from alternatively spliced regions. However, there are two major problems with this approach. First, not every isoform is associated with a unique sequence. For example, if an intron is included in one isoform and excluded in the other, sequences in the intron can be spotted to detect the intron-inclusion isoform, but there is no specific sequence that can be used to detect the intron-exclusion isoform (because all exonic sequences are common to both isoforms). Secondly, many alternatively spliced mRNAs differ in small regions (e.g. using two splice sites that are a few nucleotides away from each other, inclusion/exclusion of a tiny exon, etc.). In these cases, there would be no specific sequence to spot. Therefore, the cDNA array is not a universal approach to the alternative splicing problem.

Oligonucleotide arrays: Represented by DNA arrays manufactured by Affymetrix. DNA oligos are directly synthesized on a glass surface to produce a high-density microarray (Fodor et al., 1991). The advantage of a high density array is that it allows large redundancy and controls to be built in for each hybridization event. For example, about 20 gene-specific oligos are used to obtain an average expression profile of individual genes, and equal numbers of mismatch oligos (containing a mismatch in the central location of each specific oligo) are used to eliminate background. This technology may be adapted for alternative splicing by selecting oligos that are complementary to alternatively spliced regions. However, this strategy has the same drawback as the cDNA microarray because small alternatively spliced regions will limit the selection of a unique set of oligos.

One idea is to synthesize oligos that are complementary to specific splice junction sequences. However, there are multiple obstacles in pursuing this direction: First, one has to convince Affymetrix to manufacture such a specialized array for alternative splicing. Secondly, standard experimental protocols involve conversion of total RNA to cDNA, during which a T7 promoter sequence is incorporated into each cDNA molecule. This will allow in vitro transcription in the presence of labeled substrates to prepare targets for hybridization on arrays. However, the majority of labeled targets derived from total RNA are not specifically complementary to splice junction sequences and specific signals for splice junctions can be overwhelmed by those derived from constitutive exons. The latter problem may be overcome by adapting the strategy used in the SNP assay (Cargill et al., 1999). Specifically, one can PCR-amplify a set of

alternatively spliced regions and then hybridize the targets to an oligo microarray containing splice junction sequences. The problem is that this approach is almost as slow as RT-PCR analysis, not mentioning that it also requires specialized arrays.

Assembled Arrays: Illumina, Inc. is developing a new type of array technology, called the BeadArray™. There are two key parts to the technology, which was invented by Dr. David Walt and colleagues at Tufts University (Michael et al., 1998). The first is an array of wells patterned into an optical imaging fiber, consisting of a bundle of individual fibers arranged in a hexagonally packed matrix. Each individual fiber has a light-conducting inner core that can be chemically etched at a different rate from its surrounding cladding. As shown in Figure 1, by treating the polished end of an optical fiber with acid, an array of microwells is generated, with geometry and dimensions of the array determined by the physical specifications of the optical fiber. In the second part, the individual wells in the etched fiber can be filled with beads slightly smaller in diameter than the wells. A suspension of beads is deposited on the array, and individual spheres randomly settle into each well as liquid evaporates.

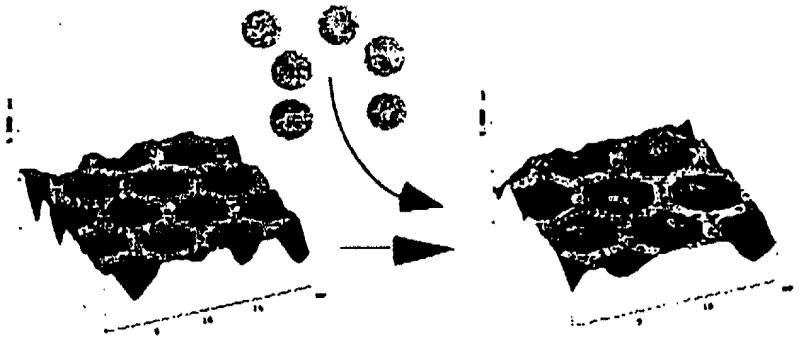


Figure 1. Atomic force micrograph of ~3.6 micron diameter microwells fabricated by chemically etching a polished 1,000 micron diameter imaging fiber (left). Each microwell is then filled with a single 3.1 micron diameter microsphere (right), on which about 10^7 molecules of a specific oligonucleotide sequence is attached.

These randomly ordered arrays offer some important advantages over conventional arrays: (1) Miniaturization and Scalability. Arrays are formed by a self-assembly process that is substantially size-independent. Currently, Illumina uses arrays with more than 13,000 wells in a 1.2 x 1.2 mm area withbeads that are ~ 5 microns in diameter. However, fiber-optic packing densities can be increased more than 100-fold over those already achieved. For example, arrays have been assembled with 0.2 micron diameter beads in 0.5 micron fibers (Michael et al., 1998). (2) Flexibility. Since arrays can be assembled rapidly, it is possible to change the composition of the probe set easily and at will. New oligonucleotide probes on beads can be prepared within one day, and assays can be revised or updated simply by the addition of new beads. The ability to easily make new arrays allows much more flexibility in designing and performing experiments in response to new data. (3) Manufacturability. A single synthesis of bead-linked oligonucleotides can provide sufficient material for millions of arrays. This unprecedented scale, coupled with the simple and rapid array assembly process, results in low manufacturing costs and highly reproducible arrays which should make the technology more accessible, and help speed the transition to clinical application.

In order to make use of the array, the identity of the beads at each location must be determined. Illumina, Inc. has developed proprietary methods of rapidly and efficiently decoding an entire array which may contain up to 2,000 unique probe sequences. In order to make the technology available to researchers worldwide, Illumina has recently partnered with PE Biosystems to commercialize the BeadArray™ technology in combination with PE Biosystems' ZipCode™ chemistry. As shown in Figure 2, hybridization to the bead arrays is straightforward and has been demonstrated in Dr. Walt's laboratory at Tufts University and at Illumina.

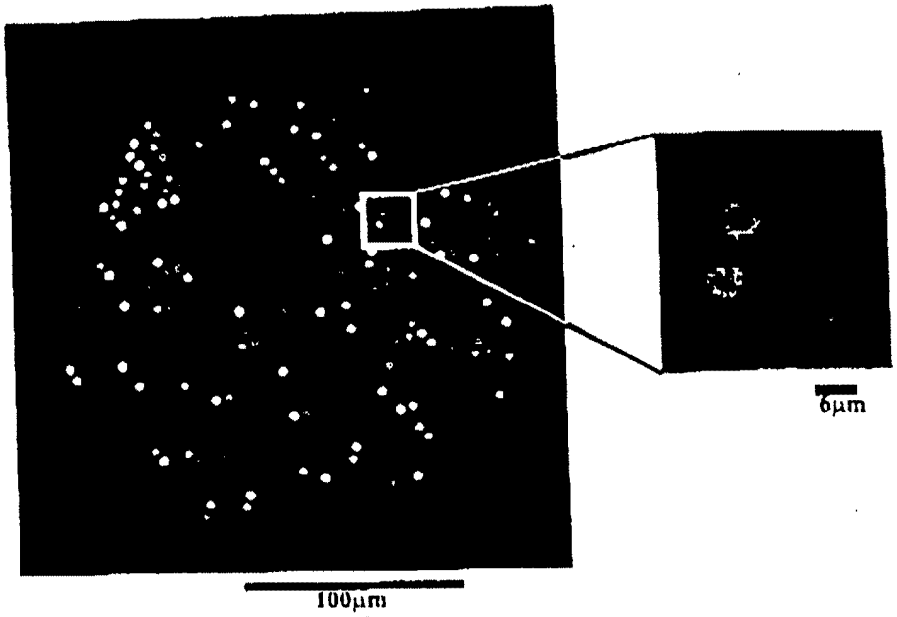


Figure 2. Image of a randomly ordered array of beads hybridized to fluorescently labeled target DNA. There are ~ 1,000 elements in this particular array, which is approximately the diameter of a single spot in a typical cDNA microarray. The chemistry in each well is physically separated from adjacent wells. Hence fluorescent signals from individual beads are well separated and distinct from each other, as can be seen from the magnified image in the figure.

The addressable array strategy:

To avoid the disadvantage of designing new arrays for each new set of targets, an alternative approach is to provide an array that is "universal" and can be used for any set of biological targets. This allows an investigator to use the same array for different target sequences, which removes the need for costly specialized designs. A universal array of this type has been described (Gerry et al., 1999). Such an array consists of a set of artificially generated probes that are sufficiently long and unique to hybridize with high specificity. These probes act as addresses or "zip-codes" on an array. In order to make use of the array, target sequences to be analyzed are linked with specific zip-code sequences (after PCR amplification, the complementary zip-code sequence will be used to hybridize to the zip-code probes on the array). Thus, any set of targets can be analyzed using the same set of zip-codes by attaching them to appropriate targets. This approach has been used for point mutation and SNP analysis (Gerry et al., 1999; Fan et al., 1999).

A somewhat similar approach has been used for a different type of application (Shoemaker et al., 1996). In this case, yeast strains were constructed to contain in vivo "tags" in order to carry out parallel analysis of yeast deletion mutants (Shoemaker et al., 1996; Winzeler et al., 1999). Briefly, the authors deleted individual open reading frames (ORFs) in *S. cerevisiae*, leaving behind two unique tag sequences in the genome to serve as molecular "bar-codes" for each specific deletion event. Each bar-code was flanked by two universal primer sites for PCR amplification from yeast genomic DNA. Hundreds of yeast strains (each containing deletion of a single ORF) were pooled to monitor their growth in a competitive fashion. At each growth point, DNA was isolated and each bar-code amplified using two universal primers. The PCR products were labeled (using biotinylated forward primer) and hybridized to a "tag" array for sorting and quantitation. Unfortunately, these arrays are not yet easily available.

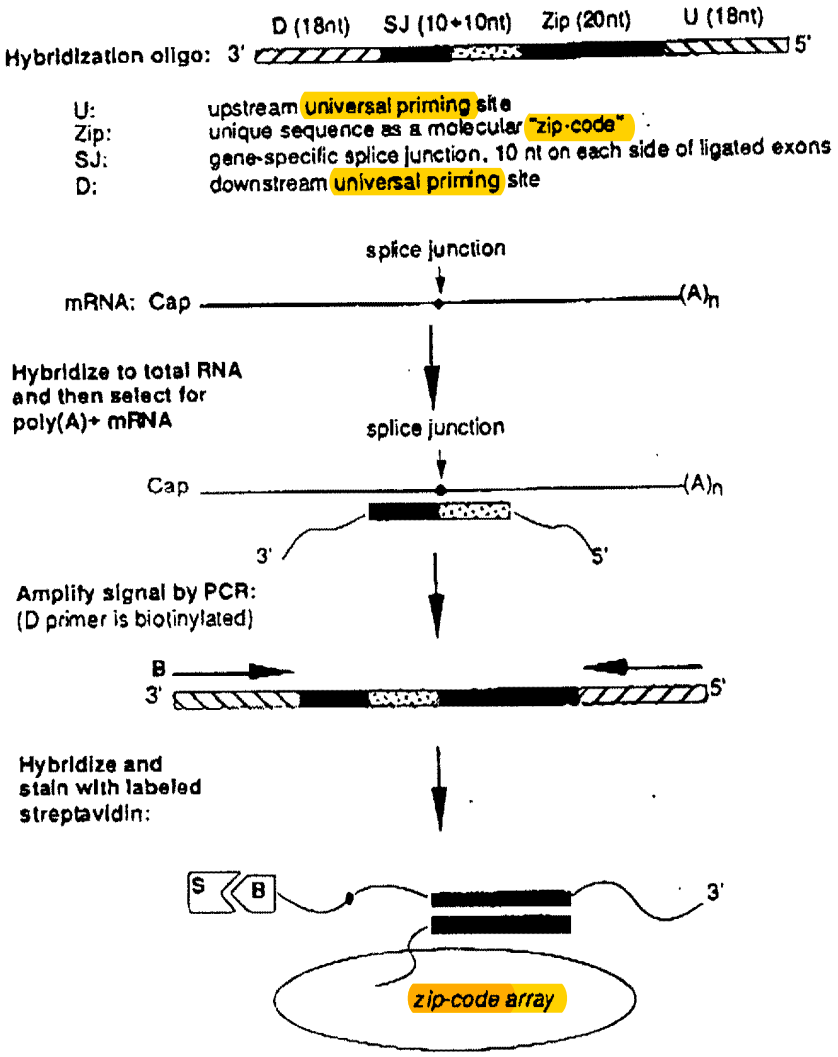
Currently, Illumina has developed universal BeadArrays™ with 128 unique addresses that hybridize efficiently and with high specificity. Development of arrays with up to 2,000 addresses is in progress. In addition, Illumina has developed an array format that allows for convenient parallel processing of many samples. This format, the Array of Arrays™, comprises multiple fiber optic bundle arrays set in a block. Each array addresses an individual well in a microtiter plate. Samples are prepared in wells in a microtiter format, and the array block is then used to analyze all the samples in parallel, instead of one at a time as in conventional array experimentation.

After examination of various microarray techniques, we conclude that it is most advantageous to make use of the universal (zip-code addressable) Illumina BeadArray™ for scaleable parallel alternative splicing analysis, as described below. The willingness of Dr. Fan to collaborate with us, Illumina's willingness to provide access to the technology, and the fact that Illumina is located nearby (~5 minutes from UCSD) are all advantages. In addition, as outlined above, we believe the BeadArray™ technology offers advantages in minaturization, flexibility, format and cost.

Our experimental strategy: The strategy is designed based on the following considerations: (1) We would like to develop an assay that can be used by individual investigators who have no access to specialized arrays. (2) We would like to avoid the requirement for PCR-amplification of alternatively spliced regions. (3) We hope to develop an assay that has maximal sensitivity. The criteria are selected to enable cost-effective large scale analysis of the role of alternative splicing events in cancer pathogenesis and progression. These requirements are fulfilled in the following experimental scheme:

As diagrammed in Figure 3 **(please keep the figure and the idea behind it confidential)**, we will synthesize oligos such that each will consist of 4 parts (from 5' to 3'; note the direction from right to left): an upstream universal 18 nt priming site (U), a unique 20 nt zip-code sequence (Zip) to target the oligo to a specific address on the array, a 20 nt sequence complementary to a specific splice junction (SJ), and a downstream universal 18 nt priming site (D). The total length will be 18+20+20+18=76mer. Many different oligos may be pooled to form groups for specific applications. Pooled oligos will be hybridized to total RNA from cells or tissues under conditions (high salt, SDS, etc.), which minimize the impact of Tm on hybridization as well as problems of non-specific sticking (to carrier RNA, to tube surface, etc.). After hybridization, we will carry out selection of poly(A)$^+$ mRNA on oligo-dT columns, and splice junction oligos annealed to mRNAs will be PCR-amplified using two universal primers (the forward primer will be biotinylated). The number of PCR cycles will be

Figure 3: A flow chart for array-based detection of RNA alternative splicing



optimized to obtain sufficient products and at the same time maintain the original ratio of individual oligos. The PCR products will be hybridized to BeadArrays™ manufactured by Illumina and hybridization signals will be detected and quantified by staining with fluorescently labeled streptavidin. Individual splicing events will then be detected and quantified by imaging the arrays. Samples from different cells or the same cells grown in different conditions will be labeled with different colors and differences in alternative splicing will be revealed by co-hybridization. As described in the Methods section, a series of controls will be included and potential pitfalls of this approach will be addressed.

Throughout this technology development phase, we are going to collaborate with Dr. Jian-Bing Fan at Illumina for a number of reasons. First, no array of the type needed for these experiments is commercially available. Dr. Fan is able to provide us with access to state-of-the-art array technology prior to its commercial introduction. Secondly, in the R21 phase the hybridization and data collection will be done on existing equipment at Illumina. As a result, the expense of carrying out the array analysis is kept to a minimum. Most importantly, we share common scientific interests in exploring new applications of array technology. We believe that this collaborative effort will benefit both basic scientific research and technology development, and lead to the provision of new tools for the scientific community.

Uniqueness of the assay: This experimental scheme diagrammed in Figure 3 is advantageous over standard array-based gene expression protocols in several respects: (1) Although we have not yet determined the minimal amount of total RNA required for each experiment, this method has the potential to detect alternative splicing events from a single cell because of signal amplification of annealed oligos after poly (A)$^+$ selection. (2) It allows oligos to hybridize directly with their RNA targets, thereby omitting a cDNA conversion step. (3) The hybridization reaction takes place in solution instead of on a surface so that DNA and RNA behave more predictably according to their thermodynamic properties. (4) The poly (A)$^+$ RNA selection step eliminates excess oligos as well as those that are nonspecifically hybridized to RNA (in the presence of rRNAs and tRNAs as competitors) or stick to the container surface. As a result, isolated oligos will reflect the level of individual splicing events in cells and the background signal should be significantly reduced. (5) After the initial poly (A)$^+$ RNA selection, the use of universal primers avoids biased signal amplification by PCR.

In conclusion, if the proposed project is successful, it will provide a powerful tool in cancer classification and target identification. The technology is novel in multiple aspects as described above and will have a broad impact on both basic and clinical research.

c. Preliminary Studies/Progress Report (evidence for feasibility)

Every array experiment involves a number of components: (1) Target identification. From over a decade of work in the splicing field, we have collected many examples of alternative splicing events and their relevance to biology and medicine. In this R21 phase, we aim to develop an array-based assay for alternative splicing using a limited number of targets. In the next R33 phase, we plan to enlarge the assay, first to include as many reported examples of alternative splicing as possible, and then move to test those derived from the search for splicing events related to specific pathogenetic processes. (2) Preparation of arrays. Since we plan to develop the alternative splicing assay on a standard platform, we will need a steady supply of the universal arrays. By jointly developing the technology with Illumina, we will have access to their arrays through a collaborative arrangement. Our collaborator, Dr. Jian-Bing Fan, has many years of experience in developing array technology previously at Affymetrix and now at Illumina. He has co-authored a series of papers on the development and application of array technology through collaborations with leaders in academia, including Drs. Eric Lander, Francis Collins, Pui-Yan Kwok, and Aravinda

Chakravarti (see Dr. Fan's CV). (3) Assay development. As detailed below, we have designed a systematic approach. Protocols for RNA isolation, poly(A)$^+$ selection, PCR amplification, and hybridization are standard. The use of arrays for quantitative analysis of RNA transcripts has been established (Schena et al., 1995; Lockhart et al., 1996). The uniqueness of our approach lies in using splice junction sequences and conducting initial hybridization with total RNA instead of using PCR-amplified products. The key to success is to develop experimental conditions to maximize the hybridization efficiency and minimize background. (4) Data analysis. Our collaborators, Dr. Michael Gribskov at the San Diego Supercomputer Center and Dr. Michael Q. Zhang at the Cold Spring Harbor Laboratory, are world experts in bioinformatics, and they will work with us in this project to develop computer programs for experimental design and data analysis. We will explore a number of academic and commercial software packages for data analysis, including GeneCluster (Whitehead, MIT) and GeneSpring (Silicon Genetics). In addition, we plan to develop our own data analysis and presentation software by modifying the current software called High-Density Array Pattern Interpreter (HAPI) under development in Dr. Gribskov's laboratory.

d. Research Design and Methods

Aim 1: Establish a set of criteria for selecting splice junction targets.

The BeadArray™ technology developed at Illumina: At Illumina, individual oligonucleotide probes are synthesized on beads. A mixture of beads is then loaded onto the tip of a fiber optic bundle to form a randomly ordered array of beads in wells. Because the array is assembled randomly, a decoding procedure is used to identify sequences at each address. A large number of the Illumina BeadArrays™ can be prepared rapidly and cost-effectively, therefore allowing parallel analysis of a large number of alternative splicing events in multiple samples. All experiments involving hybridization to the BeadArrays™, including imaging and data analysis in this R21 phase will be conducted by Dr. Jian-Bing Fan at Illumina.

Selection criteria for oligos that are complementary to specific splice junction sequences: Splice junction sequences are being compiled in our lab. In principle, each 20 nt oligo will be complementary to 10 nt on each side of a splice junction so that it will only hybridize to spliced mRNA. Such a strategy has been applied to detect co-transcriptional splicing in the nucleus (Zhang et al., 1994). To control for hybridization efficiency and specificity, we will choose target sequences with a similar Tm (annealing temperature) and minimal potential to cross-hybridize to unrelated mRNAs. For this purpose, we may slide oligos either upstream or downstream of splice junctions in order to choose sequences with relatively similar Tm values. We will also BLAST search each potential target sequence against the GenBank database to select against sequences that contain larger than 75% sequence identity to other mRNAs (in the other words, we will select sequences which contain at least 5 nt mismatches to unrelated mRNA in the database). In collaboration with Dr. Gribskov, we will develop software to aid in the selection. This software will incorporate accurate calculations of Tm based on dinucleotide composition, as well as automatic filtering of foldback sequences and screening for spurious matches to sequences in genomic databases.

In theory, only about 1% of the human genome ($3X10^9$ bp) encodes mRNA so that the sequence complexity of mRNA is about $3X10^7$ (assuming here that these sequences are randomly distributed), which is less than the sequence complexity of random 13 nt sequences ($4^{13} = 6.7X10^7$). Thus, sequences of longer than 13 nt should be unique in human mRNA. Furthermore, splicing signals are mostly conserved in introns rather than in exons and exon-exon junction sequences are relatively diverse (Burge et al., 1998). Thus, it should be feasible to select a unique sequence of 20 nt in length around the splice junction region for each alternative splicing event. In case the composition of some specific splice junction sequences still precludes

their selections as hybridization targets (for example, due to extreme Tm values and/or potential cross-hybridization), we will slide targets away from the junction regions toward alternative exon regions.

Aim 2: Prove the concept with alternative splicing reporters in transiently transfected cells.

To determine the feasibility of the proposed experimental scheme for large scale analysis, we will first use the β-thalassemia and E1A splicing reporters, which are cloned into the pcDNA3 eukaryotic expression vector, to establish experimental conditions. These splicing reporters are commonly used for splice site selection experiments in transfected cells (Caceres et al., 1994). Each pre-mRNA contains three alternative 5' splice sites, which compete for a common 3' splice site. The β-thalassemia reporter is derived from a mutant β-globin gene and the E1A reporter is a DNA virus transcription unit. Thus, they have no counterparts in most human and mouse cell lines. Moreover, it was previously shown that overexpression of a splicing regulator (such as ASF/SF2, a member of the SR family of proteins) could induce the selection of proximal splice sites in both cases (Caceres et al., 1994). Therefore, we will also be able to compare their alternative splicing in the same cell type in the presence or absence of a splicing regulator.

We will synthesize 6 specific oligos, each of which contains (from 5' to 3') the T7 promoter sequence (sense strand), a unique zip-code sequence, a target sequence complementary to a specific splice junction, and the T3 promoter sequence (antisense strand). As described in detail below, these oligos will be hybridized to total RNA from transfected and mock-transfected cells. After poly (A)$^+$ selection, the T7 and T3 promoter sequences will serve as the universal primer binding sites for PCR-amplification, and the products will be labeled and hybridized to the universal array. To improve the signal-to-noise ratio, two 18 nt oligos complementary to the priming sites will be added to the hybridization mix in both solution and array hybridization reactions to reduce nonspecific sticking of oligos through the priming sites.

A series of control oligos: To obtain meaningful hybridization results, a number of controls will be built into the assay. We are considering two types of controls:

Control for variation in gene expression: When we compare splice site selection in total RNA derived from two different cell types or from the same cell type but grown under different conditions, we need to normalize the expression level of a gene before comparing the absolute level of a specific mRNA isoform of the gene. For this purpose, an oligo complementary to a common constitutive exon sequence for each transcript will be included. The exonic regions will be selected for their G+C content, secondary structure, and uniqueness in the genome using the same criteria as those for selecting splicing junction targets. It should be pointed out that this control may not be essential if we only compare the ratios of different mRNA isoforms from different samples (for example, an increase of isoform A should be accompanied by a decrease of isoform B for a switch in splice site selection from B to A). In the R21 phase, we plan to include the expression level controls in our experiments and determine whether these controls are required for quantitative comparison. We would like to eliminate non-essential controls for economic considerations, especially for large scale assays in the R33 phase.

Control for hybridization specificity: This control is crucial because target oligos may nonspecifically stick to unrelated mRNAs (via either its zip-code or splice junction sequences or both). These nonspecific binding events will be PCR-amplified, which may overwhelm specific hybridization signals. To overcome this problem, we will include a mismatch control for each target oligo. Such a mismatch control will contain two mutations in the splice junction region, one in the middle of the upstream 10 nt arm and the other in the middle of the downstream 10 nt arm. According to a previous study of artificial mismatch hybridization, the locations of the two mismatches are sufficient to provide a critical hybridization specificity control (Guo et al., 1997). To control for the "sticky" problem, each pair of specific target sequence and its mismatch

control will be encoded with the same zip-code sequence. Specific oligos and their mismatch controls will be separately pooled and hybridized to two identical aliquots of total RNA. After poly (A)$^+$ selection and PCR-amplification, specific and control oligo pools will be labeled with two different fluorophors (say green for the specific oligo pool and red for the mismatch control pool), respectively. The products will be mixed and co-hybridized to the array. We expect greater green signals in these experiments. If a mismatch control oligo is nonspecifically stuck to RNA, however, it will be selected and amplified, and as a result, it will co-hybridize with the specific oligo on the universal array to give rise to a yellow signal. Through these control experiments, we will establish experimental conditions that achieve a satisfactory signal (generated by specific oligo targets) to noise (produced by their mismatch controls) ratio.

Experimental procedures: HeLa or COS-7 cells in 10 cm dishes (about $3X10^6$ cells) will be transfected with a splicing reporter or mock-transfected. Total RNA will be isolated using TRIzol^{TM} reagent from Gibco Life Technologies (each 10 cm culture would yield about 1 to 3 ug total RNA) and resuspended in 20 ul TE buffer. Poly (A)$^+$ selection will be conducted using the mRNA isolation kit from Miltenyi Biotec. According to instructions from the manufacturer, total RNA will be heated to 65^{0}C for 3 min. and then placed on ice before mixing with 100 ul of Binding Buffer containing 0.5 M NaCl and 1% SDS at room temperature. 25 ul of Oligo (dT) MicroBeads will be added to the mixture, and without further incubation, the mixture will be loaded onto a MACS column in a magnetic field. After extensive washing, poly(A)$^+$ mRNA retained on the column will be eluted in a low salt buffer or with H_2O.

Oligo (dT) of 24 nt in length is able to bind efficiently and specifically to the poly (A)$^+$ sequence under these conditions. Thus, splice junction oligos of 20 nt length with a high G+C content should also bind to specific mRNAs. In our application, we plan to mix 0.1 ug of total RNA from transfected or mock-transfected cells respectively with pooled oligos (0.01 pmole each, which is in large excess of RNA molecules) in the high salt binding buffer. After the poly (A)$^+$ selection, we will PCR amplify oligos attached to mRNA and analyze the PCR products on agarose gels. We expect to detect specific PCR products with mRNA from transfected cells, but not with mRNA from mock-transfected cells. We will determine the minimal number of PCR cycles that are required to amplify the signal from mRNA derived from transfected cells and the maximal number of PCR cycles without amplifying the signal associated with mRNA from mock-transfected cells. These experiments will be carried out repeatedly to establish an appropriate salt concentration, annealing temperature, and washing stringency in order to minimize background and maximize the sensitivity of the assay.

Control the specificity: Once the right conditions are identified, we will divide total RNA from transfected cells into two aliquots, one for hybridization with specific oligos and the other with corresponding mismatch controls. We will carry out poly (A)$^+$ selection and PCR amplification in parallel. Ideally, we expect to detect PCR products with specific oligos, but not their mismatch controls. Some PCR signal with mismatch controls may persist, which would indicate nonspecific sticking of one or a few control oligos. To find out which of these controls are sticky, we will label specific oligos with one color (red) and their mismatch controls with another (green), and co-hybridize them to the universal array. A red signal will indicate specific binding of a splice junction target, and a yellow signal will indicate nonspecific hybridization. We will then redesign that pair of oligos by sliding the target region upstream or downstream of the splicing junction and retest the specificity of the new target sequence. In parallel, we will also adjust experimental conditions to eliminate this sticky background problem.

Quantify the difference: Our next step is to determine whether we can quantitate alternative splicing using array technology. For this purpose, we will prepare total RNA from cells that are (1) mock-transfected, (2) transfected with a splicing reporter, and (3) co-transfected with the splicing reporter and a splicing regulator

(such as the SR protein ASF/SF2). We will carry out conventional RT-PCR analysis of the splicing products to obtain the splicing pattern in the presence or absence of the splicing regulator. On agarose gels, we will first make sure that input oligos can be isolated and amplified with mRNA from transfected or co-transfected cells, but not from mock-transfected cells. We will then label splice junction targets annealed to mRNA from transfected cells with one color and those annealed to mRNA from co-transfected cells with a different color. These samples will be co-hybridized to the universal array. The data will be compared with those obtained with conventional RT-PCR analysis.

Determine the sensitivity: Finally, we will determine the sensitivity of the proposed experimental scheme. Because the splicing reporters are overexpressed, it may be relatively easy to detect their spliced products with either PCR or using the array. To determine the sensitivity of our experimental protocol, we plan to mix total RNA from mock-transfected cells with that from transfected cells in differential ratios and determine the minimal amount of total RNA required to obtain a positive result. Because of PCR amplification after the poly $(A)^+$ selection step, it may be possible to detect specific splicing signals from total RNA isolated from a single cell.

Aim 3: Validate the technology on endogenous transcripts in model systems.

Our assay is designed to detect alternative splicing of endogenous transcripts. Initially, we plan to select a handful of well-documented alternative splicing events for the test. As listed in Table 1 on the next page, these genes are widely expressed so that we will be able to detect their splicing in commonly used human cell lines. We will focus our comparison on fibroblasts and epithelial cells so that we can monitor alternative splicing of FGFR2/KGFR as internal controls for cell-specific alternative splicing (as described in Background, FGFR2 is expressed in fibroblasts and KGFR in epithelial cells). The results of the array experiments will be verified by RT-PCR.

Specificity control using mismatch oligos: In this R21 phase, the most important work is to establish standard experimental conditions for solution and array hybridization and for PCR-amplification. In these experiments, the mismatch controls are critical. As described above, we will prepare a mismatch control for each splice junction target, and both the splice junction target and its mismatch control will be encoded with the same zip-code sequence. Specific splice junction targets will be pooled and their mismatch controls will be pooled separately. We will isolate total RNA from HeLa (epithelial) and WI-38 (fibroblast) cells. One aliquot of HeLa RNA will be hybridized to splice junction targets and the amplified products will be labeled with red. The other aliquot of HeLa RNA will be hybridized to the pool of mismatch controls and the amplified products will be labeled with green. Similarly, one aliquot of WI-38 RNA will be hybridized to splice junction targets, but the amplified products will be labeled with green and the other aliquot with the mismatch controls and the amplified products labeled with red. These four samples will be mixed in 4 pairs for co-hybridization to the universal array.

The first pair will be the mix of red (specific) and green (mismatch control) signals from HeLa and the second will be the mix of green (specific) and red (mismatch control) signals from WI-38 (note that the order of colors is reversed with their mismatch controls). These two pairs will provide controls for hybridization specificity. For example, we expect the most red hybridization signal with RNA from HeLa cells and the most green hybridization signal with RNA from WI-38 cells. A yellow signal will indicate a nonspecific sticky problem associated with that particular pair of target and its mismatch control.

It is expected that a sticky pair of oligos will have a similar nonspecific binding problem regardless of the origin of total RNA. To confirm this prediction, the third pair will include red and green mismatch control signals from both cell lines. Yellow hybridization on the array will indicate a nonspecific sticky

problem. In this case, the oligos will need to be redesigned by choosing a different region near the alternative exon or using a different zip-code sequence.

The fourth pair will be the mix of red and green splice junction targets from HeLa and WI-38 cells. This mix will report the difference of individual alternative splicing events between the two cell lines. As internal controls, KGFR should be specific for HeLa cells and thus the hybridization signal should be red. Conversely, FGFR2 should be specific for WI-38 so that the hybridization signal should be green. A yellow signal with a particular splice junction target will indicate a similar splicing event in both cell lines.

Table I. Splicing targets selected for the R21 phase studies

	Gene Name	Functional Significance	Selected Reference
1	acetylcholinesterase	synapse maturation	Luo et al., '98
2	agrin	AChR clustering at synapses	Ferns et al., '92
3	AML1	transcriptional activation	Tanaka et al., '97
4	ASF/SF2	mRNA splicing regulation	Ge et al., '91
5	Bcl-x	apoptosis regulation	Boise et al. '93
6	BRCA1	transcriptional activation	Cui et al., '98
7	c-src	signal transduction regulation	Modafferi and Black '99
8	calcium channel, alpha1A	neurotransmitterrelease	Bourinet et al., '99
9	calcium channel, alpha1B	neurotransmitterrelease	Lin et al., '99
10	caspase 1 (ICE)	apoptotic regulation	Alnemri et al. '95
11	caspase 2 (Ich-1)	apoptotic regulation	Jiang et al. '98
12	CD45	T cell maturation	Ratech et al. '97
13	clathrin light chain B	receptor-mediatedendocytosis	Daoud et al. '99
14	cytochrome P450, aromatase	steroid hormone synthesis	Utsumi et al. '96
15	estrogen receptor 1	hormone response	Balleine et al.'99
16	estrogen receptor 2	hormone response	Hanstein et al. '99
17	fas ligand	apoptotic regulation	Ruberti et al. '96
18	FGFR2	signal transduction	Luqmani et al. '95
19	fibronectin 1	wound healing	Vogelezang et al. '99
20	fyn	growth control	Weil et al. '99
21	glutamate receptor NMDAR1	neurotransmission	Koltchine et al. '96
22	Hel-N1	mRNA turnover	King '94
23	insulin receptor	signal transduction	Moller et al. '89
24	integrin beta	cell adhesion	Meredith et al. '95
25	jun kinase 2	transcriptionalcofactor	Bost et al. '99
26	MUC1	cell surface tumor marker	Baruch et al. '99
27	myosin heavy chain	muscle contraction	Haase and Morano '96
28	NCAM	cell adhesion	Rafuse and Landmesser '96
29	nNOS	neurotransmission	Wang et al. '99
30	p15 CDK inhibitor 2B (INK4b)	cell cycle control	Tsubari et al. '97
31	p16 CDK inhibitor 2A (ARF)	cell cycle control	Robertson and Jones '99
32	presenilin 2	apoptotic regulation	Sato et al. '99
33	prostate-specificantigen	cell surface tumor marker	Su et al. '95
34	SMN	spinal motor neuron survival	Lorson et al. '99
35	SRp40	mRNA splicing regulation	Du et al. '98
36	tau	neuronal maturation	Varani et al. '99
37	telomerase	chromosomal integrity and growth control	Ulaner et al. '98
38	transformer 2 beta	mRNA splicing regulation	Daoud et al. '99
39	tropomyosin 1 (alpha)	muscle contraction	Kashiwada et al.'97
40	tropomyosin 2 (beta)	muscle contraction	Gimona et al. '95
41	troponin T3	muscle contraction	Ogut et al. '99
42	VEGFR-1	angiogenesis and vascular permeability	He et al. '99
43	Wilms tumor 1	transcriptional regulation	Webster et al. '97

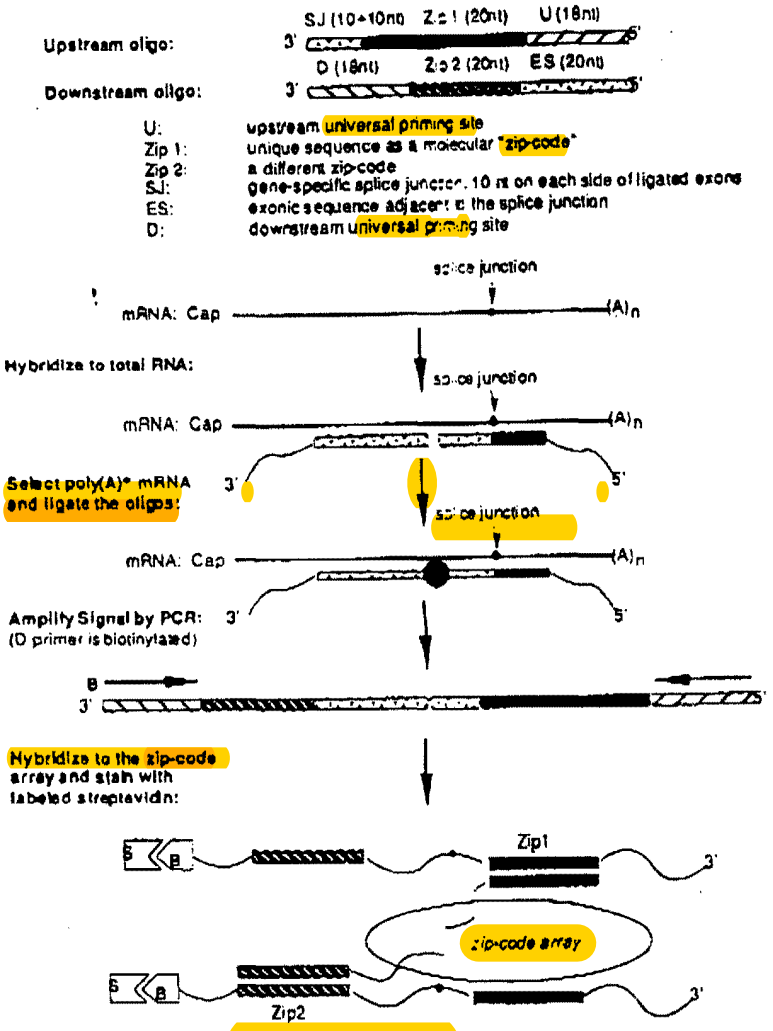
Determination of the sensitivity of the assay with endogenous transcripts: Although we may obtain sufficient information about the sensitivity of the assay by titrating RNA from transfected cells as described above, we plan to determine the assay's sensitivity using endogenous transcripts. By Northern blotting analysis or by quantitative RT-PCR, we will determine the abundance of individual transcripts in the group

we selected for these initial phase studies. By comparison with an RNA standard, we should be able to estimate the copy number per cell of individual transcripts. We will then make a serial dilution of total RNA from HeLa cells and determine the minimal amount of RNA required to obtain a reliable result on alternative splicing of endogenous transcripts. This will indicate how many cells are required and also estimate how many copies of each transcript are required for our assay.

Potential pitfalls and solutions: An alternative approach: The major potential pitfall of our approach is nonspecific annealing of oligos. As a result, those oligos will be amplified, which may surpass specific signals resulting from specific annealing. This may be especially problematic with rare mRNA isoforms. Although we have designed critical controls for each oligo target, it will seriously jeopardize our experimental strategy if a large number of oligo targets have the sticky problem. In considering an alternative approach, we may modify our experimental procedure in the following way:

As shown in Figure 4 (**please keep the figure and the idea behind it confidential**), we will make two oligos as a set for each splicing event instead of synthesizing a single long oligo. The upstream oligo will consist of three parts (from 5' to 3'): the 18 nt upstream universal priming site, a 20 nt zip-code sequence, and a 20 nt sequence complementary to a specific splicing junction sequence. Similarly, the downstream oligo will also have three parts (from 5' to 3'): a 20 nt sequence complementary to the downstream exonic sequence that is immediately adjacent to the region covered by the upstream oligo, a different 20 nt zip-code sequence, and the 18 nt downstream universal priming site. The downstream oligo will be kinased (so that it contains a 5'-phosphate) before use. These target sets will be hybridized to total RNA. Paramagnetic oligo (dT) beads will be added to the mix. After loading to the poly $(A)^+$ selection column and extensive washing in the magnetic field, the column will be removed from the magnetic filed so that the oligo (dT) beads is no longer magnetic, which will allow the recovery of the attached mRNA with annealed oligos (along with the oligo (dT) beads) by washing the column with buffer. This procedure is essential to make sure that oligos remain hybridized to their RNA targets. We will then carry out the ligation reaction at room temperature (or higher) using T4 DNA ligase to link the upstream and the downstream oligos that are bridged together through spliced mRNA. This reaction should work since many investigators in the splicing field have been ligating RNA fragments bridged by DNA

Figure 4: The two-oligo hybridization and ligation strategy



sequences with T4 DNA ligase (Moore and Sharp, 1992). If necessary, we may further increase the ligation stringency and efficiency by using thermostable Tsc DNA ligase at higher ligation temperatures, which will be determined by experiments, and the result will be compared with those obtained with T4 DNA ligase. After the ligation step, oligos will be PCR-amplified using two universal primers and the products will be applied to the universal array.

This procedure will prevent oligos that are nonspecifically stuck elsewhere from being ligated and subsequently amplified, thereby eliminating the background problem. Because of this critical advantage, we may not need mismatch controls for each set of oligos (although we shall experimentally determine if this is the case). Thus, this assay strategy may be more economical. Furthermore, each ligated and amplified target will have two zip-code sequences so that we can detect them on two separate probes on the array, which will reduce hybridization variation associated with individual zip-code sequences. I would like to emphasize here that although we describe this alternative approach in this third aim which focuses on alternative splicing of endogenous transcripts, it may be advantageous to compare the single oligo method with this double oligo approach in the second aim. We therefore plan to test both experimental procedures throughout this R21 phase.

In conclusion, we will establish an experimental procedure that combines efficiency and accuracy to conduct parallel analysis of alternative splicing events in model cellular systems in this R21 phase, and then extend the assay to large scale analysis in the next R33 phase.

Milestones for the R21 phase.

We anticipate reaching two milestones to conclude the R21 phase in about a year. Specifically:

1. We should be able to develop the assay so that we can use it to determine splice site selection of the chosen alternative splicing reporters in transfected cells. The results should match those obtained with conventional analyses.
2. A standardized set of controls and experimental conditions should be established and optimized for detecting the splicing pattern of endogenous transcripts in model cell lines. Again, the results should match those generated by conventional methods such as RT-PCR or RNase protection.

The R33 phase studies

a. Specific Aims

The assay developed in the R21 phase for large scale analysis of alternative splicing will pave the way for an organized effort to apply the technology to both basic and clinical research. We plan to pursue the following areas through a multilateral collaboration involving Drs. Jian-Bing Fan and Mark Chee at Illumina, Dr. Michael Gribskov at the San Diego Supercomputer Center, Dr. Michael Q. Zhang at Cold Spring Harbor laboratory, Dr. David Tarin at the UCSD Cancer Center, and our laboratory in the Department of Cellular and Molecular Medicine, UCSD.

Aim 1: Construction of a database for experimental design and data presentation.

Our assay for parallel analysis of alternative splicing requires a database for target collection and classification for experimental design as well as for documentation of the experimental data. Although a database for alternative splicing was recently published, it is incomplete and has a quality problem because construction of this database relies solely on computer analysis, which often amplifies artifacts in original sequence databases. In this proposal, we describe a strategy to combine computer analysis with manual annotation for target identification and data analysis.

Aim 2: Parallel analysis of alternative splicing in cell and animal models.

We plan to enlarge our assay in a step-wise fashion to include most reported alternative splicing

events in mice and humans. We will group genes according to their functions and relative abundance in cells and validate target design for each gene by assaying their alternative splicing in model cell lines. As a test case, we plan to apply the technology to our basic research in order to identify alternative splicing events that are responsive to signal transduction pathways or regulated by a specific splicing regulator in cells and/or in knockout mice.

Aim 3: Applications in cancer research: developing new tools for disease classification.

In this application, we propose a new approach in which changes in alternative splicing patterns are compared as independent variables with morphological and clinical parameters. We plan to collaborate with Dr. Tarin's Molecular Pathology group to analyze their collection of breast and prostate cancer samples. The main objective is to find sets of molecular and cellular characteristics which can assess more accurately the aggressiveness of the tumor. If successful, we may also extend the assay to other diseases.

Aim 4: Assay development for both small and large-scale applications.

We believe that the technology described in this proposal has wide application in both basic and clinical research. To make this technology useful and accessible to a variety of researchers, we will work closely with our collaborators at Illumina to develop various combinations of splice-site assays for small and large-scale applications of the technology in both academia and industry.

b. Background and Significance

Alternative splicing represents a major challenge in the post-genome era. In the recent "Genome Issue" of Science, an article used a gene called erythrocyte protein 4.1 to illustrate that because many exons that are known to be included or excluded during RNA processing, as many as 900 different transcripts may be generated through alternative splicing (Pennisi, 1999). As a community, we do not even have a database to list reported alternative splicing events in the literature, not to mention those derived from comparison with EST databases. This problem is further complicated by numerous sequencing errors in the existing DNA and protein databases, and as a result, artifacts are amplified when alternatively spliced transcripts are deduced from these databases under the assumptions that differences in the encoded sequence reflect true alternative splicing. Clearly, some efforts are being made at NCBI to deal with this problem. In the meantime, however, we would like to build up a working database to include well-documented alternative splicing events to aid our experimental design and data analysis. Although this approach may overlap with other similar efforts in database construction, we believe that we are in a unique position to combine database building and data mining with a practical experimental approach. We hope that our initiative can eventually be integrated into a national consortium to systematically attack the alternative splicing problem. In the meantime, the new technology described in this proposal may find a variety of immediate applications in both basic and clinical research, some of which are described below.

Problems confronting basic research of splicing mechanisms and regulation: A large number of our colleagues in the splicing field are studying the mechanisms and regulation of constitutive and alternative splicing. However, as a group, we are approaching the problem using about a dozen model systems. Many potential splicing regulators are identified through biochemical purification and molecular cloning. However, in most cases, endogenous targets for a particular splicing regulator remain elusive. For example, SR proteins are a family of serine/arginine-rich RNA binding proteins, which are not only required for constitutive splicing, but also shown to affect splice site selection in a dosage dependent manner in vitro and in transfected cells (see review by Fu, 1995). Likewise, heterogeneous nuclear ribonucleoproteins (hnRNPs) play a variety of functions in eukaryotic cells (see review by Krecic and Swanson, 1999). A subset of them are shown to counteract SR proteins in splice site selection. The problem is that no one knows the

endogenous substrates or the unique functions and substrate specificities of SR or hnRNP proteins in a particular cell type and/or during development. Thus, a large scale parallel analysis of alternative splicing is much needed to couple with gene targeting (knockout or knockin) techniques to identify specific alternative splicing events that are controlled by a particular splicing regulator.

Some alternative splicing events may be targets for oncogenic transformation or regulated by various signaling pathways and the functions of some specific alternative transcripts may be directly responsible for altered cell growth and differentiation. However, little is known in this area of research. Armed with a large scale detection method for alternative splicing, we will be able to quickly sort out specific alternative splicing events that are responsive to or associated with treatment with certain growth factors or drugs, activation of oncogenes or inactivation of antioncogenes, changes in a developmental program, etc. These studies may in turn provide important clues to the mechanisms of specific cellular responses, and the information may be critical for devising strategies for disease prevention and treatment. In a way, functional genomics will and has been changing our approaches to fundamental biological problems from hypothesis-driven searching to search-driven hypothesis testing.

Novel avenue for disease classification: The identification, classification and prognostic evaluation of tumors has until now depended on histopathological criteria. The purpose of a classification scheme is to identify subgroups of tumors with homogenous properties, which can be further studied and compared with each other. Such classification has been an essential first step in identifying the causes of various types of cancer and in predicting their clinical behavior. However, molecular and biochemical characteristics are not revealed by these approaches Therefore, the current classification of tumors, although useful, is insufficiently sensitive for prognostic assessment of individual patients and for probing the causal mechanisms involved. Clearly, integration of a broad range of information from genetic, biochemical and morphological approaches is needed. What is needed is a strategy which is independent of conventional morphologic classification schemes and which identifies a new spectrum of molecular and cellular characteristics. Because alternative splicing events often alter the functions of their corresponding gene products, a detailed study of alternative splicing events in selected , staged tumor samples compared to matched normal tissues from the same patient offers a novel approach to this objective.

Monitoring global changes in gene expression has recently been applied to disease diagnosis and classification in leukemia (Golub et al., 1999). In this exciting paper, the authors demonstrated the feasibility of distinguishing two types of leukemia, ALL (acute lymphoblastic leukemia) and AML (acute myeloid leukemia), based on the expression profile of a group of genes. A mathematical model was developed to predict the disease type without prior pathological diagnosis. Although this approach may be extended to many other diseases, it is unclear at present whether it is universally applicable to every disease. Thus, development of additional molecular classification methods may be complementary to the approach, and in combination, they may be more powerful and accurate and require fewer genes in a test group, making it more feasible for clinical applications.

Because numerous genes are alternatively spliced and many alternative splicing events have been linked to cell growth, differentiation and transformation, some may be associated with specific disease states and therefore could serve as a distinct indicator for a disease and/or different stages of a disease. A particularly clear example appears in the disorderly expression of the human CD44 gene in several common types of cancer (Matsumura and Tarin, 1992, Matsumura et al., 1994). Inappropriate overproduction of many splice variants is seen in malignant cells, but not in matched normal cells from the same patient. In particular, as briefly described earlier (in the background section for the R21 phase), inclusion of a specific exon (v6) appears to correlate with the metastatic potential of a tumor cell (Gunthert et al., 1992). Thus,

CD44 alternative splicing may be a marker not only for a tumor, but also a indicator for the development stage of a tumor. These findings strongly suggest that, with advanced multiplex methods, it should be possible to find specific unusual patterns of alternative splicing associated with particular stages in the induction and progression of malignancy. Our collaborative arrangement with surgical pathologists at the UCSD Cancer Center who have access to large quantities of fresh and frozen tissue specimens provides a unique opportunity for successful conduct of the proposed studies.

Potential applications in multiple research fronts: Because of the unique design of our experimental approach using target-specific oligonucleotide sequences attached to unique zip-code sequences, the technology should be applicable to many other types of assays. For example, one might select specific sequences in a group of transcripts and monitor their expression in cells and tissues. Although a variety of array-based assays are available for gene expression analysis, our technology may be more sensitive and is not restricted to a particular set of targets, and thus, may be more appealing to individual investigators. For example, T cells and B cells undergo a series of DNA recombination events during their maturation and activation. To monitor these developmental and differentiation events, scientists currently rely on Northern blot analysis using oligonucleotides complementary to specific combined VDJ regions to monitor the expression of rearranged immunoglobulin and T cell receptor genes, which requires a large amount of cellular RNA. The technique described in this proposal can be easily adapted to monitor VDJ recombination in a small amount of sample. In another example, it was recently discovered that human neural cadherin-like cell adhesion protein genes are organized in a similar fashion to the T cell receptor and immunoglobulin genes, containing variable and constant regions (Wu and Maniatis, 1999). Remarkably, while each transcript uses a unique set of exons for variable regions, a large number of transcripts share the same three exons for the constant region. It is unclear currently whether the joining is through DNA recombination, cis-splicing or trans-splicing (see discussion by Wu and Maniatis, 1999). Because each such cadherin-like molecule is thought to function in a specific neuron, it will require a highly sensitive assay to study such a cell-specific event. The technique developed here allows the selection of specific junctional sequences to detect a specific set of those molecules. It will be a powerful tool in determining how these recombination/splicing events take place and control the function of these cadherin-like molecules in specific neurons.

Development of a general platform for both small and large-scale applications: Because our experimental design is based on the use of universal arrays, it combines maximal flexibility with minimal cost in small scale applications for individual investigators. It also allows parallel analysis of a large number of alternative splicing events in a large number of samples. Thus, this scheme may be modified to establish an automated procedure for high throughput drug screening, which is a unique technical advantage of the array technology developed at Illumina. Our collaboration therefore may lead to further development based on a specific set of alternative splicing events for drug screening in the future.

c. Preliminary Studies/Progress Report

We plan to reach the milestones set for the R21 phase in about one year. If successful, it will provide strong preliminary results to justify the transition of the project to this R33 phase.

d. Research Design and Methods

Aim 1: Database construction

The project will begin with the selection of specific alternative splicing events to initiate the assay development described in the R21 phase. In the R33 phase, we will enlarge the scope of our assay and

apply the technology to scientific research and disease classification in order to demonstrate its usefulness and potential. To continue our experimental progress, we will need computer-assisted target identification, data analysis, and data documentation in order to pursue the alternative splicing research in a systematic way. Our goal is to construct a database which contains both reported alternative isoforms of individual genes and our own experimental data on the expression of these isoforms. The database will be available over the internet to allow access to the national scientific community, and more importantly, to permit the scientific community to assist in adding information to the database (community-based curation).

Inadequate information from currently available alternative splicing databases: We will take advantage of previous efforts in database construction to speed up our target selection process. A few alternative splicing databases of varying quality have already been constructed. One was recently published (Gelfand et al., 1999), which is derived from reclassification of the Swissprot protein database (currently, it contains about 1700 proteins sequences from various species). We examined the database and found that it does not contain useful information for our purpose because the database is assembled based only on clustering of closely related proteins sequences, often merging members of multigene families into clusters. Thus, those clusters may not correspond to alternatively spliced genes. Separately, the Jackson laboratory is maintaining a publicly accessible mouse alternative splicing database, which is updated weekly. At present, the database contains about 800 genes that have been reported to produce alternative transcripts. However, not only is the database limited to the mouse, but it also omits critical sequence information and contains no annotation detailing the structure of the alternatively spliced products. Thus, these databases are not suitable for our purpose.

Our current effort in database construction: Rather than waiting for someone else to construct a high quality database, in June of 1998 Dr. Yeakley in our group started a collection of alternatively spliced human genes reported in the literature, and has since collected information on about 650 genes. While in the process of annotating these reported alternative splicing events, we learned recently that Dr. Michael Q. Zhang at Cold Spring Harbor Laboratory has been building an alternative splicing database (ASEDB) in collaboration with Dr. Stefan Stamm at the Max-Planck Institute and Dr. Kenta Nakai now at University of Tokyo, and the database will be published soon. This work was started in 1993 when Drs. Stamm and Zhang were investigating neural specific alternative splicing (Stamm et al. 1994). So far, their database has the highest quality because it is based on manual collection and annotation from the literature. Currently, ASEDB has 848 entries, representing 33 species including human, mouse, chicken and fruit fly, and 32 tissue types. Six alternative splicing mechanisms are documented in the database: mutual exclusion, cassettes, alternative 3', alternative 5', alternative 3' with polyadenylation and retained intron. In addition, it includes information about splicing enhancers and silencers found in exons or introns, which have been shown to enhance or inhibit the splicing activity of neighboring splice site, respectively. Computer programs are provided to analyze the properties of alternatively spliced exons. *Histogram of scores* displays the distribution of consensus scores of 5' and/or 3' splice sites. *Distribution around the splice site* displays the nucleotide composition around the 5' and 3' splice sites. *Exon length* shows the distribution of exon length for the selected group. *n-tuple* computes the frequencies of n-tuples (oligomers of sizes 2-5 bp). Various searching programs enable browsing the data from different aspects. Prediction tools such as exon finder and enhancer/silencer predictor are also provided. ASEDB is now only accessible to a few participating labs (through passwords). Since ASEDB contains only small subset of the available data , it needs to be substantially expanded. More importantly, it has to be refined for our purpose by adding missing information, such as the gene structure (if known), patterns of alternative splicing, sequences of alternative

intron-exon junctions and mutations/SNPs (for later disease studies). Therefore, we plan to join forces with Dr. Zhang's group to construct an updated new database: a human/mouse alternative splicing database.

In the meantime, we are selecting an initial set of targets from our list and from Dr. Zhang's database for assay development based on the following criteria: (1) Only mouse and human sequences will be collected. (2) Only genes that have cDNA and genomic sequences available will be included. (3) Gene expression and alternative splicing patterns are relatively well documented in the literature. (4) Some biological functions have been assigned to alternative transcripts. The examples listed in Table I represent some of these initial selections.

A strategy to collect published alternative splicing events: Dr. Michael Gribskov at the San Diego Supercomputer Center is an expert on sequence alignment and pattern recognition. As a local expert with a long term interest in mRNA isoforms, he will join forces with us for database construction. He is also playing a key role in establishing functional genomics training programs at UCSD for undergraduate and graduate students. In keeping with that effort, we plan to work together to recruit a sizable number of undergraduate students at UCSD, train them to extract the required information from the literature, to combine with information from cDNA and genomic sequence databases, and have them report to Dr. Yeakley in my lab for review and quality control. Since splicing information is uniformly annotated in the public sequence databases, and there are many types of errors in the existing annotation, manual updating and data integrity checking are absolutely necessary. We realize that collection and annotation of just the published alternative splicing examples will be a major undertaking and that eventually a database must be built allowing community data entry and annotation in order to minimize duplicated efforts and to maximize its utility. During this project, we will experiment with distributive database format and networked annotation tools between UCSD and CSHL.

Development of computational tools for database construction and presentation: While we focus our effort on bench work, Drs. Gribskov and Zhang will collaborate to design and implement a relational database for alternative splicing based on carefully curated information from the scientific literature (beginning with the 650 references already collected in our group). The development of this database will be based on a common data model, possibly represented as an XML DTD, that will be jointly developed by the UCSD and CSHL groups. The database will be mirrored at both UCSD and CSHL sites. All application tools will be web-based, which will be available to the international scientific community over the internet both as a resource, and to act as a means of collecting information directly from research scientists.

Dr. Zhang plans to continue using *mySQL* as the RDBMS and using the Apache web server on a Unix platform. Program tools will be a combination of Perl/CGI, TK/Perl and Java. Dr. Zhang will also develop a stand-alone graphical user interface for browsing the sequence data. It will be a modified version of SEView that can represent known or predicted sequence elements through a WWW browser. Independence from the source data's format will be achieved through description language and ad hoc translators, which will make the system versatile and flexible.

To facilitate manual collection and annotation of published data on alternative splicing, Dr. Gribskov plans to develop computational tools to assist in automatically extracting information from the literature and for dynamically linking information in the alternative splicing database to other electronic data. These tools will provide a simple "fill-in-the-blanks" form for entering precise information about experimentally observed alternative splicing systems. Such tools are critical to implementing a high quality database as they ensure that the correct information is acquired, and they enforce a restricted vocabulary used to describe the alternatively spliced genes. In turn, the restricted vocabulary is essential for efficient searching and retrieval

of information from the database. The database will also link each entry to related cDNA and EST sequences. Graphical tools depicting the alternative splicing patterns will be implemented to provide an easily understandable interface for laboratory biologists. The UCSD group is experienced with the *mySQL* RDBMS and which will further enhance the ability of the UCSD and CSHL groups to collaborate.

Long term approaches to database construction: Dr. Zhang's group has initiated a major effort in the integration of existing splice site databases. In 1990, Stephens and Schneider (1990) built a human splice junction database that presently consists of approximately 1800 non-redundant, experimentally verified donor and 1700 acceptor sites. Unfortunately, the nomenclature for identifying entries in Genbank (from using locus names to accession numbers) and other entry changes render this data unusable to the public. In a recent EBI effort (Thanaraj 1999), a clean data set of EST-confirmed human splice sites has been created which contains (1) 619 donor sites; (2) 623 acceptor sites; of these (3) 392 pairs are flanking constitutively spliced exons and (4) 209 pairs flanking alternatively spliced exons. Dr. Zhang's team will extract useful information from these database resources and integrate it to his database (ASEDB) maintained at the Cold Spring Harbor Laboratory, which will continue to be updated from our collection of alternatively spliced human genes and the mouse genes listed at the Jackson Laboratory.

A more powerful approach is to analyze ESTs and to detect alternatively spliced mRNA directly. Several efforts are underway, including the Unigene project at NCBI, the GeneIndex project at TIGR, and the STACK project at the South African Nation Bioinformatics Institute (see Bouck et al. 1999 for a review and comparison of these indexes). The approach in use at STACK appears best suited to detecting alternatively spliced genes because possible splicing alternatives are tracked and recorded during the clustering process. Briefly, the STACK approach begins by clustering overlapping ESTs using a high-throughput matching algorithm called D2-cluster. The ESTs are carefully filtered to remove possible primer and vector sequences, and assembled into clusters comparable to "contigs" in shotgun sequencing projects. During this process, alternatively spliced genes are detected and preserved for analysis. Dr. Gribskov has been involved in the STACK project and the STACK software is currently being imported to a multiprocessor Sun architecture and will be available for the proposed project in this R33 phase for target selection.

Dr. Gribskov's team is highly experienced in sequence comparison and database implementation. Their current projects include the protein kinase resource (http://www.sdsc.edu/kinases), the plant phosphoprotein functional genomics database (http://www.sdsc.edu/mpr/plant_pk). and homophila, (http://www.sdsc.edu/mpr/homophila) a database that links genes involved in human disease to their homologs in *Drosophila*. Many of these disease-related genes are likely to show alternative splicing, and this information will allow us to design assays for their alternative splicing for disease classification and prediction as described in Aim 3 of this proposal.

In conclusion, this multilateral collaboration takes advantage of combined and highly complementary expertise to approach a fundamental biological problem. We are committed to working together to construct an alternative splicing database that is coupled with experimental data.

Aim 2: Parallel analysis of alternative splicing in cell and animal models

Studies proposed in this aim represent the refinement and expansion of the experiments described in the R21 phase. In addition, we plan to take advantage of the biological systems that are already established in our laboratory as a test case for the new technology to demonstrate the usefulness and robustness of the proposed approach.

Systematic expansion of assay targets: In this R33 phase, we plan to enlarge the scope of our alternative splicing by including as many alternative splicing events reported in the literature as possible. For practical reasons, we will examine 50 to 100 genes at a time in the same manner as in the R21 phase for the first set of genes. Two alternative experimental strategies (the one oligo vs. two oligo approaches) will be tested in the R21 phase. Through the proposed experiments, we will establish a standardized procedure for large scale analysis in this R33 phase. Our experiments will require a large number of DNA oligos, and their quality is critical. After considering various options for oligo preparations, we conclude that it is most economical to contract oligo synthesis to MWG Biotech, where, at no extra charge, each oligo synthesized is purified using their proprietary technology to eliminate all failure products, and the purity of every oligo is determined by mass spectrometric analysis. These quality controls are essential in our application to minimize nonspecific hybridization and interference by failure products during oligo synthesis.

Cell line studies: Because many genes are differentially expressed in different cells and tissues, we will conduct parallel analyses of each set of alternative splicing events on a panel of human cell lines, including HeLa (cervical epithelial), 293 (kidney epithelial), WI-38 (normal fibroblast), Jurkat (lymphoblast), Weri-1 (retinoblastoma), and MG-63 (osteosarcoma). We have chosen these cells because they are routinely cultured in our lab and they come from a variety of tissues. Obviously, the assay may be extended to many other human cell lines. These studies will allow comparison of information from cell lines with that derived from human tissues. In this regard we plan to prepare total RNA from various normal human tissues and obtain a tissue-specific alternative splicing profile for individual genes. These normal tissue-specific alternative splicing profiles will serve as references to corresponding changes in a specific disease when applying the assay to human disease classification and prediction, as described in the next aim.

For basic research, we are most interested in applying the assay to addressing how alternative splicing is regulated. Like transcriptional regulation, many alternative splicing events may be subject to control through various signal transduction pathways. We plan to first determine alternative splicing events that are regulated (individually or co-regulated) by specific growth factors and/or responsive to the activation of specific kinase pathways. By using translation inhibitors, we will determine whether a regulated splicing event is dependent on ongoing protein synthesis. We are particularly interested in identifying alternative splicing events that are responsive to a stimulus in the absence of protein synthesis, which will allow us to establish a model system for studying the regulation of alternative splicing by signaling.

Animal studies: Our lab has been studying a number of splicing factors called SR proteins, which are not only essential for constitutive splicing, but also involved in regulated splicing (see a review by Fu, 1995). More recently, our lab has identified and characterized a family of kinases that are specific regulators of the SR splicing factors. However, all experiments so far (by us and by our colleagues in the splicing field) that address the function of these mammalian splicing factors and regulators have been performed in vitro or in transfected cells. It is therefore unclear how these splicing regulators function in animals and whether they are part of regulatory pathways in normal development and/or involved in certain diseases. To address these important biological questions, we have been developing animal models to investigate the in vivo functions of various splicing regulators, and the preliminary results are highly encouraging. For example, it appears that knocking out just one copy of the SC35 gene (an SR family member) suppresses female development in mice. When both copies of the gene are deleted in thymus (using the Lox-Cre inducible knockout system), T cell maturation appears to be impaired during the transition from double positive $CD4^+CD8^+$ to single positive $CD4^+CD8^-$ and $CD4^-CD8^+$ stages (unpublished observations). These experiments demonstrate the importance of the SC35 gene in development and differentiation. However, the mechanisms leading to these

phenotypes remain to be elucidated. Using the parallel analysis strategy proposed here, we will be able to address alternative splicing events that occur in the absence of a specific regulator. This represents a long term goal for our basic research.

Aim 3: Applications in cancer research: developing new tools for disease classification

The main goal of this proposal is to develop a novel system for molecular classification of cancer. As described in Background/Significance, alternative splicing has been linked to development, differentiation, and disease. A large number of alternative splicing events are frequently associated with human cancer. Based on these observations, it should be feasible to use a sizable group of alternative splicing events to classify and predict different kinds of cancer or the same type of cancer in different stages.

Target selection: We plan to select targets for disease classification and prediction in two phases. First, we plan to focus on alternative splicing events that have already been linked to a disease(s). As listed in Table I in the R21 phase, we have already selected a number of genes that appear to produce disease-specific isoforms. For example, alternative splicing of the AML1 gene appears to be involvedwith a subclass of acute myelogenous leukemia (Tanaka et al., 1995) and alternative splicing of the gene for PSA (prostate specific antigen) seems to correlate with prostate cancer (Su et al., 1995). In this R33 expansion phase, we will continue to enlarge the sample size. For instance, CD44 alternative splicing was not selected in the R21 assay development phase because it contains ten distinct variable exons and therefore will produce over 1000 combinations. Once the assay is established in the R21 phase, CD44 will be selected as one of the primary targets for cancer classification because it has been shown to be one of the most reliable tumor indicators. Here, we are particularly interested in testing the hypothesis that inclusion of certain CD44 exons is directly linked to a tumor and/or metastatic potential of the tumor by a large scale parallel analysis of alternative splicing. By examining many reported disease-associated alternative splicing events in this early R33 phase, we hope to nail down some critical disease-associated alternative splicing events and use them for disease classification and prediction.

Our next target selection phase will include other reported alternative splicing detected in our database construction efforts. We believe that it is important to conduct a comprehensive search, rather than solely focus on disease-associated genes because specific alternative splicing events may be relevant to certain diseases, but they may not yet have been tested or reported in the literature. This is illustrated by the recent publication on ALL/AML distinction by gene expression monitoring (Golub et al., 1999). An elevated expression of some genes in ALL or AML appear to make sense, but many others have no apparent connection to any kind of leukemia based on previous functional studies.

Preclinical and clinical applications: By collaboration with Dr. David Tarin's group, we plan to focus on breast and prostate cancer because of their availability through the UCSD Cancer Center. Dr. Tarin's group has already compiled tissue banks containing hundreds of breast cancer samples. Prostate cancer samples are also being systematically collected for their histopathological studies. Many of these cases have matched tissue samples obtained from the primary tumor, normal tissue, and metastases. Therefore, we are in a posittion to identify tumor specific expression patterns and differences related to tumor behavior.

Once a battery of specific targets are found, we plan to take advantage of the Molecular Pathology Shared Resource at UCSD to obtain dissected malignant tissue by Laser Capture Microscopy so that we can identify specific cell and tissue types that give rise to detected alterations in alternative splicing. In addition, we will determine the sensitivity of our assay by mixing normal and cancerous tissues in different ratios. If an assay can be established to detect tumor-specific alternative splicing events in a very small amount of

diseased tissue in the presence of a large amount of normal tissue, it will find wide application in clinical tumor diagnosis. A sensitive PCR-based assay for measuring human telomerase has demonstrated such power (Kim et al., 1994). It was shown that 10 immortal cells in a pool of 10,000 normal cells were sufficient for detection of telomerase activity. Thus, telomerase could be detected if 1 in 1000 cells was immortal, which is therefore highly useful in examining tumor samples without preparing a homogeneous cell population for the assay. Since telomerase activity is associated with most tumors, it is a good indicator, but cannot be used to classify tumors. Our goal is to develop an equally sensitive assay which can be used not only to detect a tumor, but also to identify the type and stage of the tumor. This method may be particularly useful in diagnosis of prostate cancer, which may have similar pathological grades, but follow highly variable clinical courses. A molecular classifier would therefore be highly beneficial in devising a therapeutic strategy.

Once the assay is developed and optimized in our laboratories, we will collaborate with physicians to explore its application in clinical diagnosis and compare the results with those by conventional pathological analysis. For this purpose, fresh material will be collected through the UCSD Cancer Center tissue procurement resource led by Dr. Wasserman and we have IRB approval for this.

Data analysis and presentation: Our experiments are designed to discover unique alternative splicing events that are strongly correlated with a particular type of disease. Because the array-based experiments will generate a large amount of data for analysis, we will need to develop tools to analyze and present the data. For this purpose, we will work closely with our collaborators, Drs. Zhang and Gribskov to construct a web site for raw data as well as for methodological details. Dr. Gribskov's group is actively working to develop new approaches in this area. These approaches will be integrated with an online database, making the microarray experimental results themselves available. This database will conform to relevant standards for exchange of array data, currently under discussion in the microarray community as defined and accepted.

To determine the statistical significance of a correlative event, we will need a set of analytical tools to treat raw data. A paper by Golub and his colleagues was recently published, in which they described one such set of calculation formulae based on "neighborhood analysis" (Golub et al., 1999). These formulae will be very useful for treating our experimental data. In carrying out such neighborhood analysis, we will first determine the level of correlation [$P(g, c)$] of individual alternative splicing events with a particular cancer type in comparison with corresponding healthy tissue. This will allow us to identify a group of informative genes that can be used as predictors. We will initially focus on the identification of such predictors for breast and prostate cancer using CD44 and PSA alternative splicing. It is not known, however, whether these alternative splicing events are sufficient to serve as predictors for specific cancers. For this purpose, the predication strength (PS) for a set of informative alternative splicing events will be calculated according to Golub et al. (1999).

Because the Golub method is restricted to comparison between two types of diseases, it may not be suitable for general cancer classification and a multivariate statistical classification will have to be developed for analysis of alternative splicing data from heterogeneous cancer samples. Dr. Zhang has extensive experiences in analyzing large-scale gene expression data (Zhang, 1999a). He was the first to combine array clustering and regulatory element analysis (Zhang, 1999b). He has used both hierarchical Pearson correlation clustering (Spellman et al., 1998) and Kohonen Self-Organizing Map (SOM) clustering algorithms (Zhu and Zhang, 1999). Very recently, in collaboration with physicists, he has applied a novel SPC (Super-Paramagnetic Clustering) algorithm to study genome-wide gene expression profiles (Getz et al. 1999). In order to experiment with a multitude of multivariate statistical classification techniques on alternative splicing array analysis, Dr. Zhang will test CART (Classification and Regression Trees) technology

(Breiman et al., 1984) because it has the potential to identify a few most relevant mRNA transcripts from a background of thousands for sensitive and robust molecular classifications. CART is an unsupervised technology and it will automatically perform cross-validations for statistical significance.

To develop predictors for various cancers, Dr. Gribskov also has a wealth of experience in this area because the procedure is similar to those used to develop predictors for gene functions based on a set of (unaligned) training sequences. In essence, it should be possible to adapt the unsupervised learning procedure already in use for discovering motifs in protein and nucleic acid sequences to discover patterns of alternative splicing associated with different cancer types and/or the same type of cancer at different malignant stages. The similarity is more than superficial; sequence motifs are represented as an array of probability vectors where each vector describes the probability of finding each amino acid residue (or base) at each position of the motif. In the case of gene expression or alternative splicing, the predictive pattern is an array of probability vectors where the vectors are the probabilities of each alternatively spliced transcript. Expectation-maximization and related algorithms should therefore be applicable to these data just as they are to protein sequence data. Other standard clustering technologies such as the k-means algorithm, principal components analysis, and less widely used techniques such as adaptive clustering are also likely to be useful in this analysis. Drs. Gribskov and Zhang will work closely to develop a standard set of classification tools for quick and accurate disease diagnosis.

In conclusion, the feasibility of molecular disease classification and prediction has been demonstrated using the method ofmonitoring overall gene expression. The method described in this proposal is based on RNA splicing, which is catalyzed by a distinct machinery from transcription in cells. Alternative splicing represents another level of control for gene expression and changes in alternative splicing may take place without a global alteration of gene expression. Thus, our assay will be complementary to those based on transcription and a combinatory use of different assay systems may be more accurate and robust in future disease classification and prediction.

Aim 4: Assay development for both small and large-scale applications.

Unlike conventional cDNA microarrays and standard gene expression chips, the method described in this proposal does not need a massive number of PCR amplifications of specific targets nor preparation of specialized arrays. Thus, the technology can be adopted by individual investigators. For this to happen, it will be necessary to ensure robust, reproducible, and highly accurate results in a wide variety of experimental settings. To make this a reality, the fourth aim will focus on assay development, capitalizing on the investment in the first three aims of the R33 phase, which set the stage for large scale analysis of alternative splicing events. The ability to carry out many assays in parallel will allow us to develop efficient algorithms for assay optimization.

For assay development, the alternative splicing assays will be grouped into functionally relevant sets that will be useful for focused research (e.g. based on homology; or association with a particular pathway or disease; or expressed in particular tissues of interest), as well as for more global studies. Genes grouped according to their structure and functional relationships will include kinases, phosphatases, oncogenes/tumor suppressor genes, cell surface receptors, etc. The assays will then be optimized to achieve a high degree of reliability and specificity within a set. For application in basic research, we plan to gradually expand our assay to include alternative splicing events reported in the literature and then extend to computer-predicted alternative splicing of genes. Based on our results, we will prepare combinations of these targets for different applications and make them available for other investigators.

By working closely with our collaborators at Illumina, we would also like to develop an experimental procedure for large scale applications. For instance, Illumina will soon develop their array sets in microtiter

plates to allow parallel analysis of a large number of targets from multiple samples or different experimental conditions. This Array of Arrays™ format will be powerful in high throughput drug screening. To conduct such a parallel alternative splicing assay, we plan to test the following experimental strategy: (1) Grow chosen cells on microtiter plates for treatment with different drugs. (2) Conduct in situ cell lysis followed by addition of specific target oligos plus oligo (dT) magnetic beads. (3) Insert a strong magnet underneath each plate and carry out stringent washing. (4) Add universal primers and transfer the plates for PCR amplification. (5) Apply PCR products to the array plates for hybridization, staining, and scanning. If experimental conditions are worked out, all the procedures can be handled by a robot. In this R33 phase, we would like to test the feasibility of this experimental scheme. If the technology can be developed, it will be not only powerful for high throughput drug screening in industry, but also very useful for handling a large number of clinical samples in disease diagnosis and classification.

Milestones for the R33 phase:

Although it is difficult to accurately predict the outcome of the R33 phase research, we are confident in achieving the following milestones at the conclusion of this project:

1. A comprehesive and high quality database for alternative splicing in humans and mice will be created, which will become an international resource for biomedical research.
2. The array experimental procedures will be standardized for large scale analysis of alternative splicing. Our assay strategy can be adapted in many other array applications.
3. A large number of alternative splicing events reported in the literature and/or deduced from EST clustering will be re-investigated using array experiments, and the experimental data coupled with the database. The experimental data and related information will be accessible to the research community through internet.
4. This genomic approach will generate novel insights into the regulation of alternative splicing through our cell line and animal studies.
5. We will develop cancer classification tools based on mRNA isoforms. If successful, these tools should find wide application in cancer diagnosis and prediction.
6. Array technology will be further developed toward an automated system for large scale analysis and drug screening.

Time table and responsibilities: A brief summary.

As the PI, I will be responsible for overall supervision of the project and integration of collaborative efforts. In the R21 phase, I will actually work at the bench, with Dr. Yeakley in my group, to carry out the initial experiments and Dr. Jian-Bing Fan at Illumina will perform all array experiments. We plan to conclude the R21 phase in about a year. In the following three years, during the R33 phase, Dr. Yeakley, Dr. Fan, and I will continue to function as the core experimental team. In addition, Dr. Yeakley's responsibility will include organizational efforts in annotation of reported alternative splicings. At this point, we will need an experienced technician to join the team. Dr. Gribskov's team will develop computational tools for experimental design. They will play a major role in organizing efforts in database building and conduct EST analysis. Dr. Zhang' team will integrate various splice site databases and develop a database graphical interface. Drs. Zhang and Gribskov will collaborate closely to build the alternative splicing database, implement various electronic and experimental data, and develop tools for array analysis and for disease classification and prediction. Dr. Tarin will coordinate the clinical activities. Drs. Wasserman and Goodison in his group will provide normal and disease tissue samples and conduct morphological analyses to characterize their malignant grades and stages. They will play a key role in cancer classification and prediction.

e. Human Subjects: We are not going to directly collect human tissues from patients for our studies. Normal and cancer samples will be provided to us by our collaborators at the UCSD Cancer Center where collected tissues are stored. In the future, if we are at a stage to examine tissue samples from patients, we will work closely with physicians at the UCSD Cancer Center to obtain cancer samples and all procedures will comply with relevant university and government guidelines in handling human tissues.

f. Vertebrate Animals: This project will involve collecting mouse tissues. Procedures and precautions which will be taken in handling animals are described in our animal protocol B99116, which has been approved by the Animal Subjects Committee at the University of California, San Diego. Specifically:

1. We will use knockout mice we are generating in my laboratory at UCSD to provide tissue samples for array analysis. These mice are derived from BL/6 and 129 strains of various ages and both sexes in which a splicing factor or regulator is deleted. We estimate that we will need about 100 to 200 mice per year for the experiments proposed in this project.
2. These mice were created for other studies in our lab under the animal protocol B99116 and these animals happen to be valuable as a model system for the proposed studies in this project. Specifically, we will prepare fibroblast cells from embryos to examine the alternative splicing pattern. We will also collect T cells from blood, spleen, and thymus for similar studies.
3. The University of California at San Diego has full-time on-site veterinarians who will have responsibility for overseeing all laboratory animal care.
4. Since all of our experiments will be conducted using animal tissues, we will avoid any discomfort and injury to the mice before sacrifice.
5. We will use CO_2 gas as the method of choice for euthanasia before cervical dislocation. This method is selected because it is humane and simple and is consistent with the recommendations of the Panel on Euthanasia of the American Veterinary Medical Association.

g. Literature Cited:

Alnemri, E.S., Fernandes-Alnemri, T., & Litwack, G. (1995) Cloning and expression of four novel isoforms of human interleukin-1 beta converting enzyme with different apoptotic activities. *J. Biol. Chem.*, **270**:4312-4317.

Balleine, R.L., Hunt, S.M., & Clarke, C.L. (1999) Coexpression of alternatively spliced estrogen and progesterone receptor transcripts in human breast cancer. *J. Clin. Endo. Metab.*, **84**:1370-1377.

Baruch, A., Hartmann, M., Yoeli, M., Adereth, Y., Greenstein, S., Stadler, Y., Skornik, Y., Zaretsky, J., Smorodinsky, N.I., Keydar, I., & Wreschner, D.H. (1999) The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein. *Cancer Res.*, **59**:1552-1561.

Boise, L.H., González-García, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nuñez, G., & Thompson, C.B. (1993) bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death *Cell*, **74**:597-608.

Bost, F., McKay, R., Bost, M., Potapova, O., Dean, N.M., & Mercola, D. (1999) The Jun kinase 2 isoform is preferentially required for epidermal growth factor-induced transformation of human A549 lung carcinoma cells. *Mol. Cell. Biol.*, **19**:1938-1949.

Bourinet, E., Soong, T.W., Sutton, K., Slaymaker, S., Mathews, E., Monteil, A., Zamponi, G.W., Nargeot, J., & Snutch, T.P. (1999) Splicing of alpha 1A subunit gene generates phenotypic variants of P- and Q-type calcium channels. *Nat Neurosci*, **2**:407-415.

Bouck. J., Yu, W., Gibbs, R., & Worley, K. (1999). Comparison of gene indexing databases. *Trend Genetics* 15:159-162.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees, Pacific Grove: Wadsworth.

Burge, C. B., Padgett, R. A., & Sharp, P. A. (1998) Evolutionary fates and origins of U12-type introns. *Mole. Cell* 2:773-785.

Cáceres, J. F., Stamm, S., Helfman, D. M., & Krainer, A. R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science*, 265:1706-1709.

Cargill, M. et al., (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* 22:231-238.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. *Science* 274:610-614.

Cui, J.Q., Wang, H., Reddy, E.S., & Rao, V.N. (1998) Differential transcriptional activation by the N-terminal region of BRCA1 splice variants BRCA1a and BRCA1b. *Oncol. Rep.*, 5:585-589.

Daoud, R., Da Penha Berzaghi, M., Siedler, F., Hübener, M., & Stamm, S. (1999) Activity-dependent regulation of alternative splicing patterns in the rat brain. *Eur. J. Neurosci.*, 11:788-802.

Du, K., Leu, J.I., Peng, Y., & Taub, R. (1998) Transcriptional up-regulation of the delayed early gene HRS/SRp40 during liver regeneration. Interactions among YY1, GA-binding proteins, and mitogenic signals. *J. Biol. Chem.*, 273:35208-35215

Fan, J-B. et al., (1999). Scaleable parallel genotyping using high-density oligonucleotide tag arrays. Manuscript in preparation.

Ferns, M., Hoch, W., Campanelli, J.T., Rupp, F., Hall, Z.W., & Scheller, R.H. (1992) RNA splicing regulates agrin-mediated acetylcholine receptor clustering activity on cultured myotubes. *Neuron*, 8:1079-1086.

Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., & Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767-773.

Fu, X-D. (1995). The superfamily of arginine/serine-rich splicing factors. *RNA* 1:663-680.

Gelfand, M. S., Dubchak, I., Dralyuk, I., & Zorn, M. (1999). ASDB: database of alternatively spliced genes. *Nucleic Acids Research* 27:301-302.

Ge, H., Zuo, P., & Manley, J.L. (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell*, 66:373-382.

Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G., & Barany, F. (1999). Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol. Biol.* 292:251-262.

Getz, G., Levine, E., Domany, E., & Zhang, M. Q. (1999). Super-paramagnetic clustering of yeast gene expression profiles. Submitted.

Gimona, M., Watakabe, A., & Helfman, D.M. (1995) Specificity of dimer formation in tropomyosins: influence of alternatively spliced exons on homodimer and heterodimer assembly. *Proc. Natl. Acad. Sci.*, 92:9776-9780.

Golub, T. R. et al., (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531-536.

Goodison, S. & Tarin, D. (1998). Current status of CD44 variant isoforms as cancer diagnostic markers. *Histopathology* 32:1-6.

Gunthert, U., et al., (1992). A new variant of glycoprotein CD44 confers matastatic potential to rat carcinoma cells. *Cell* **65**:13-24.

Guo, Z., Liu, Q., & Smith, L. M. (1997). Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nature Biotech.* **15**:331-335.

Haase, H., & Morano, I. (1996) Alternative splicing of smooth muscle myosin heavy chains and its functional consequences. *J Cell. Biochem.*, **60**:521-528.

Hanstein, B., Liu, H., Yancisin, M.C., & Brown, M. (1999) Functional analysis of a novel estrogen receptor-beta isoform. *Mol. Endo.*, **13**:129-137.

He, Y., Smith, S.K., Day, K.A., Clark, D.E., Licence, D.R., & Charnock-Jones, D.S. (1999) Alternative splicing of vascular endothelial growth factor (VEGF)-R1 (FLT-1) pre-mRNA is important for the regulation of VEGF activity. *Mol. Endo.*, **13**:537-545.

Jiang, Z.H., Zhang, W.J., Rao, Y., & Wu, J.Y. (1998) Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors. *Proc. Natl. Acad. Sci.*, **95**:9155-9160.

Jiang, Z. H., & Wu, J. Y. (1999). Alternative splicing and programmed cell death. *Proc. Society for Exp. Biol. and Med.* **220**:64-72.

Kashiwada, K., Nishida, W., Hayashi, K., Ozawa, K., Yamanaka, Y., Saga, H., Yamashita, T., Tohyama, M., Shimada, S., Sato, K., & Sobue, K. (1997) Coordinate expression of alpha-tropomyosin and caldesmon isoforms in association with phenotypic modulation of smooth muscle cells. *J. Biol. Chem.*, **272**:15396-15404.

King, P.H. (1994) Hel-N2: a novel isoform of Hel-N1 which is conserved in rat neural tissue and produced in early embryogenesis. *Gene*, **151**:261-265.

Koltchine, V.V., Anantharam, V., Bayley, H., & Treistman, S.N. (1996) Alternative splicing of the NMDAR1 subunit affects modulation by calcium. *Brain Res. Mol.Brain Res.*, **39**:99-108.

Krecic, A. M. & Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* **11**:363-371.

Lin, Z., Lin, Y., Schorge, S., Pan, J.Q., Beierlein, M., & Lipscombe, D. (1999) Alternative splicing of a short cassette exon in alpha1B generates functionally distinct N-type calcium channels in central and peripheral neurons. *J. Neurosci.*, **19**:5322-5331.

Lipshutz, R.J. et al., (1999). High density synthetic oligonucleotide arrays. *Nature Genet.* **21**:20-24.

Lockhart, D. J. et al., (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotech.* **14**:1675-1680.

Lorson, C.L., Hahnen, E., Androphy, E.J., & Wirth, B. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci.*, **96**:6307-6311.

Luo, Z.D., Camp, S., Mutero, A., & Taylor, P. (1998) Splicing of 5' introns dictates alternative splice selection of acetylcholinesterase pre-mRNA and specific expression during myogenesis. *J. Biol. Chem.*, **273**:28486-28495.

Luqmani, Y.A., Mortimer, C., Yiangou, C., Johnston, C.L., Bansal, G.S., Sinnett, D., Law, M., & Coombes, R.C. (1995) Expression of 2 variant forms of fibroblast growth factor receptor 1 in human breast. *Intl. J. Cancer*, **64**:274-279.

Matsumura, Y. & Tarin, D. (1992). Significance of CD44 gene products for cancer diagnosis and disease evaluation. *Lancet* **340**:1053-1058.

Matsumura, Y., Hanbury, D., Smith, J., & Tarin, D. (1994). Non-invasive detection of malignancy by identification of unusual CD44 gene activity in exfoliated cancer cells. *BMJ* (Clinical Research Ed.), **308**:619-624.

Meredith, J. Jr., Takada, Y., Fornaro, M., Languino, L.R., & Schwartz, M.A. (1995) Inhibition of cell cycle progression by the alternatively spliced integrin beta 1C. *Science*, **269**:1570-1572.

Michael, K. L., Taylor, L. C., Schultz, S. L., Walt, D. R. (1998). Randomly ordered addressable high-density optical sensor arrays. *Analytical Chemistry* **70**:1242-1248.

Miki, T., et al., (1992). Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene. *Proc. Natl. Acad. Sci. USA* **89**:246-250.

Modafferi, E.F., & Black, D.L. (1999) Combinatorial control of a neuron-specific exon. *RNA*. 5:687-706.

Moller, D.E., Yokota, A., Caro, J.F., & Flier, J.S. (1989) Tissue-specific expression of two alternatively spliced insulin receptor mRNAs in man. *Mol. Endo.*, 3:1263-1269.

Moore, M. J. & Sharp, P. A. (1992). Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science* **256**:992-997.

Pennisi, E. (1999). Keeping genome databases clean and up to date. *Science* **286**:447-450.

Ogut, O., Granzier, H., & Jin, J.P. (1999) Acidic and basic troponin T isoforms in mature fast-twitch skeletal muscle and effect on contractility. *Am. J. Physiol.*, 276:C1162-1170.

Rafuse, V.F., & Landmesser, L. (1996) Contractile activity regulates isoform expression and polysialylation of NCAM in cultured myotubes: involvement of Ca2+ and protein kinase C. *J. Cell Biol.*, **132**:969-983.

Ratech, H., Denning, S., & Kaufman, R.E. (1997) An analysis of alternatively spliced CD45 mRNA transcripts during T cell maturation in humans. *Cell. Immunol.*, **177**:109-118.

Robertson, K.D., & Jones, P.A. (1999) Tissue-specific alternative splicing in the human INK4a/ARF cell cycle regulatory locus. *Oncogene*, **18**:3810-3820.

Ruberti, G., Cascino, I., Papoff, G., & Eramo, A. (1996) Fas splicing variants and their effect on apoptosis. *Adv. Exper. Med. Biol.*, **406**:125-134.

Sato, N., Hori, O., Yamaguchi, A., Lambert, J.C., Chartier-Harlin, M.C., Robinson, P.A., Delacourte, A., Schmidt, A.M., Furuyama, T., Imaizumi, K., Tohyama, M., & Takagi, T. (1999) A novel presenilin-2 splice variant in human Alzheimer's disease brain tissue. *J. Neurochem.*, 72:2498-505.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray *Science* **270**:467-470.

Shoemaker, D.D. et al. (1996). Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**:450-456.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. & Futcher, B. (1998). Comprehensive identification of cell cycle regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell.* 9:3273-3297.

Stamm, S., Zhang, M. Q., Marr, T., & Helfman, D. M. (1994). A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucl. Acid. Res.* **22**:1515-1526.

Stephens, R. M. & Schneider, T. D. (1990). Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.* **228**:1124-1136.

Su, S. L., Huang, I. P., Fair, W. R., Powell, C. T., & Heston, W. D. (1995). Alternatively spliced variants of prostate-specific membrane antigen RNA: ratio of expression as a potential measurement of progression. *Cancer Res.* **55**:1441-1443.

Tanaka, T., Tanaka, K., Ogawa, S., Kurokawa, M., Mitani, K., Yazaki, Y., Shibata, Y., & Hirai, H. (1997) An acute myeloid leukemia gene, AML1, regulates transcriptional activation and hemopoietic myeloid cell differentiation antagonistically by two alternative spliced forms. *Leukemia*, 11 Suppl 3:299-302.

Tanaka, T. et al., (1995). An Acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. *EMBO J.* **14**:341-350.

Thanaraj, T. A. (1999). A clean data set of EST-confirmed splice sites from Homo sapiens and standards for clean-up procedures. *Nucl. Acid. Res.* **27**:2627-2637.

Tsubari, M., Tiihonen, E., & Laiho, M. (1997) Cloning and characterization of p10, an alternatively spliced form of p15 cyclin-dependent kinase inhibitor. *Cancer Res.*, **57**:2966-2973.

Ulaner, G.A., Hu, J.F., Vu, T.H., Giudice, L.C., & Hoffman, A.R. (1998) Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts. *Cancer Res.*, **58**:4168-72.

Utsumi, T., Harada, N., Maruta, M., & Takagi, Y. (1996) Presence of alternatively spliced transcripts of aromatase gene in human breast cancer. *J. Clin. Endo. Metab.*, **81**:2344-2349.

Varani, L., Hasegawa, M., Spillantini, M.G., Smith, M.J., Murrell, J.R., Ghetti, B., Klug, A., Goedert, M., & Varani, G. (1999) Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc. Natl. Acad. Sci.*, **96**:8229-8234.

Vogelezang, M.G., Scherer, S.S., Fawcett, J.W., & ffrench-Constant, C. (1999) Regulation of fibronectin alternative splicing during peripheral nerve repair. *J.Neurosci. Res.*, **56**:323-333.

Wang, D. G. et al., (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**:1077-1082.

Wang, Y., Newton, D.C., & Marsden, P.A. (1999) Neuronal NOS: gene structure, mRNA diversity, and functional relevance. *Crit. Rev. Neurobiol.*, **13**:21-43.

Webster, N.J., Kong, Y., Sharma, P., Haas, M., Sukumar, S., & Seely, B.L. (1997) Differential effects of Wilms tumor WT1 splice variants on the insulin receptor promoter. *Biochem. Mol. Med.*, **62**:139-150.

Weil, R., Levraud, J.P., Dodon, M.D., Bessia, C., Hazan, U., Kourilsky, P., & Israël, A. (1999) Altered expression of tyrosine kinases of the Src and Syk families in human T-cell leukemia virus type 1-infected T-cell lines. *J. Virol.*, **73**:3709-3717.

Winzeler, E.A. et al., (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**:901-906.

Wu, Q. & Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97:779-790.

Zhang, G., Taneja, K. L., Singer, R. H., & Green, M. R. (1994). Localization of pre-mRNA splicing in mammalian nuclei. *Nature* **372**:809-812.

Zhang, M. Q. (1999a). Large scale gene expression data analysis: a new challenge to computational biologists. *Genome Research* 9:681-688.

Zhang, M. Q. (1999b). Promoter analysis of co-regulated genes in the yeast. *Computers and Chemistry* **23**:233-250.

Zhu, J. & Zhang, M. Q. (1999). Cluster, function and promoter: analysis of yeast expression array. Pacific Symposium Biocomputing'2000 (Accepted).

h. Consortium/Contractual Arrangements:

We will collaborate with Dr. Jian-Bing Fan of Illumina, Inc. for the array technology portions of this application. For database construction and analysis of generated array data, we will collaborate with Dr.

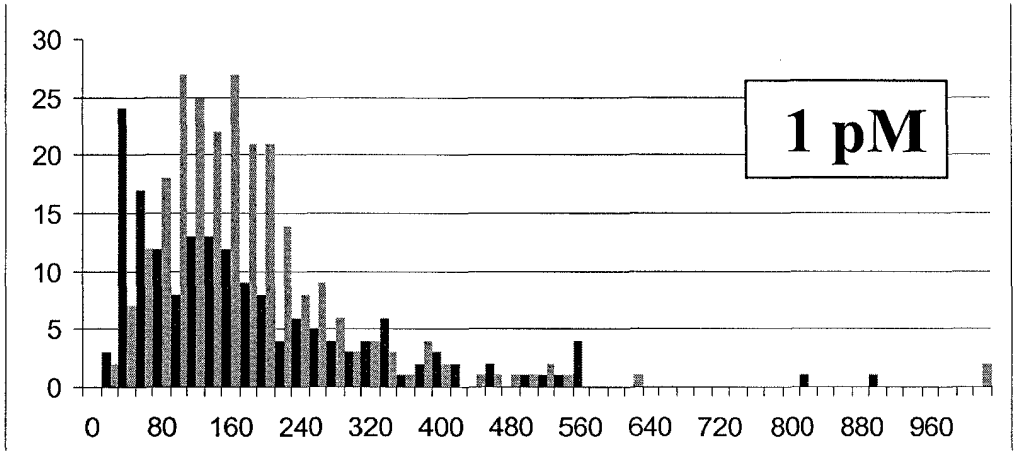
Michael Q. Zhang at the Cold Spring Harbor Laboratory. To simplify the budgetary issue, these collaborators will be treated as subcontractors.

In addition, we will collaborate with Dr. Michael Gribskov at the San Diego Supercomputer Center, who also holds a joint appointment in the Department of Biology, UCSD. This collaboration will allow us to work closely together at all stages of this project from database construction and mining to development of computational tools for experimental design and data analysis. We will also collaborate with Dr. David Tarin' group at the UCSD Cancer Center to take advantage of their expertise in tumor pathology, and through them, we will obtain all cancer samples for our analysis.

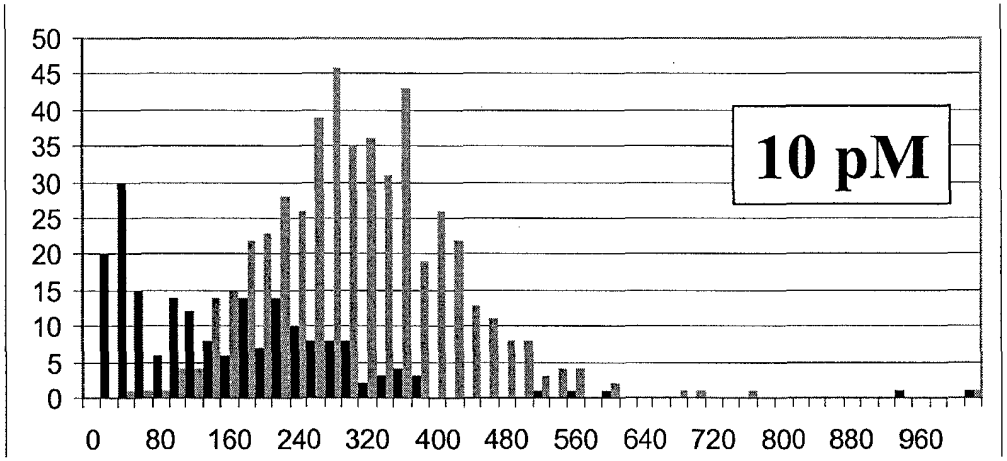
Letters of collaboration from all parties follow.

I. Consultants: None.

RNA Hybridization Specificity



~ 5 copies/cell

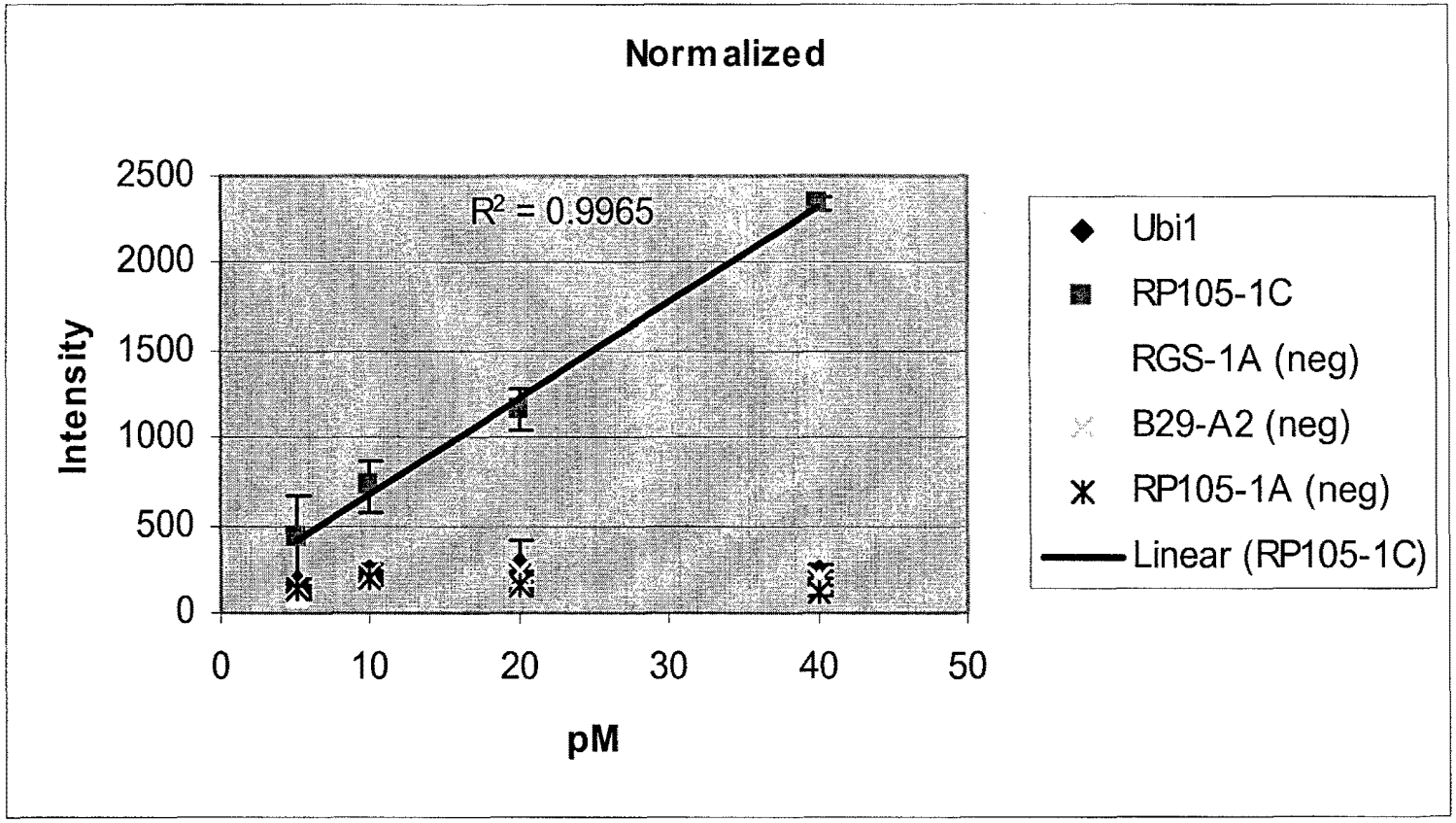


~ 50 copies/cell

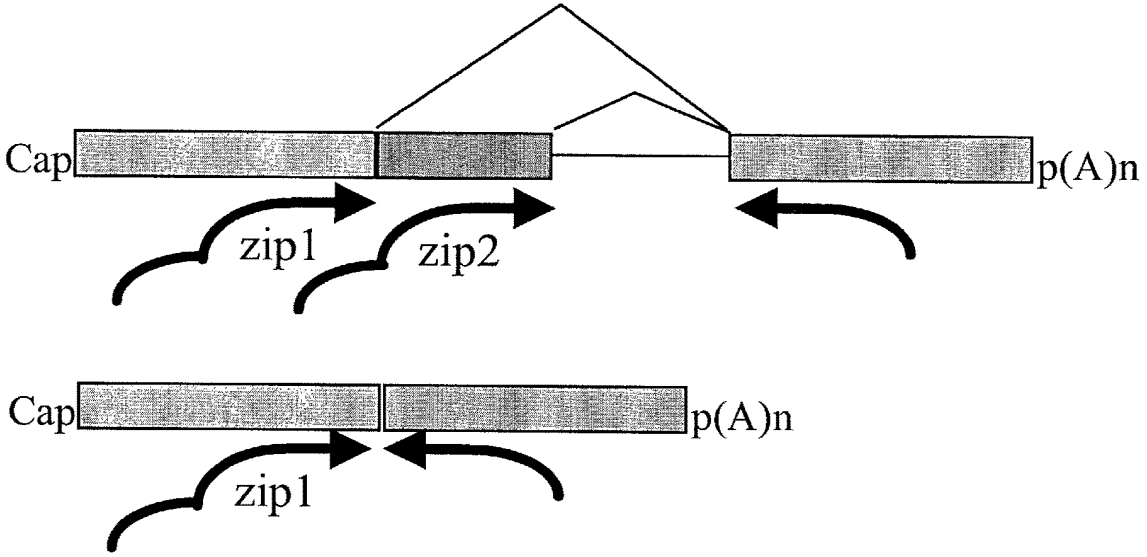
background

spiked

Detection of Differences in RNA Concentration



Array-Based Detection of Alternative mRNA Splicing



Step 2: PolyA selection

⇩

Step 3: RNA-dependent oligo ligation

⇩

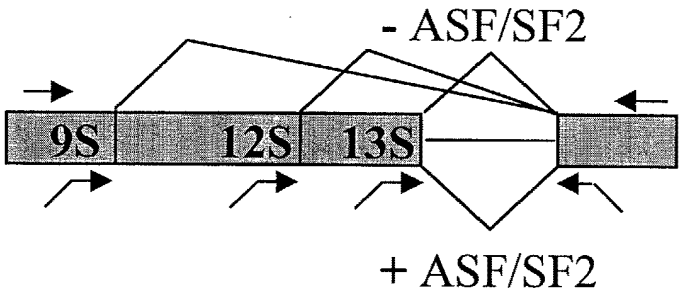
Step 4: PCR amplification using universal primers

⇩

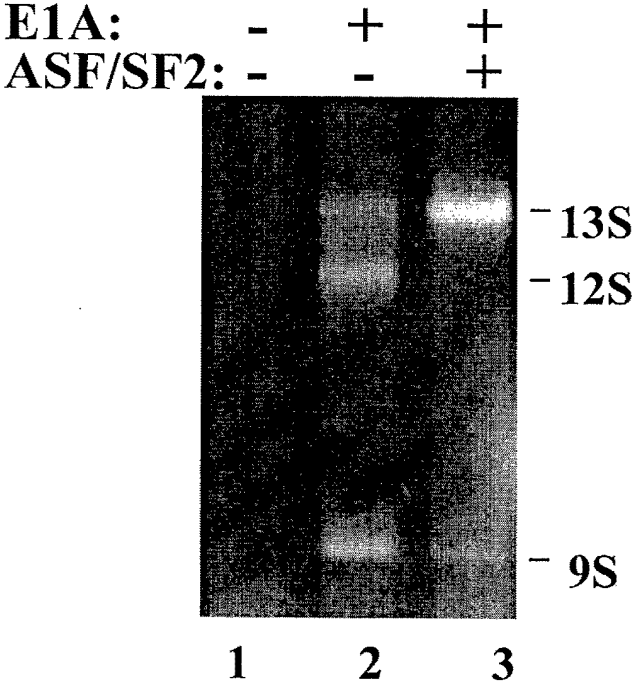
Step 5: Hybridization to Zip-code arrays

Induced Switch of Splice Sites in Transfected Cells

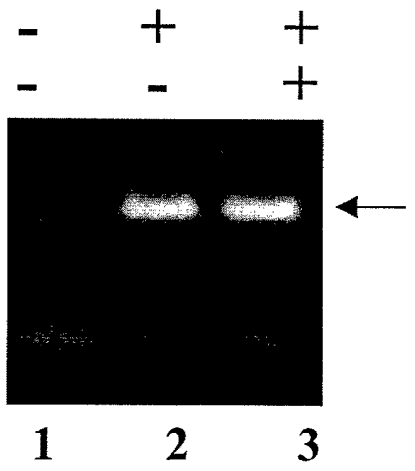
A. E1A splicing reporter



B. RT-PCR analysis

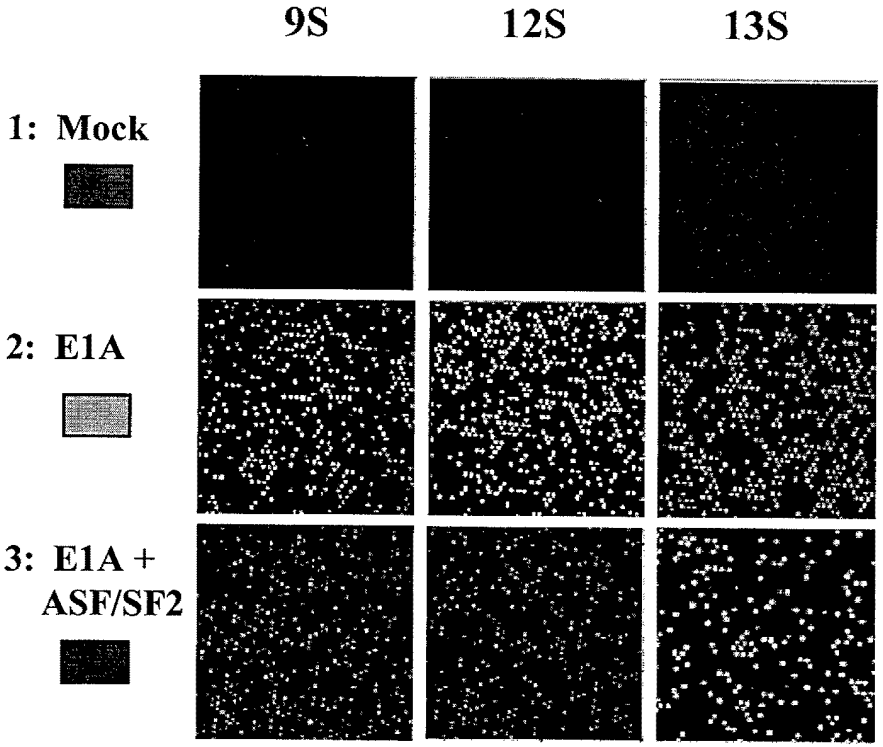


C. RNA-dep. ligation & PCR amplification



Detection of Alternative Splicing on Zip-Code Arrays and Comparison with RT-PCR

A. Array Images



B. Quantitative Comparison

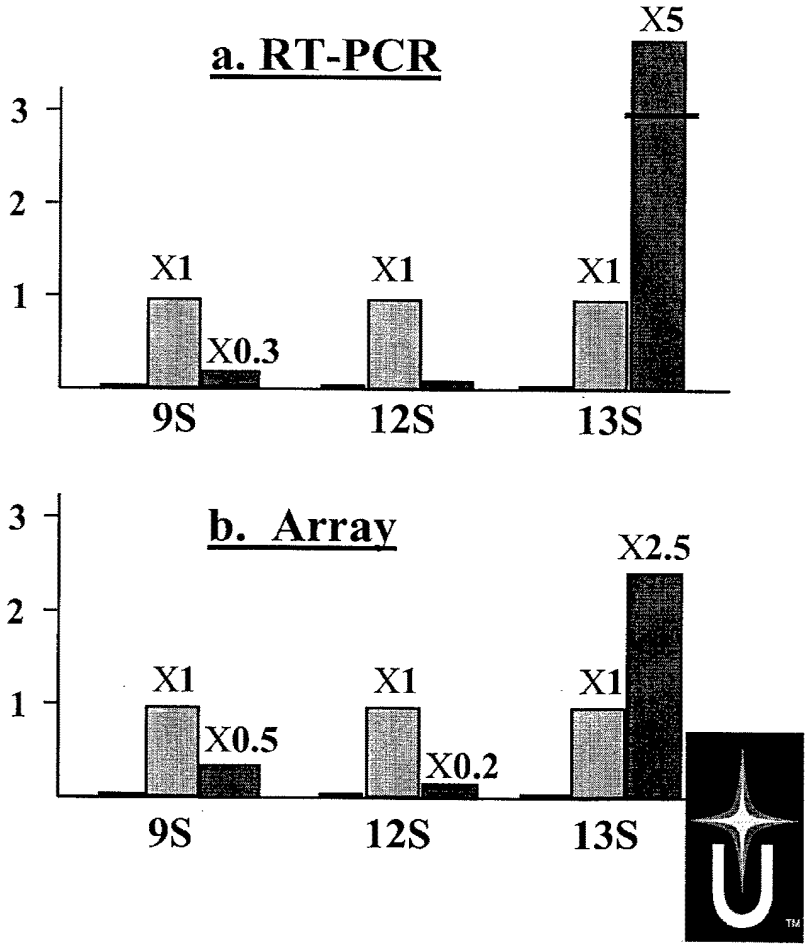


EXHIBIT 11

US007955794B2

(12) **United States Patent**
Shen et al.

(10) **Patent No.: US 7,955,794 B2**
(45) **Date of Patent: *Jun. 7, 2011**

(54) **MULTIPLEX NUCLEIC ACID REACTIONS**

(75) Inventors: **Mun-Jui Richard Shen**, Poway, CA (US); **Arnold Oliphant**, Poway, CA (US); **Scott L. Butler**, San Diego, CA (US); **John E. Stuelpnagel**, Encinitas, CA (US); **Mark S. Chee**, Del Mar, CA (US); **Kenneth M. Kuhn**, San Diego, CA (US); **Jian-Bing Fan**, San Diego, CA (US)

(73) Assignee: **Illumina, Inc.**, San Diego, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 824 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **10/177,727**

(22) Filed: **Jun. 20, 2002**

(65) **Prior Publication Data**

US 2003/0211489 A1 Nov. 13, 2003

Related U.S. Application Data

(60) Provisional application No. 60/234,143, filed on Sep. 21, 2000, provisional application No. 60/234,732, filed on Sep. 22, 2000, provisional application No. 60/297,609, filed on Jun. 11, 2001, provisional application No. 60/311,271, filed on Aug. 9, 2001, provisional application No. 60/336,958, filed on Dec. 3, 2001, provisional application No. 60/305,118, filed on Jul. 12, 2001, provisional application No. 60/341,827, filed on Dec. 17, 2001.

(51) **Int. Cl.**

C12Q 1/68	(2006.01)
C12P 19/34	(2006.01)
C12M 1/34	(2006.01)
C07H 21/02	(2006.01)
C07H 21/04	(2006.01)

(52) **U.S. Cl.** **435/6**; 435/91.2; 435/91.51; 435/287.2; 435/91.1; 536/23.1; 536/24.3; 536/24.33

(58) **Field of Classification Search** 435/6, 91.1, 435/91.2, 91.51, 183, 283.1, 287.1, 287.2; 436/94, 501; 536/231, 24.3, 24.33, 25.3

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,200,110	A *	4/1980	Peterson et al.	128/634
4,499,052	A *	2/1985	Fulwyler	422/52
4,563,419	A	1/1986	Ranki et al.	
4,582,789	A	4/1986	Sheldon	
4,682,895	A *	7/1987	Costello	356/402
4,687,732	A	8/1987	Ward et al.	
4,751,177	A	6/1988	Stabinsky	
4,785,814	A *	11/1988	Kane	128/634
4,822,746	A *	4/1989	Walt	436/528
4,824,789	A *	4/1989	Yafuso et al.	436/68
4,876,187	A	10/1989	Duck et al.	
4,882,269	A	11/1989	Schneider et al.	
4,883,750	A	11/1989	Whiteley et al.	
4,921,805	A	5/1990	Gebeyehu et al.	
4,988,617	A *	1/1991	Landegren et al.	435/6
4,999,306	A *	3/1991	Yafuso et al.	436/68
5,002,867	A *	3/1991	Macevicz	435/6
5,011,769	A *	4/1991	Duck et al.	435/6
5,028,545	A *	7/1991	Soini	436/501
5,104,791	A	4/1992	Abbott et al.	
5,105,305	A *	4/1992	Betzig et al.	359/368
5,114,864	A *	5/1992	Walt	436/528
5,132,242	A *	7/1992	Cheung	436/501
5,143,853	A *	9/1992	Walt	436/501
5,175,082	A	12/1992	Jeffreys	
5,185,243	A *	2/1993	Ullman et al.	435/6
5,194,300	A *	3/1993	Cheung	427/213.31
5,232,829	A	8/1993	Longiaru et al.	
5,244,636	A *	9/1993	Walt et al.	422/82.07
5,244,813	A *	9/1993	Walt et al.	436/172
5,250,264	A *	10/1993	Walt et al.	422/82.07
5,252,494	A *	10/1993	Walt	436/528
5,254,477	A *	10/1993	Walt	436/172
5,298,741	A *	3/1994	Walt et al.	250/227.23
5,302,509	A *	4/1994	Cheeseman	435/6
5,314,809	A	5/1994	Erlich et al.	
5,320,814	A *	6/1994	Walt et al.	422/82.07
5,357,590	A *	10/1994	Auracher	385/33
5,380,489	A *	1/1995	Sutton et al.	422/68.1
5,387,505	A	2/1995	Wu et al.	
5,399,491	A *	3/1995	Kacian et al.	435/91.21
5,403,711	A	4/1995	Walder et al.	
5,409,818	A *	4/1995	Davey et al.	435/91.21
5,427,930	A *	6/1995	Birkenmeyer et al.	435/91.52
5,435,724	A *	7/1995	Goodman et al.	433/215
5,445,934	A *	8/1995	Fodor et al.	435/6

(Continued)

FOREIGN PATENT DOCUMENTS

EP	0139489	5/1985

(Continued)

OTHER PUBLICATIONS

Abel et al., "Fiber-Optic Evanescent Wave Biosensor for the Detection of Oligonucleotides," *Anal. Chem.*, 8:2905-2912 (1996).

Abramson and Myers, "Nucleic acid amplification technologies," *Current Opinion in Biotechnology*, 4:41-47 (1993)

Anonymous, "Fluorescent Microspheres," Tech. Note #19, Bangs Laboratories, Inc. (Fishers, IN) Feb. 1997.

Anonymous, "Microsphere Selection Guide," Bangs Laboratories, Inc. (Fishers, IN) Sep. 1998.

Bangs, "Immunological Applications of Microspheres," *The Latex Course*, Bangs Laboratories, Inc.(Carmel, IN) Apr. 1996.

Barany, "Genetic disease detection and DNA amplification using cloned thermostable ligase," *Proc. Natl. Acad. Sci. USA*, 88:189-193 (1991).

(Continued)

Primary Examiner — Frank W Lu
(74) *Attorney, Agent, or Firm* — Jones Day

(57) **ABSTRACT**

The invention is directed to a variety of multiplexing methods used to amplify and/or genotype a variety of samples simultaneously.

22 Claims, 14 Drawing Sheets

U.S. PATENT DOCUMENTS

5,455,166	A		10/1995	Walker
5,474,895	A	*	12/1995	Ishii et al. 435/6
5,481,629	A	*	1/1996	Tabuchi 385/14
5,494,798	A	*	2/1996	Gerdt et al. 435/6
5,494,810	A		2/1996	Barany et al. 435/91.52
5,496,997	A		3/1996	Pope 250/227.21
5,503,980	A		4/1996	Cantor
5,512,490	A		4/1996	Walt et al. 436/171
H1531	H		5/1996	Blumentals
5,516,635	A		5/1996	Ekins et al. 435/6
5,521,065	A		5/1996	Whiteley et al.
5,541,311	A		7/1996	Dahlberg et al. 536/23.7
5,545,531	A		8/1996	Rava et al. 435/6
5,554,516	A		9/1996	Kacian et al. 435/91.21
5,565,324	A		10/1996	Still et al. 435/6
5,567,587	A	*	10/1996	Kohne 435/6
5,573,907	A		11/1996	Carrino et al. 435/6
5,573,909	A		11/1996	Singer et al. 435/6
5,575,849	A		11/1996	Honda et al. 118/44
5,578,832	A		11/1996	Trulson et al. 250/458.1
5,593,840	A		1/1997	Bhatnagar et al. 435/6
5,595,890	A		1/1997	Newton et al. 435/91.2
5,604,097	A		2/1997	Brenner et al.
5,632,957	A		5/1997	Heller et al.
5,633,972	A		5/1997	Walt et al. 385/116
5,639,603	A		6/1997	Dower et al. 435/6
5,656,241	A		8/1997	Seifert et al. 422/82.06
5,660,988	A		8/1997	Duck et al. 435/6
5,679,524	A		10/1997	Nikiforov et al. 435/6
5,690,894	A		11/1997	Pinkel et al. 422/68.1
5,705,365	A		1/1998	Ryder et al. 435/91.1
5,710,029	A		1/1998	Ryder et al. 435/911
5,733,729	A		3/1998	Lipshutz et al. 435/6
5,744,305	A		4/1998	Fodor et al. 435/6
5,763,175	A		6/1998	Brenner
5,780,231	A		7/1998	Brenner
5,792,607	A		8/1998	Backman et al. 435/6
5,795,714	A		8/1998	Cantor et al.
5,795,716	A		8/1998	Chee et al. 435/6
5,800,992	A		9/1998	Fodor et al. 435/6
5,804,376	A		9/1998	Braxton et al.
5,814,524	A		9/1998	Walt et al. 436/518
5,830,655	A		11/1998	Monforte et al.
5,830,711	A		11/1998	Barany et al. 435/91.1
5,834,758	A		11/1998	Trulson et al. 250/201.2
5,840,256	A		11/1998	Demers et al. 422/102
5,846,710	A		12/1998	Bajaj 435/6
5,846,717	A		12/1998	Brow et al. 435/6
5,846,719	A		12/1998	Brenner et al.
5,849,544	A		12/1998	Harris
5,853,989	A		12/1998	Jeffreys et al.
5,854,033	A		12/1998	Lizardi 435/91.2
5,854,684	A		12/1998	Stabile et al. 356/440
5,856,083	A		1/1999	Chelsky et al. 435/4
5,858,732	A		1/1999	Solomon et al. 435/91.2
5,863,708	A		1/1999	Zanzucchi et al. 430/320
5,866,321	A		2/1999	Matsue et al.
5,869,252	A		2/1999	Bouma et al. 435/6
5,871,928	A		2/1999	Fodor et al. 435/6
5,874,219	A		2/1999	Rava et al. 435/6
5,876,924	A		3/1999	Zhang et al. 435/5
5,888,723	A		3/1999	Sutton et al. 435/5
5,888,779	A		3/1999	Kacian et al. 435/91.2
5,888,819	A		3/1999	Goelet et al. 435/5
5,900,481	A		5/1999	Lough et al. 536/55.3
5,935,793	A		8/1999	Wong 435/6
5,942,391	A		8/1999	Zhang et al. 435/6
5,952,174	A		9/1999	Nikiforov et al. 435/6
5,974,164	A		10/1999	Chee 382/129
5,976,797	A		11/1999	Mitsuhashi et al.
5,994,066	A		11/1999	Bergeron et al. 435/6
5,998,175	A		12/1999	Akhavan-Tafti
6,001,564	A		12/1999	Bergeron et al. 435/6
6,013,440	A		1/2000	Lipshutz et al.
6,013,456	A		1/2000	Akhavan-Tafti 435/6
6,017,738	A		1/2000	Morris et al.
6,023,540	A		2/2000	Walt et al. 385/12
6,025,601	A		2/2000	Trulson et al. 250/461.2
6,027,889	A		2/2000	Barany et al. 435/6
6,040,138	A		3/2000	Lockhart et al. 435/6
6,045,996	A		4/2000	Cronin et al.
6,051,380	A		4/2000	Sosnowski et al. 435/6
6,054,564	A		4/2000	Barany et al. 536/22.1
6,060,245	A		5/2000	Sorge
6,066,454	A		5/2000	Lipshutz 435/6
6,083,763	A		7/2000	Balch 436/518
6,090,558	A		7/2000	Butler et al. 435/6
6,096,496	A		8/2000	Frankel
6,110,678	A		8/2000	Weisburg et al. 435/6
6,124,102	A		9/2000	Fodor et al. 435/7.1
6,143,495	A		11/2000	Lizardi et al. 435/6
6,172,218	B1		1/2001	Brenner 536/25.4
6,183,960	B1		2/2001	Lizardi 435/6
6,185,561	B1		2/2001	Balaban et al. 707/6
6,188,783	B1		2/2001	Balaban et al. 382/129
6,197,506	B1		3/2001	Fodor et al. 435/6
6,210,884	B1		4/2001	Lizardi 435/6
6,221,603	B1		4/2001	Mahtani
6,225,064	B1		5/2001	Uematsu et al.
6,225,625	B1		5/2001	Pirrung et al. 250/302
6,229,911	B1		5/2001	Balaban et al. 382/128
6,251,639	B1		6/2001	Kurn 435/91.2
6,261,776	B1		7/2001	Pirrung et al. 435/6
6,268,148	B1		7/2001	Barany et al. 435/6
6,274,323	B1		8/2001	Bruchez et al.
6,280,935	B1		8/2001	Macevicz 435/6
6,280,949	B1		8/2001	Lizardi
6,284,465	B1		9/2001	Wolber
6,291,166	B1		9/2001	Gerdes et al.
6,291,183	B1		9/2001	Pirrung et al. 435/6
6,306,643	B1		10/2001	Gentalen et al.
6,308,170	B1		10/2001	Balaban 707/3
6,309,822	B1		10/2001	Fodor et al. 435/6
6,316,229	B1		11/2001	Lizardi et al. 435/91.1
6,327,410	B1		12/2001	Walt et al.
6,335,165	B1		1/2002	Navot et al.
6,342,389	B1		1/2002	Cubicciotti
6,346,413	B1		2/2002	Fodor et al. 435/287.2
6,355,432	B1		3/2002	Fodor et al. 435/6
6,372,813	B1		4/2002	Johnson et al.
6,399,365	B2		6/2002	Besemer et al. 435/287.2
6,403,320	B1		6/2002	Read et al. 435/6
6,416,952	B1		7/2002	Pirrung et al. 435/6
6,440,667	B1		8/2002	Fodor et al. 435/6
6,451,536	B1		9/2002	Fodor et al. 435/6
6,491,871	B1		12/2002	Fodor et al. 422/63
6,534,293	B1		3/2003	Barany et al.
6,544,739	B1		4/2003	Fodor et al. 435/6
6,576,424	B2		6/2003	Fodor et al. 435/6
6,604,902	B2		8/2003	Norris et al. 414/331.05
6,610,482	B1		8/2003	Fodor et al. 435/6
6,618,679	B2		9/2003	Loehrlein et al. 702/20
6,627,402	B2		9/2003	Wallace 435/6
6,646,243	B2		11/2003	Pirrung et al. 250/200
6,812,005	B2		11/2004	Fan et al.
7,582,420	B2	*	9/2009	Oliphant et al. 435/6
2002/0150921	A1		10/2002	Barany et al.
2002/0168645	A1		11/2002	Taylor
2002/0177141	A1		11/2002	Chee et al. 435/6
2004/0101835	A1		5/2004	Willis et al.
2006/0275782	A1		12/2006	Gunderson et al.

FOREIGN PATENT DOCUMENTS

EP	0238332	9/1987
EP	0 269 764	6/1988
EP	0 392 546	10/1990
EP	0 478 319	4/1992
EP	0 320 308	11/1993
EP	0 336 731	5/1994
EP	0 246 864	7/1994
EP	0614987	9/1994
EP	0 357 336	10/1994
EP	0 439 182	4/1996
EP	0 723 146	7/1996
EP	0 799 897	10/1997
EP	1 121 465	9/2002
GB	2156074	10/1985

WO	WO 89/09835	10/1989
WO	WO 89/11101	11/1989
WO	WO 89/12696	12/1989
WO	WO 90/01069	2/1990
WO	WO 90/01564	2/1990
WO	WO 91/06678	5/1991
WO	WO 91/17442	11/1991
WO	WO 93/02360	2/1993
WO	WO 93/25563	12/1993
WO	WO 94/02515	2/1994
WO	WO 95/04160	2/1995
WO	WO 95/16918	6/1995
WO	WO 95/21271	8/1995
WO	WO 95/25538	9/1995
WO	WO 96/03212	2/1996
WO	WO 96/15271	5/1996
WO	WO 96/17958	6/1996
WO	WO 96/30392	10/1996
WO	WO 96/31622	10/1996
WO	WO 97/14028	4/1997
WO	WO 97/14928	4/1997
WO	WO 97/31256	8/1997
WO	WO 97/40385	10/1997
WO	WO 97/45559	12/1997
WO	WO 97/46704	12/1997
WO	WO 98/04746	2/1998
WO	WO 98/13523	4/1998
WO	WO 98/31836	7/1998
WO	WO 98/37230	8/1998
WO	WO 98/40726	9/1998
WO	WO 98/50782	11/1998
WO	WO 98/53093	11/1998
WO	WO 98/53300	11/1998
WO	WO 98/59243	12/1998
WO	WO 99/18434	4/1999
WO	WO 99/39001	8/1999
WO	WO 99/53102	10/1999
WO	WO 99/60170	11/1999
WO	WO 99/64867	12/1999
WO	WO 99/67414	12/1999
WO	WO 99/67641	12/1999
WO	WO 00/04372	1/2000
WO	WO 00/13004	3/2000
WO	WO 00/14193	3/2000
WO	WO 00/16101	3/2000
WO	WO 00/39587	7/2000
WO	WO 00/47996	8/2000
WO	WO 00/48000	8/2000
WO	WO 00/58516	10/2000
WO	WO 00/63437	10/2000
WO	WO 00/71243	11/2000
WO	WO 00/71992	11/2000
WO	WO 00/71995	11/2000
WO	WO 00/75373	12/2000
WO	WO 01/06012	1/2001
WO	WO 01/62961	8/2001
WO	WO 02/057491	7/2002
WO	WO 02/061143	8/2002

OTHER PUBLICATIONS

Barnard and Walt, "A Fibre-Optic Chemical Sensor with Discrete Sensing Sites," *Nature*, 53:338-340 (1991).

Broude et al, "Enhanced DNA sequencing by hybridization," *Proc. Natl. Acad. Sci. USA*, 91:3072-3076, (1994).

Chen et al., "A Microsphere-Based Assay for Multiplexed Single Nucleotide Polymorphism Analysis Using Single Base Chain Extension," *Genome Res*., 10:549-557 (2000).

Czarnik, "Illuminating the SNP Genomic Code," *Modern Drug Discovery*, 1:49-55 (1998).

Drmanac and Crkvenjakov, "Sequencing by Hybridization (SBH) with Oligonucleotide Probes as an Integral Approach for the Analysis of Complex Genomes," *International Journal of Genome Research*, I:59-79 (1992).

Drmanac et al., "Prospects for a Miniaturized, Simplified and Frugal Human Genome Project," *Scientia Yugoslavica*, 16:97-107 (1990).

Drmanac et al., "Sequencing by Hybridization," Automated DNA Sequencing and Analysis, ed. M. Adams, C. Fields and J. Venter (1994).

Drmanac et al., "Sequencing by Oligonucleotide Hybridization: A Promising Framework in Decoding of the Genome Program," The First International Conference on Electrophoresis, Supercomputing and the Human Genome, Proceeding on the Apr. 10-13, 1990 Conference at Florida State University Ed. C. Cantor and H. Lim.

Ferguson et al., "A Fiber-Optic DNA Biosensor Microarray for the Analysis of Gene Expression," *Nat. Biotechnol*. 14:1681-1684 (1996).

Fuh et al., "Single Fibre Optic Fluorescence pH Probe," *Analyst*, 112:1159-1163 (1987).

Hatch et al., "Rolling circle amplification of DNA immobilized on solid surfaces and its application to multiplex mutation detection," *Genetic Analysis: Biomolecular Engineering*, 15:35-40 (1999).

Healey and Walt, "Development of a Penicillin Biosensor Using a Single Optical Imaging Fiber," *SPIE*, 2388:568-573 (1995).

Healey and Walt, "Improved Fiber-Optic Chemical Sensor for Penicillin," *Anal. Chem*., 67:4471-4476, (1995).

Healey et al, "Fiberoptic DNA Sensor Array Capable of Detecting Point Mutations," *Anal. Biochem*., 251:270-279 (1997).

Hirschfeld et al., "Laser-Fiber-Optic "Optrode" for Real Time In Vivo Blood Carbon Dioxide Level Monitoring," *Journal of Lightwave Technology*, LT-5(7):1027-1033 (1987).

Iannone et al., "Multiplexed Single Nucleotide Polymorphism Genotyping by Oligonucleotide Ligation and Flow Cytometry," *Cytometry*, 39:131-140 (2000).

Lizardi et al., "Mutation detection and single-molecule counting using isothermal rolling-circle amplification," *Nature Genetics*, 19:225-232 (1998).

Lyamichev et al., "Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes," *Nat. Biotechnol*., 17:292-296 (1999).

Metzker et al., "Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'- triphosphates," *Nucleic Acids Res*., 22:4259-4267 (1994).

Michael et al., "Fabrication of Micro- and Nanostructures Using Optical Imaging Fibers and Their Use as Chemical Sensors," Proceedings of the Third National Symposium on Microstructures and Microfabricated Systems, eds. P.J. Hesketh et al., v. 97-5, Electrochemical Society Proceedings, 152-157, (Aug. 1997).

Michael et al., "Making Sensors out of Disarray: Optical Sensor Microarrays," *SPIE*, 3270:34-41 (1998).

Michael et al., "Randomly Ordered Addressable High-Density Optical Sensor Arrays," *Anal. Chem*., 70: 1242-1248 (1998).

Mignani et al., "In-Vivo Biomedical Monitoring by Fiber-Optic Systems," *Journal of Lightwave Technology*, 13: 1396-1406 (1995).

Myer and Day, "Synthesis and application of circularizable ligation probes," *BioTechniques*, 30:584-593, (2001).

Nickerson, "Gene probe assays and their detection," *Current Opinion in Biotechnology*, 4:48-51 (1993).

Nilsson et al., "Padlock probes: circularizing oligonucleotides for localized DNA detection," *Science*, 265:2085-2088 (1994).

Pantano and Walt, "Ordered Nanowell Arrays," *Chem. Mater*., 8:2832-2835 (1996).

Peterson and Vurek, "Fiber-Optic Sensors for Biomedical Applications," *Science*, 13:123-127(1984).

Peterson et al., "Fiber Optic pH Probe for Physiological Use," *Anal. Chem*., 52:864-869 (1980).

Piunno et al., "Fiber-Optic DNA Sensor for Fluorometric Nucleic Acid Determination," Anal. Chem., 67:2635-2643 (1995).

Pope, "Fiber Optic Chemical Microsensors Employing Optically Active Silica Microspheres," *SPIE*, 2388:245-256 (1995).

Schafer and Hawkins, "DNA variation and the future of human genetics," *Nature Biotechnology*, 16:33-39 (1998).

Shoemaker et al., "Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy," *Nature Genetics*, 14:450-456 (1996).

Thomas et al., "Amplification of padlock probes for DNA diagnostics by cascade rolling circle amplification or the polymerase chain reaction," *Arch. Pathol. Lab. Med*., 123:1170-1176 (1999).

Walt, "Fiber Optic Imaging Sensors," *Accounts of Chemical Research*, 31:267-278 (1998).

Walt, "Fiber-Optic Sensors for Continuous Clinical Monitoring," *Proceedings of the IEEE*, 80:903-911.

Xu and Kool, "High sequence fidelity in a non-enzymatic DNA autoligation reaction," *Nucleic Acids Research*, 27:875-881 (1999).
Berg, et al., "Hybrid PCR sequencing: sequencing of PCR products using a universal primer," *BioTechniques*. 17(5):896-901 (Nov. 1994).
Boguszewski, et al., "Cloning of Two Novel Growth Hormone Transcripts Expressed in Human Placenta" *J Clin Endocrinol Metab*. 83(8): 2878-2885 (1998).
Fan, et al., "Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays" *Genome Res*. 10(6):853-860 (Jun. 2000).
Fodor, et al., "Light-directed, spatially addressable parallel chemical synthesis" *Science* 51(4995):767-73 (Feb. 15, 1991).
Hendrickson et al., "High sensitivity multianalyte immunoassay using covalent DNA-labeled antibodies and polymerase chain reaction," *Nucleic Acids Res*. 23(3):522-529 (1995).
Hirschhorn, et al., "SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping" *Proc Natl Acad Sci U S A*. 97 (22): 12164-12169 (2000).
Hsuih, et al., "Novel, ligation-dependent PCR assay for detection of hepatitis C in serum" *J Clin Microbiol*. 34(3):501-507 (Mar. 1996).
Khanna, et al., "Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors" *Oncogene* 18(1):27-38 (Jan. 7, 1999).
Kozal, et al., "Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays" *Nat. Med*. 2(7):753-759 (Jul. 1996).
Ladner, et al., "Multiplex detection of hotspot mutations by rolling circle-enabled universal microarrays" *Lab. Invest*. 81(8):1079-1086 (2001).
Pease, et al., "Light-generated oligonucleotide arrays for rapid DNA sequence analysis" *Proc. Natl. Acad. Sci*. 91(11):5022-5026 (1994).
Seradyn, "Sera-Mag Streptavadin Magnetic Microparticles" *Particle Technology* 1-7 (Nov. 1996).
Smith, et al., "Fluorescence detection in automated DNA sequence analysis" Nature 321 (6071):674-679 (1986).
Strachan, "A rapid general method for the identification of PCR products using a fibre-optic biosensor and its application to the detection of *Listeria" Lett Appl Microbiol*. 21(1):5-9 (1995).
Syvänen, AC., "From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms" *Hum. Mut*. 13(1):1-10 (1999).
Walt, "Molecular Biology: Bead-based fiber-optic arrays" *Science* 287: 451-452 (1999).
Nikiforov et al., "Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms," *Nuc Acids Res*. 22(20):4167-4175 (1994).
Shuber et al., "High throughput parallel analysis of hundreds of patient samples for more than 100 mutations in multiple disease genes," *Hum Mol Genet*. 6(3):337-347 (1997).
Office Action dated Mar. 17, 2008, in U.S. Appl. No. 11/238,826.

* cited by examiner

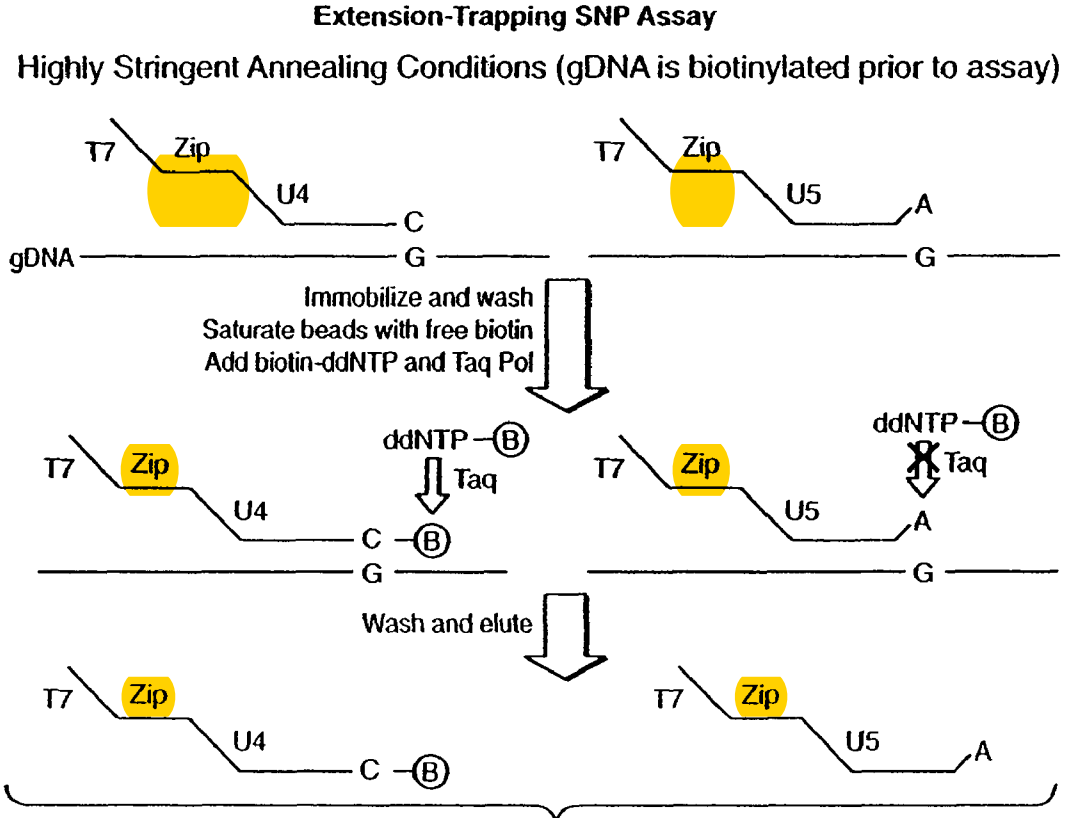

Extension-Trapping SNP Assay
Highly Stringent Annealing Conditions (gDNA is biotinylated prior to assay)
T7
Zip
U4
C
gDNA
G
T7
Zip
U5
A
G
Immobilize and wash
Saturate beads with free biotin
Add biotin-ddNTP and Taq Pol
ddNTP—B
Taq
T7
Zip
U4
C—B
G
T7
Zip
U5
A
ddNTP—B
Taq
G
Wash and elute
T7
Zip
U4
C—B
T7
Zip
U5
A

FIG._1A

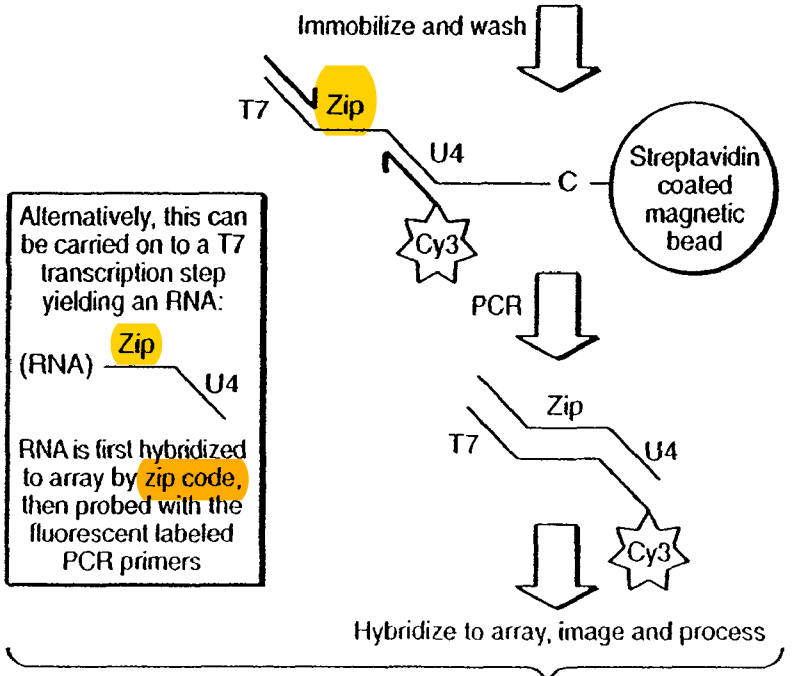

Immobilize and wash
T7
Zip
U4
C
Streptavidin coated magnetic bead
Cy3
Alternatively, this can be carried on to a T7 transcription step yielding an RNA:
(RNA)
Zip
U4
RNA is first hybridized to array by zip code, then probed with the fluorescent labeled PCR primers
PCR
Zip
T7
U4
Cy3
Hybridize to array, image and process

FIG._1B

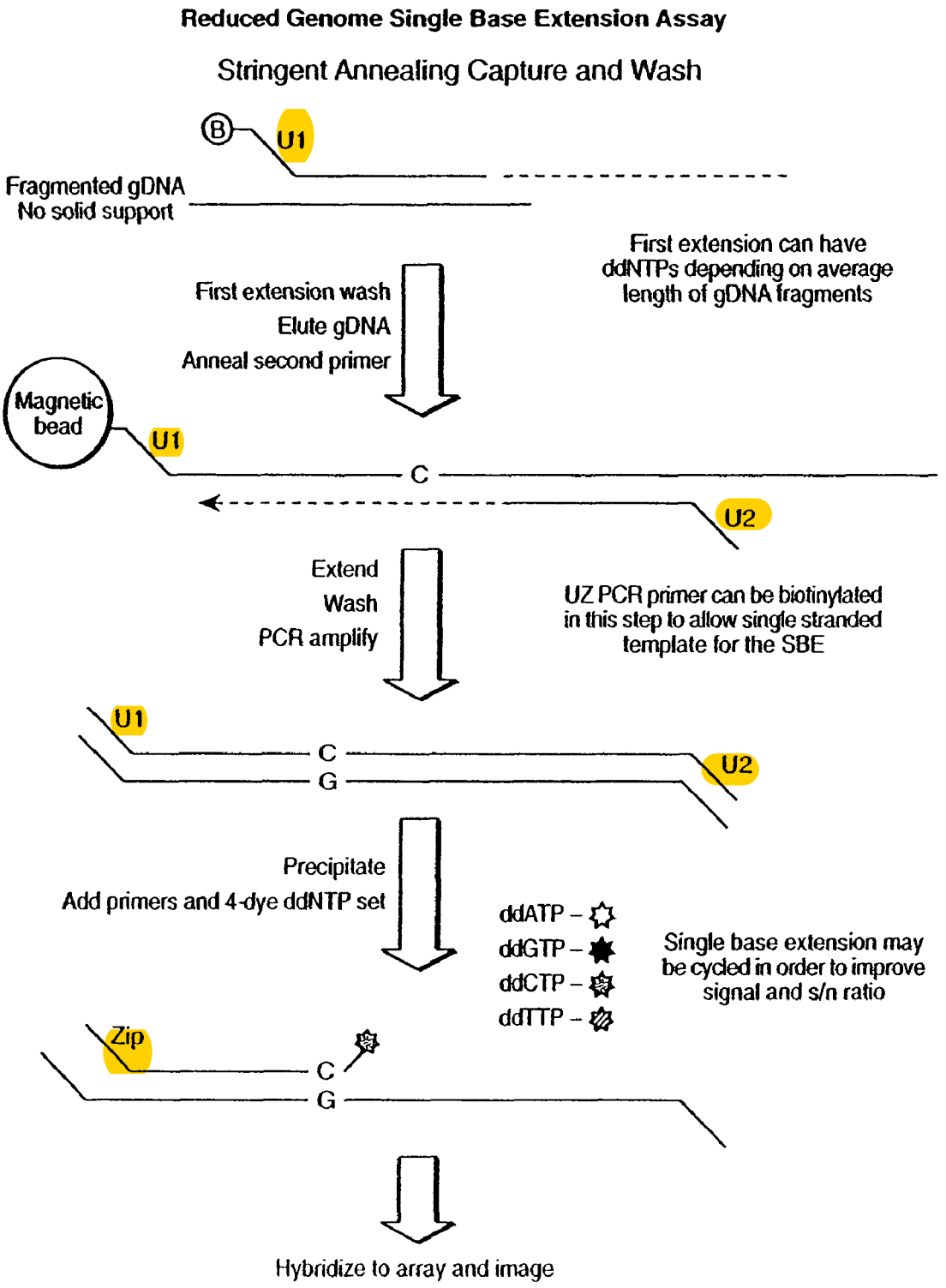

Reduced Genome Single Base Extension Assay
Stringent Annealing Capture and Wash
B
U1
Fragmented gDNA
No solid support
First extension can have ddNTPs depending on average length of gDNA fragments
First extension wash
Elute gDNA
Anneal second primer
Magnetic bead
U1
C
U2
Extend
Wash
PCR amplify
UZ PCR primer can be biotinylated in this step to allow single stranded template for the SBE
U1
C
G
U2
Precipitate
Add primers and 4-dye ddNTP set
ddATP –
ddGTP –
ddCTP –
ddTTP –
Single base extension may be cycled in order to improve signal and s/n ratio
Zip
C
G
Hybridize to array and image

FIG._2

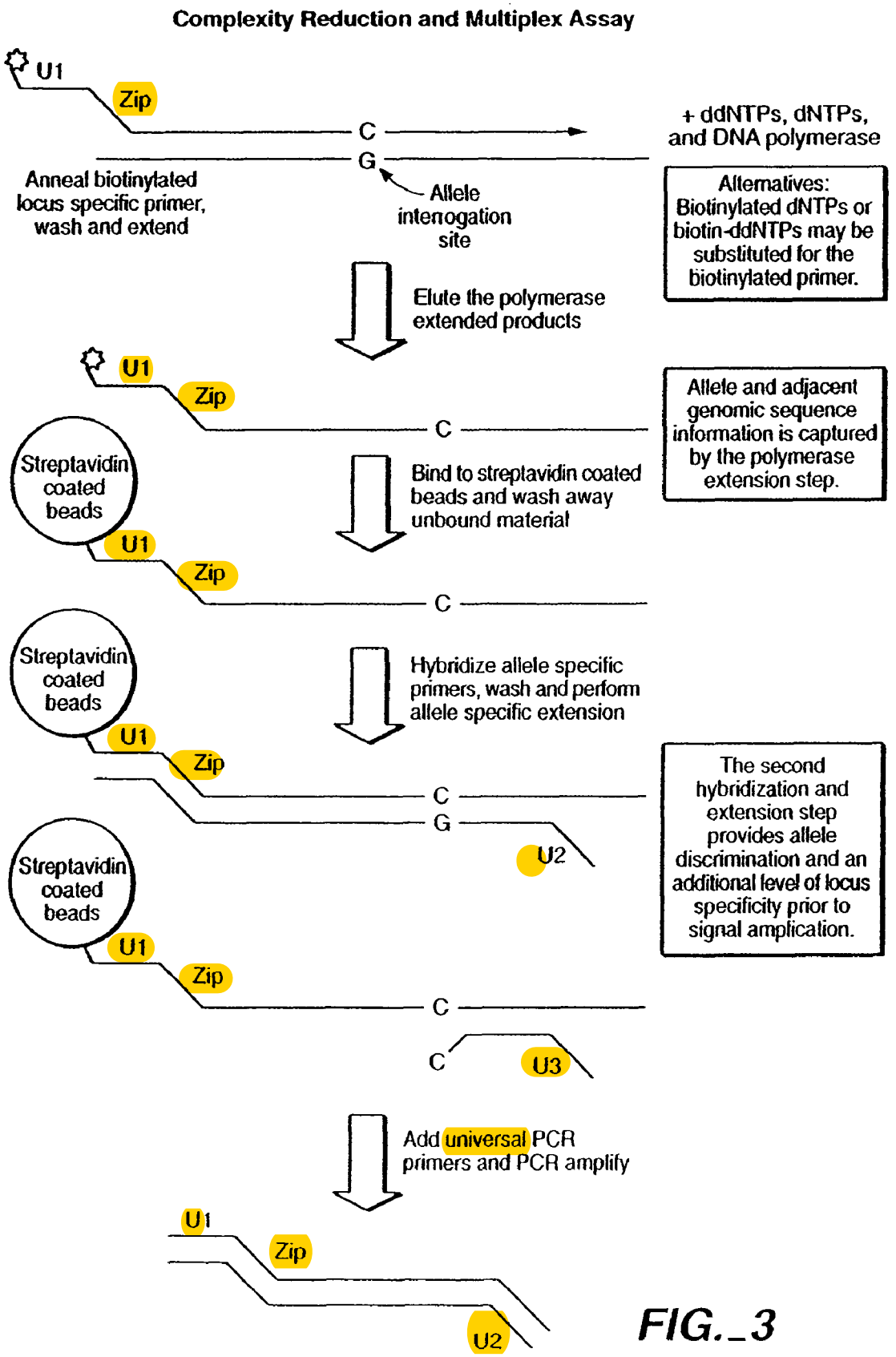

Complexity Reduction and Multiplex Assay
U1
Zip
C
+ ddNTPs, dNTPs, and DNA polymerase
G
Anneal biotinylated locus specific primer, wash and extend
Allele interrogation site
Alternatives: Biotinylated dNTPs or biotin-ddNTPs may be substituted for the biotinylated primer.
Elute the polymerase extended products
U1
Zip
C
Allele and adjacent genomic sequence information is captured by the polymerase extension step.
Streptavidin coated beads
Bind to streptavidin coated beads and wash away unbound material
U1
Zip
C
Streptavidin coated beads
Hybridize allele specific primers, wash and perform allele specific extension
U1
Zip
C
G
U2
The second hybridization and extension step provides allele discrimination and an additional level of locus specificity prior to signal amplication.
Streptavidin coated beads
U1
Zip
C
C
U3
Add universal PCR primers and PCR amplify
U1
Zip
U2

FIG._3

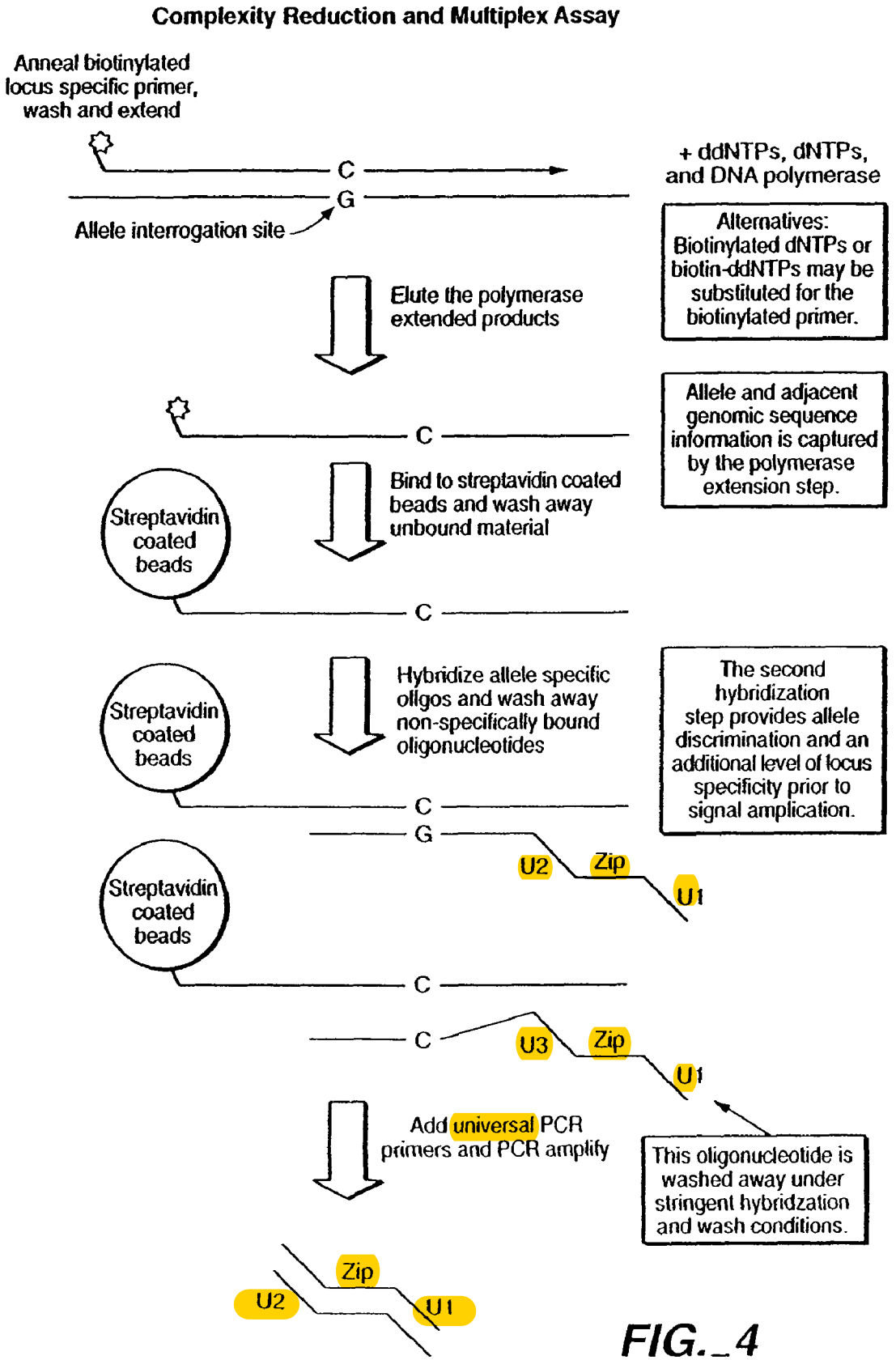

Complexity Reduction and Multiplex Assay
Anneal biotinylated locus specific primer, wash and extend
C
G
Allele interrogation site
+ ddNTPs, dNTPs, and DNA polymerase
Alternatives: Biotinylated dNTPs or biotin-ddNTPs may be substituted for the biotinylated primer.
Elute the polymerase extended products
C
Allele and adjacent genomic sequence information is captured by the polymerase extension step.
Streptavidin coated beads
Bind to streptavidin coated beads and wash away unbound material
C
Streptavidin coated beads
Hybridize allele specific oligos and wash away non-specifically bound oligonucleotides
The second hybridization step provides allele discrimination and an additional level of locus specificity prior to signal amplication.
C
G
U2
Zip
U1
Streptavidin coated beads
C
C
U3
Zip
U1
Add universal PCR primers and PCR amplify
This oligonucleotide is washed away under stringent hybridzation and wash conditions.
Zip
U2
U1

FIG._4

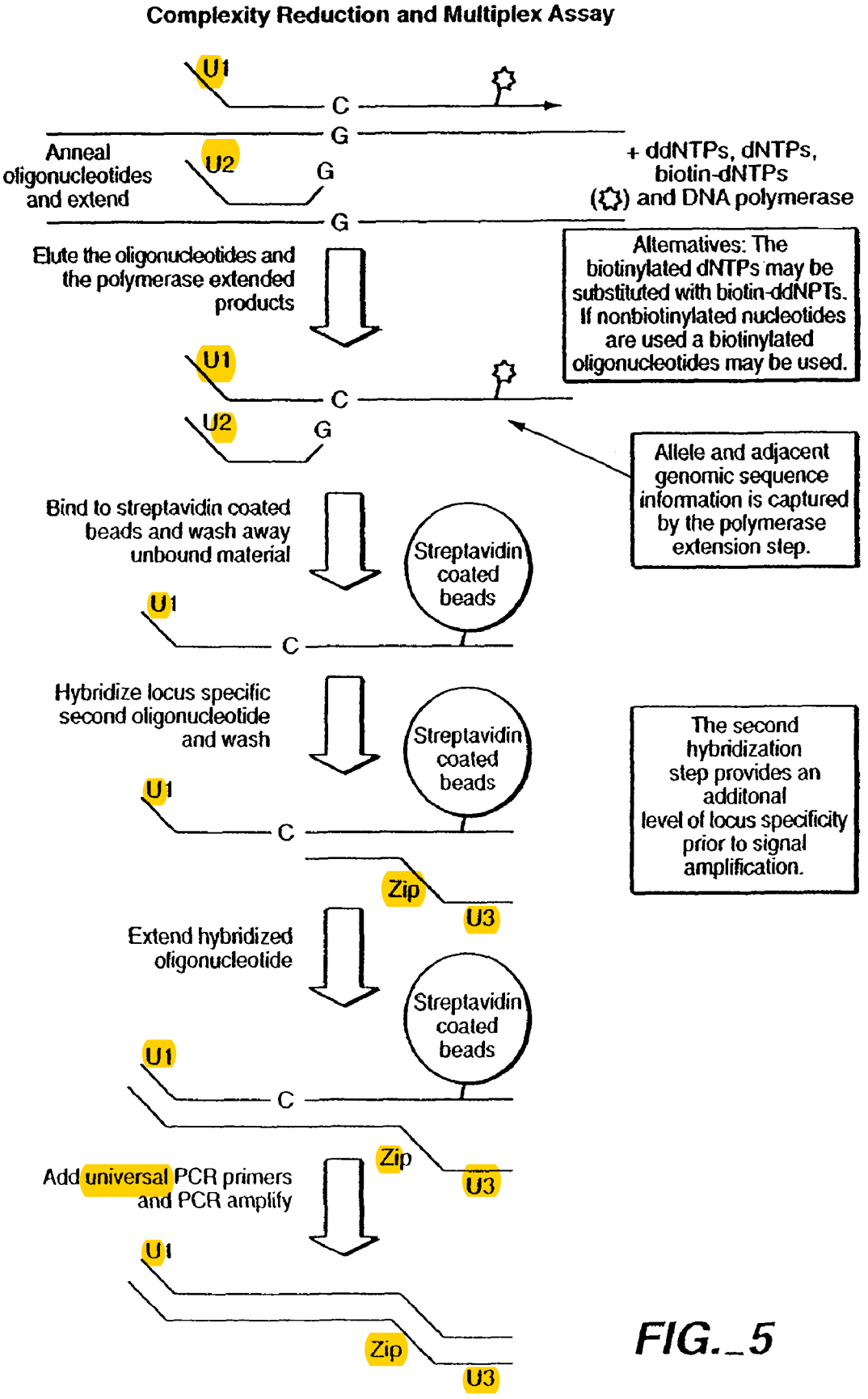

Complexity Reduction and Multiplex Assay
U1
C
G
Anneal oligonucleotides and extend
U2
G
G
+ ddNTPs, dNTPs, biotin-dNTPs (✡) and DNA polymerase
Elute the oligonucleotides and the polymerase extended products
Alternatives: The biotinylated dNTPs may be substituted with biotin-ddNPTs. If nonbiotinylated nucleotides are used a biotinylated oligonucleotides may be used.
U1
C
U2
G
Allele and adjacent genomic sequence information is captured by the polymerase extension step.
Bind to streptavidin coated beads and wash away unbound material
Streptavidin coated beads
U1
C
Hybridize locus specific second oligonucleotide and wash
Streptavidin coated beads
U1
C
Zip
U3
The second hybridization step provides an additonal level of locus specificity prior to signal amplification.
Extend hybridized oligonucleotide
Streptavidin coated beads
U1
C
Zip
U3
Add universal PCR primers and PCR amplify
U1
Zip
U3

FIG._5

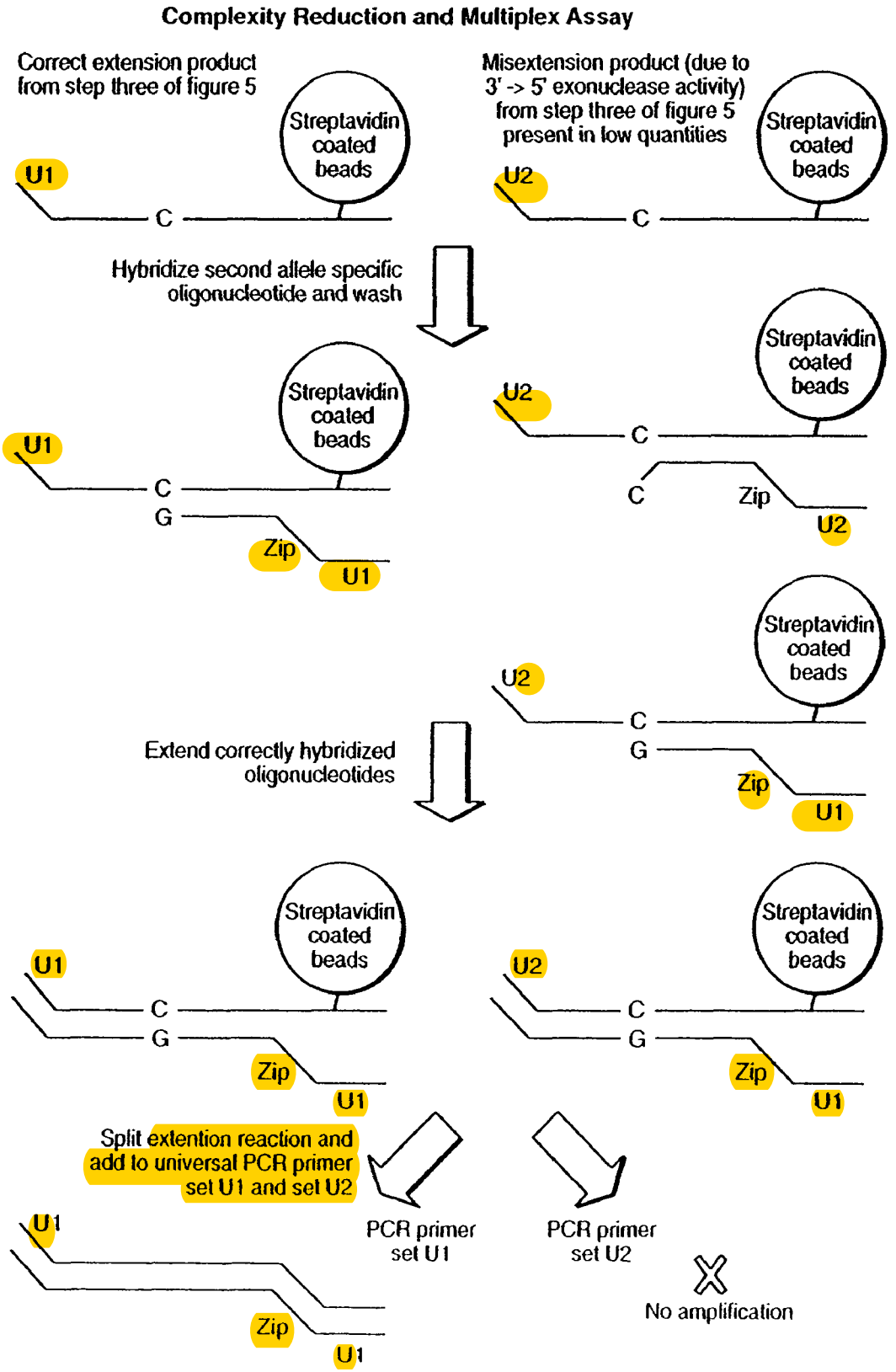

Complexity Reduction and Multiplex Assay
Correct extension product from step three of figure 5
Streptavidin coated beads
U1
C
Misextension product (due to 3' -> 5' exonuclease activity) from step three of figure 5 present in low quantities
Streptavidin coated beads
U2
C
Hybridize second allele specific oligonucleotide and wash
Streptavidin coated beads
U1
C
G
Zip
U1
Streptavidin coated beads
U2
C
C
Zip
U2
Extend correctly hybridized oligonucleotides
Streptavidin coated beads
U2
C
G
Zip
U1
Streptavidin coated beads
U1
C
G
Zip
U1
Streptavidin coated beads
U2
C
G
Zip
U1
Split extention reaction and add to universal PCR primer set U1 and set U2
PCR primer set U1
PCR primer set U2
U1
Zip
U1
No amplification

FIG._6

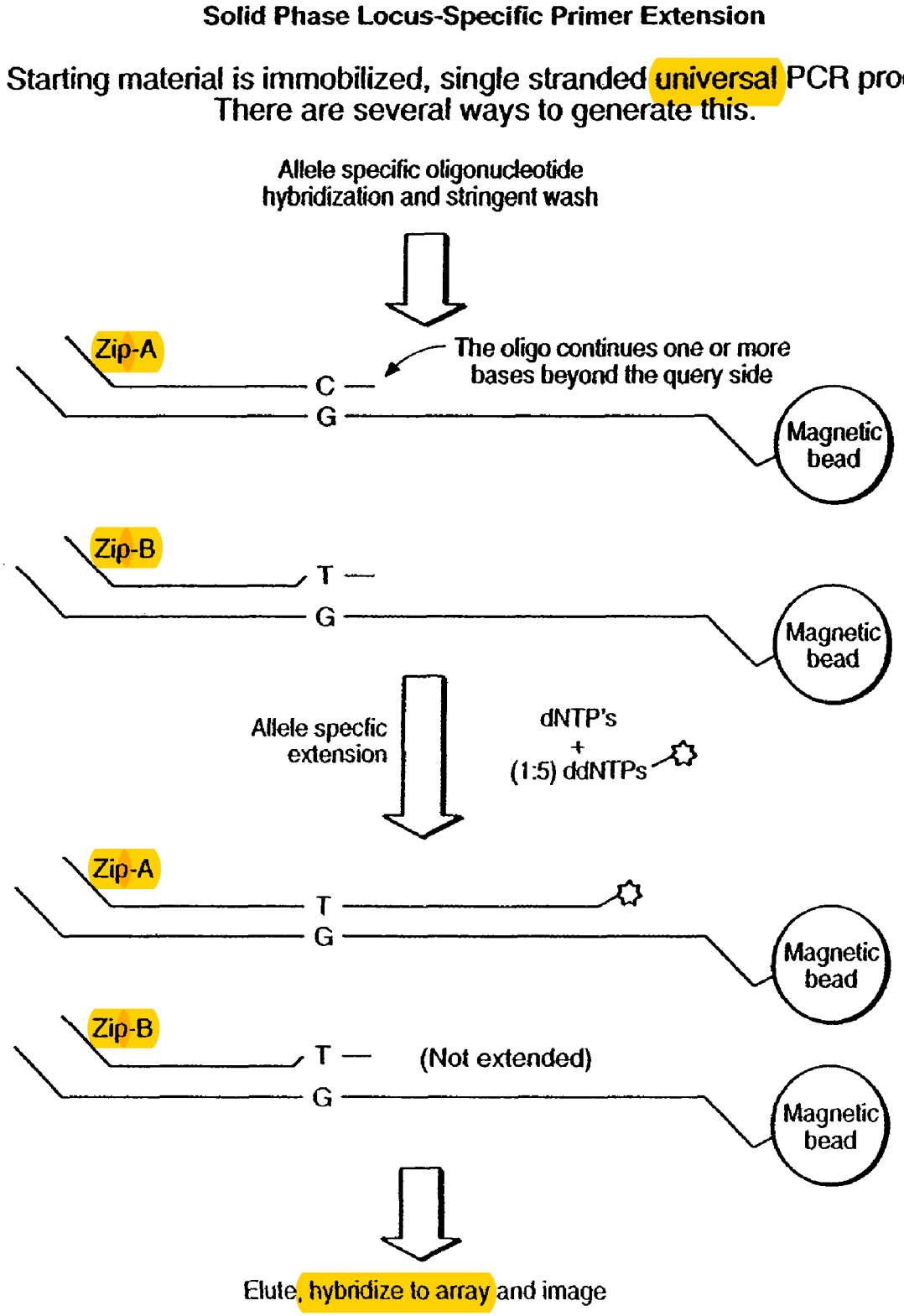

Solid Phase Locus-Specific Primer Extension
Starting material is immobilized, single stranded universal PCR product.
There are several ways to generate this.
Allele specific oligonucleotide
hybridization and stringent wash
Zip-A
C
G
The oligo continues one or more
bases beyond the query side
Magnetic
bead
Zip-B
T
G
Magnetic
bead
Allele specfic
extension
dNTP's
+
(1:5) ddNTPs
Zip-A
T
G
Magnetic
bead
Zip-B
T
(Not extended)
G
Magnetic
bead
Elute, hybridize to array and image

FIG._7

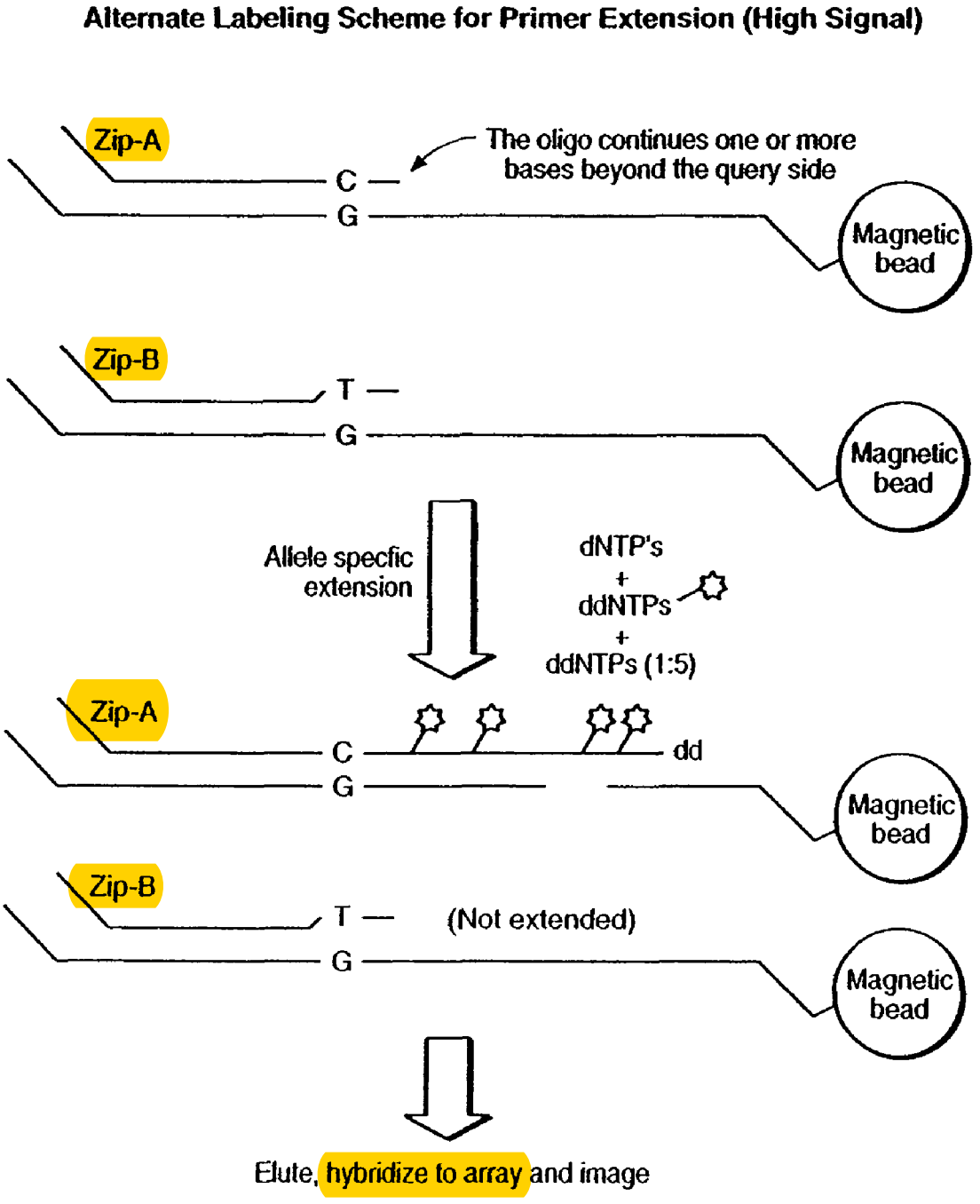

Alternate Labeling Scheme for Primer Extension (High Signal)
Zip-A
C —
G
The oligo continues one or more bases beyond the query side
Magnetic bead
Zip-B
T —
G
Magnetic bead
Allele specfic extension
dNTP's
+
ddNTPs
+
ddNTPs (1:5)
Zip-A
C
G
dd
Magnetic bead
Zip-B
T —
(Not extended)
G
Magnetic bead
Elute, hybridize to array and image

FIG._8

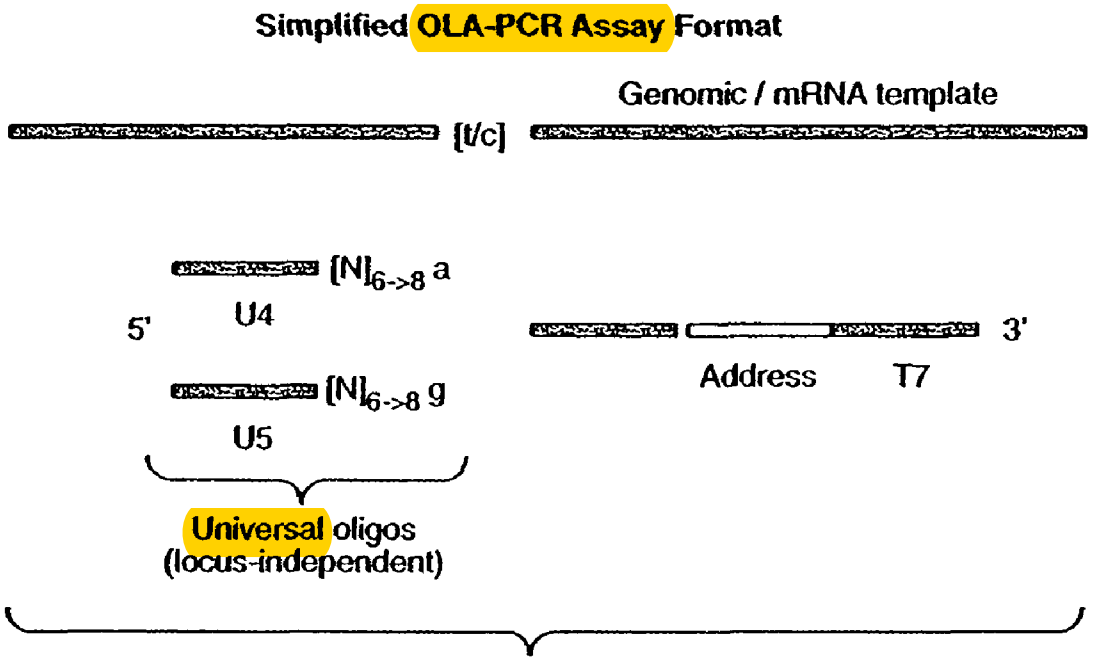

Simplified OLA-PCR Assay Format
Genomic / mRNA template
[t/c]
[N]6->8 a
U4
5'
[N]6->8 g
U5
Universal oligos
(locus-independent)
3'
Address
T7

FIG._9

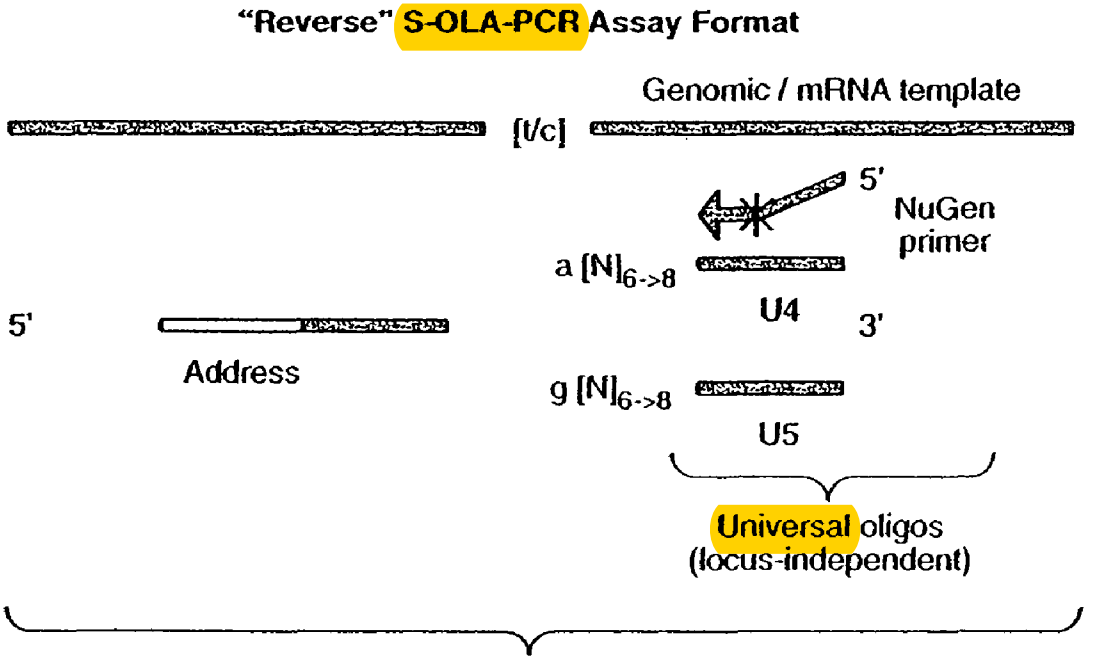

"Reverse" S-OLA-PCR Assay Format
Genomic / mRNA template
[t/c]
5'
NuGen
primer
a [N]6->8
U4
3'
5'
Address
g [N]6->8
U5
Universal oligos
(locus-independent)

FIG._10

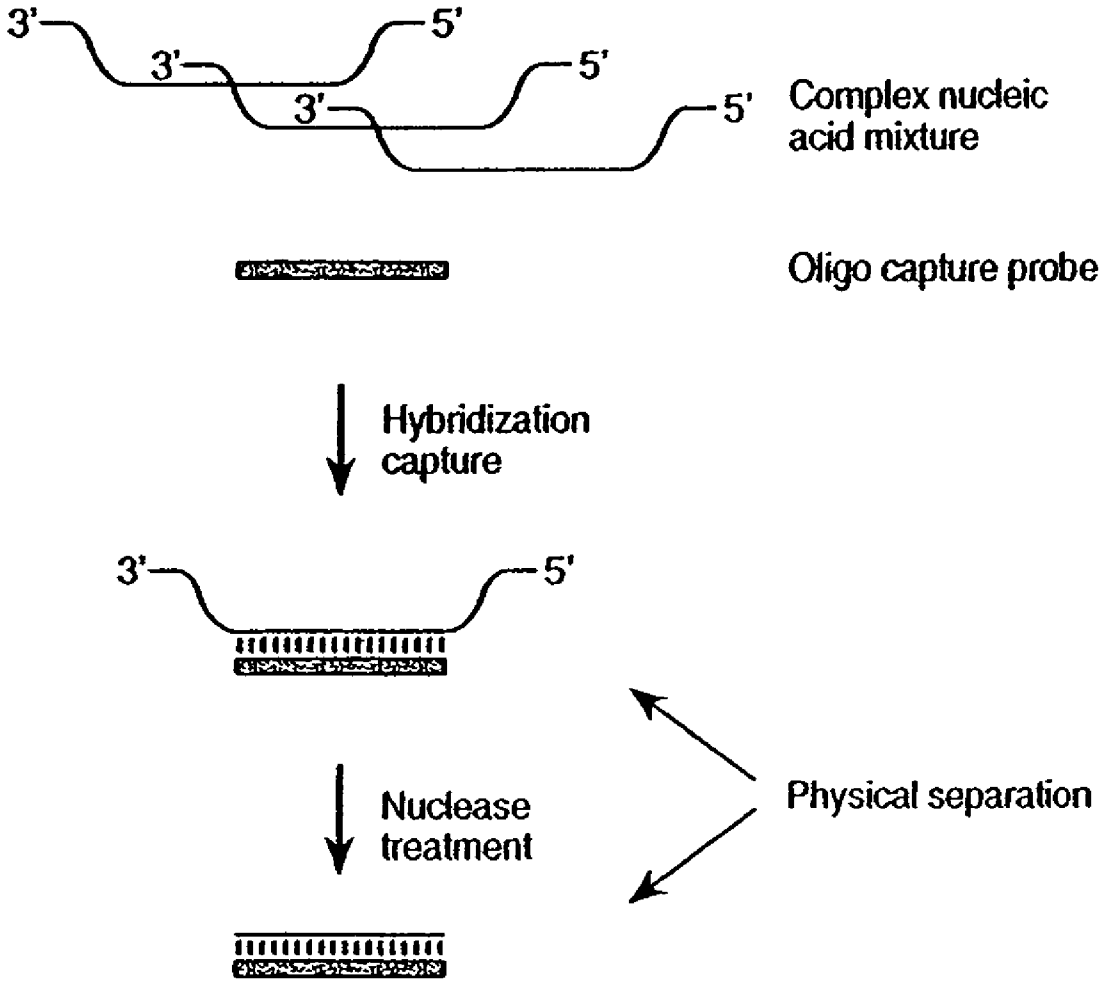

3'
5'
3'
5'
3'
5'
Complex nucleic acid mixture
Oligo capture probe
Hybridization capture
3'
5'
Nuclease treatment
Physical separation

FIG._11

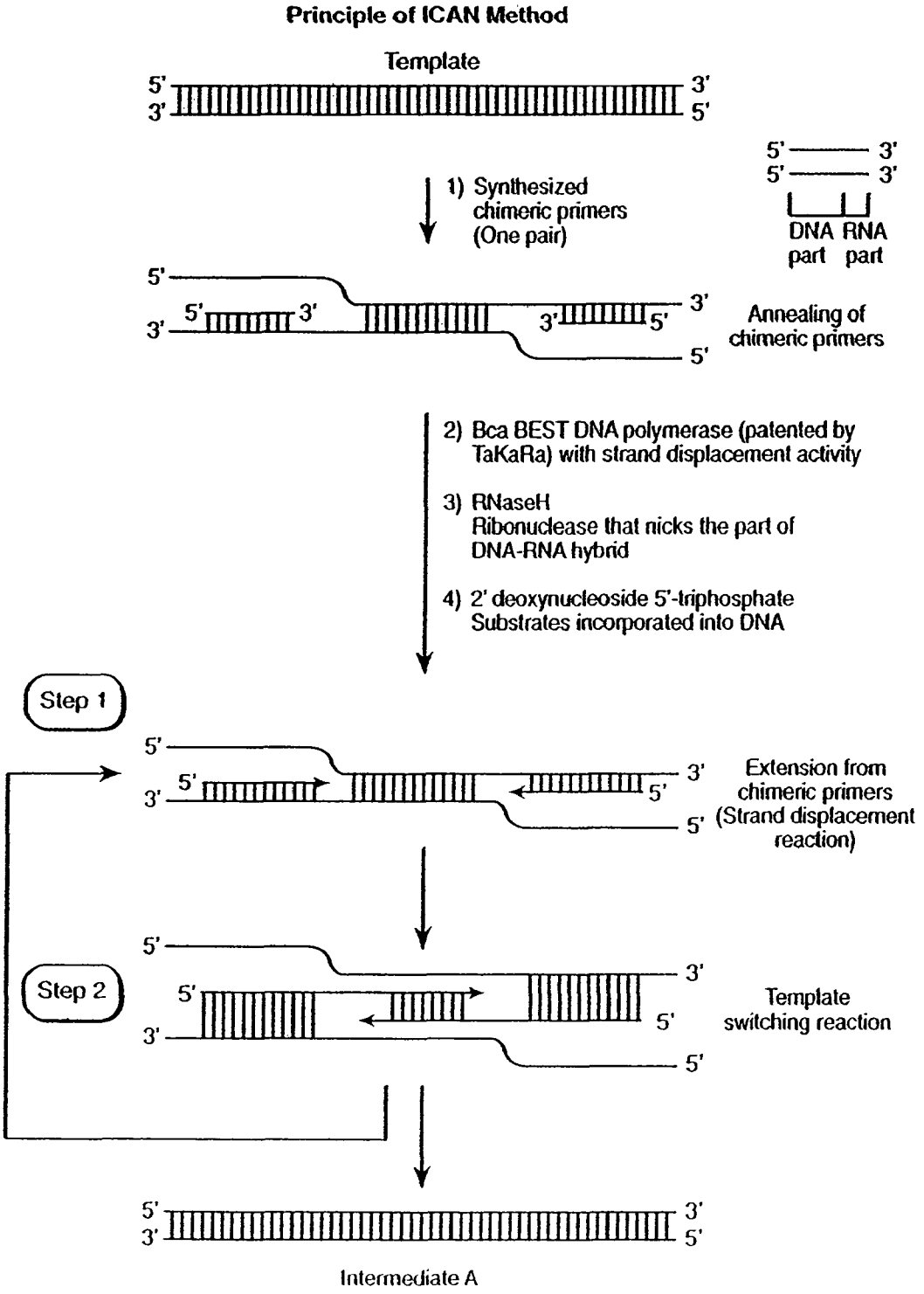

Principle of ICAN Method
Template
5'
3'
3'
5'
1) Synthesized chimeric primers (One pair)
5'
3'
DNA part
RNA part
Annealing of chimeric primers
2) Bca BEST DNA polymerase (patented by TaKaRa) with strand displacement activity
3) RNaseH
Ribonuclease that nicks the part of DNA-RNA hybrid
4) 2' deoxynucleoside 5'-triphosphate
Substrates incorporated into DNA
Step 1
Extension from chimeric primers (Strand displacement reaction)
Step 2
Template switching reaction
Intermediate A

FIG._12A

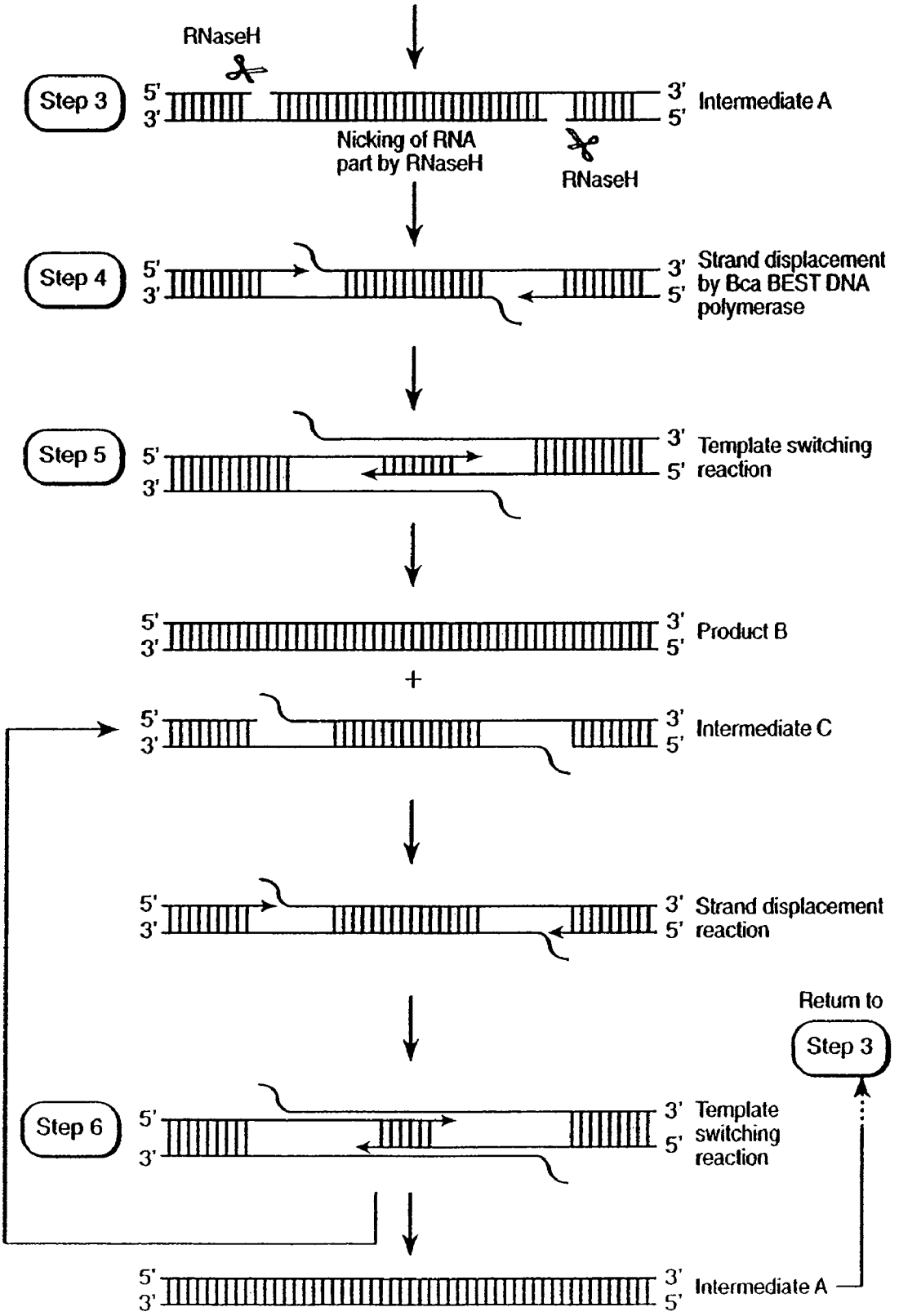

RNaseH
Step 3
5'
3'
3'
5'
Intermediate A
Nicking of RNA part by RNaseH
RNaseH
Step 4
Strand displacement by Bca BEST DNA polymerase
Step 5
Template switching reaction
Product B
+
Intermediate C
Strand displacement reaction
Return to Step 3
Step 6
Template switching reaction
Intermediate A

FIG._12B

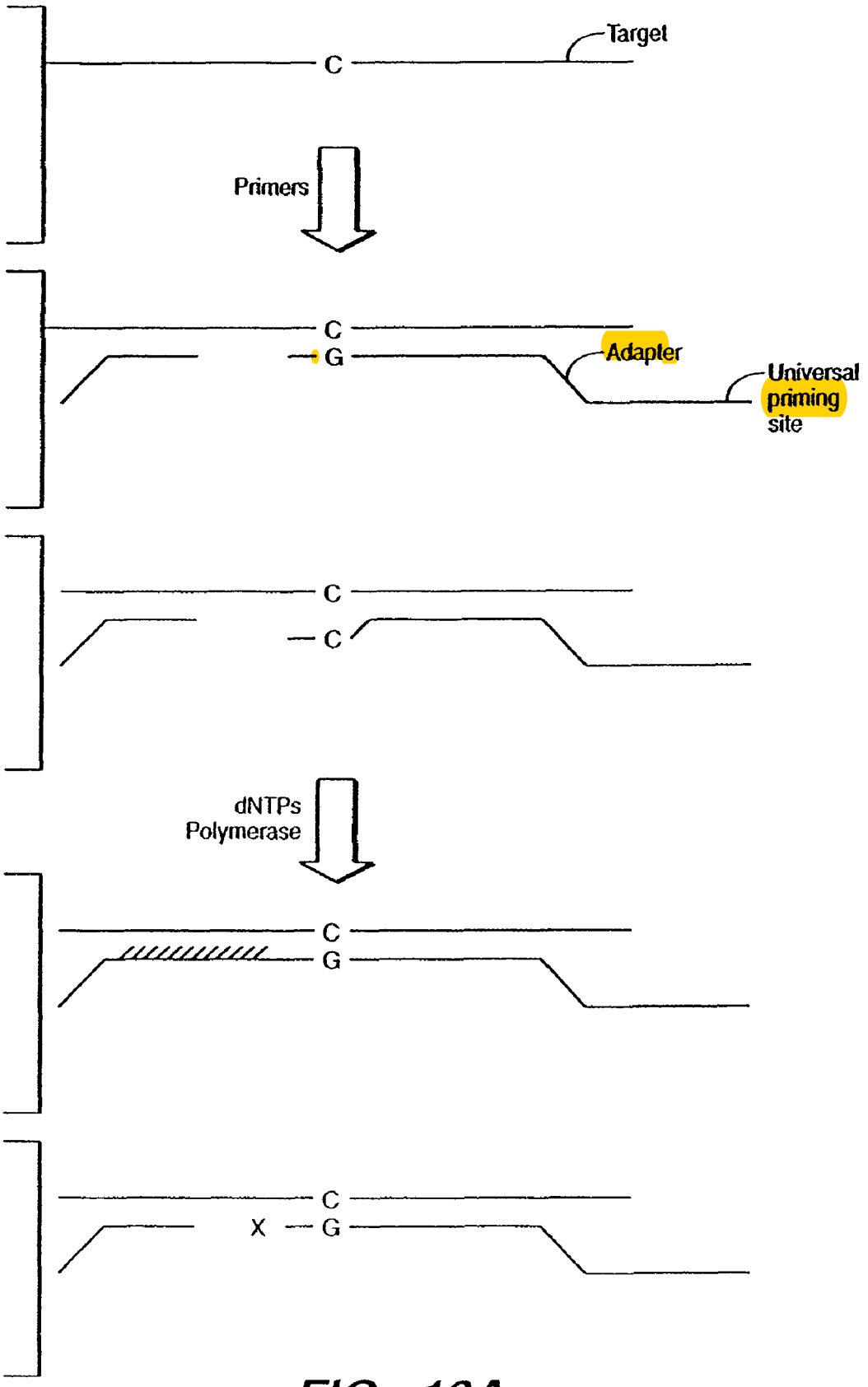

Target
C
Primers
C
G
Adapter
Universal priming site
C
C
dNTPs
Polymerase
C
G
C
X
G

FIG._13A

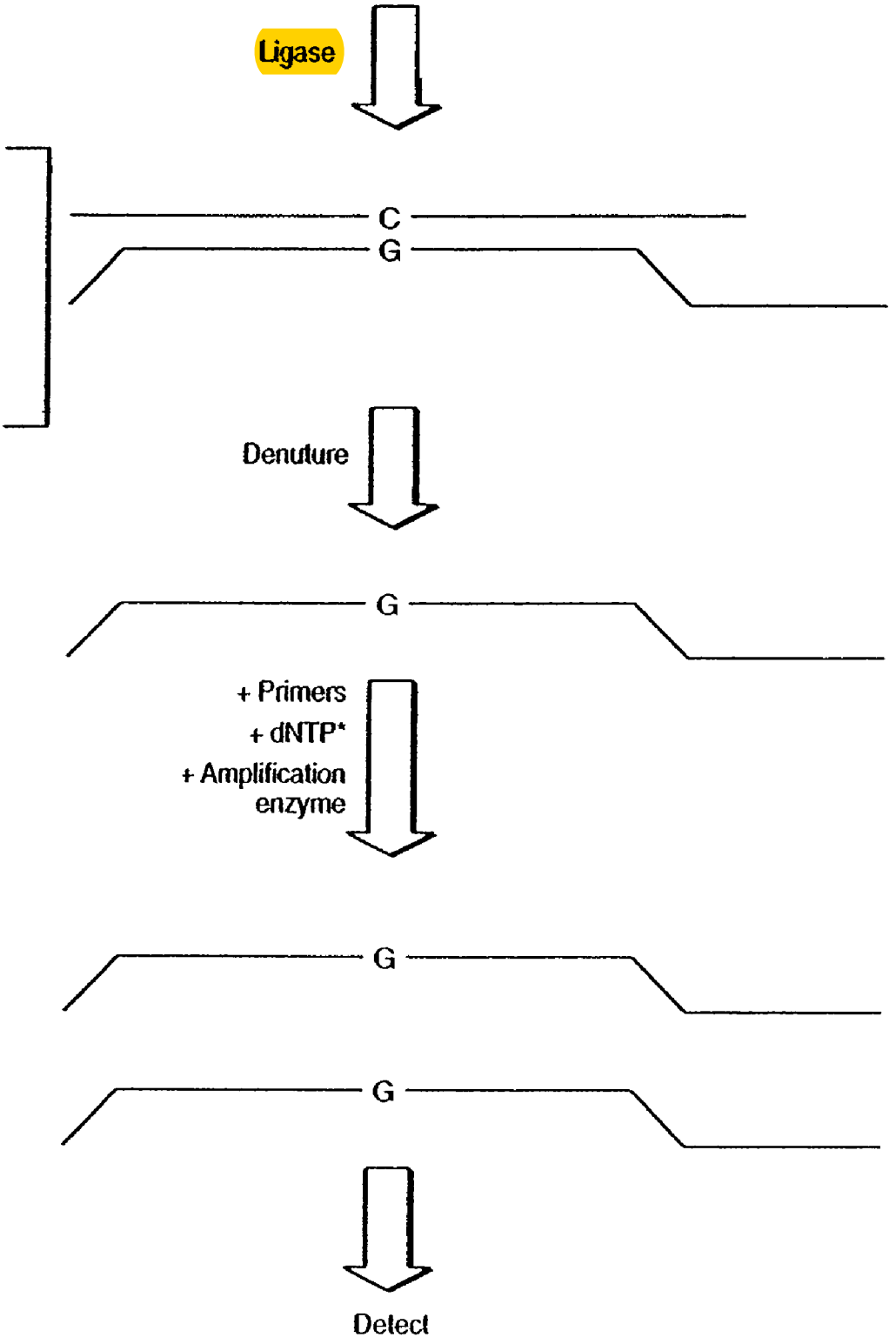

Ligase
C
G
Denuture
G
+ Primers
+ dNTP*
+ Amplification enzyme
G
G
Detect

FIG._13B

MULTIPLEX NUCLEIC ACID REACTIONS

The present application claims the benefit of U.S. Application Ser. Nos. 60/234,143, filed on Sep. 21, 2000, 60/234,732, filed on Sep. 22, 2000, 60/297,609, filed on Jun. 11, 2001, 60/311,271, filed on Aug. 9, 2001, 60/336,958, filed on Dec. 3, 2001, 60/305,118, filed on Jul. 12, 2001, and 60/341,827, filed on Dec. 17, 2001 and claims priority to Ser. No. 09/779,376, now abandoned, filed on Feb. 7, 2001 and WO 01/57269, filed on Feb. 7, 2001, Ser. No. 09/915,231, now U.S. Pat. No. 6,890,741, filed on Jul. 24, 2001 and Ser. No. 09/931,285, now U.S. Pat. No. 6,931,884, filed on Aug. 16, 2001, all of which are expressly incorporated herein by reference.

Portions of this invention were made with government support under HG02003 awarded by the National Human Genome Research Institute and CA81952 awarded by the National Cancer Institute. The government has certain rights in the invention.

FIELD OF THE INVENTION

The invention is directed to a variety of multiplexing methods used to amplify and/or genotype a variety of samples simultaneously.

BACKGROUND OF THE INVENTION

The detection of specific nucleic acids is an important tool for diagnostic medicine and molecular biology research. Gene probe assays currently play roles in identifying infectious organisms such as bacteria and viruses, in probing the expression of normal genes and identifying mutant genes such as oncogenes, in typing tissue for compatibility preceding tissue transplantation, in matching tissue or blood samples for forensic medicine, and for exploring homology among genes from different species.

Ideally, a gene probe assay should be sensitive, specific and easily automatable (for a review, see Nickerson, Current Opinion in Biotechnology 4:48-51 (1993)). The requirement for sensitivity (i.e. low detection limits) has been greatly alleviated by the development of the polymerase chain reaction (PCR) and other amplification technologies which allow researchers to amplify exponentially a specific nucleic acid sequence before analysis as outlined below (for a review, see Abramson et al., Current Opinion in Biotechnology, 4:41-47 (1993)).

Currently, a variety of biochips comprising substrates with associated nucleic acids are used in a variety of nucleic acid detection systems, including the detection, quantification, sequence determination and genotyping of a nucleic acid target sequences. However, sample preparation for these high density chips remains an issue.

Accordingly, it is an object of the invention to provide a number of methods directed to the multiplexing amplification and/or genotyping reactions of target sequences to create amplicons that can subsequently be detected on an array.

SUMMARY OF THE INVENTION

In accordance with the objects outlined above, the present invention provides a method of detecting target sequences in a sample comprising providing a first solid support comprising at least a first and a second target sequence, contacting the first and second target sequences with first and second probes, respectively, wherein each of the first and second probes comprise a first universal priming site, a target specific domain substantially complementary to at least a portion of the target sequence, to form first and second hybridization complexes, respectively, removing unhybridized probes, contacting the first and second hybridization complexes with a first enzyme to form modified first and second probes, respectively contacting the modified first and second probes with at least a first primer that hybridizes to the universal priming site NTPs, and an extension enzyme, wherein the first and second modified probes are amplified to form first and second amplicons, respectively, and detecting the amplicons.

In addition the invention provides a method of detecting target sequences in a sample comprising providing a first solid support comprising at least a first and a second target sequence, contacting the first and second target sequences with first and second probes, respectively, wherein each of the first and second probes comprise a first universal priming site, a target specific domain substantially complementary to at least a portion of the target sequence, to form first and second hybridization complexes, respectively, removing unhybridized probes, contacting the first and second probes with at least a first universal primer that hybridizes to the universal priming site, NTPs and an extension enzyme, wherein the first and second probes are extended to form first and second modified probes, respectively, contacting the first and second modified probes with at least third and fourth probes, respectively, wherein the modified first and second probes comprise a detection position, the third and fourth probes each comprise an interrogation position, and a second enzyme, wherein the second enzyme only modifies the third and fourth probes if there is perfect complementarity between the bases at the interrogation position and the detection position, forming third and fourth modified probes, and detecting the third and fourth modified probes.

In addition the invention provides a method comprising providing a plurality of target nucleic acid sequences each comprising from 3' to 5' a first, second and third target domain, the first target domain comprising a detection position, the second target domain being at least one nucleotide contacting the target nucleic acid sequences with sets of probes for each target sequence, each set comprising a first probe comprising from 5' to 3' a first domain comprising a first universal priming sequence, and a second domain comprising a sequence substantially complementary to the first target domain of a target sequence, and an interrogation position within the 3' four terminal bases, a second probe comprising a first domain comprising a sequence substantially complementary to the third target domain of a target sequence, to form a set of first hybridization complexes, contacting the first hybridization complexes with an extension enzyme and dNTPs, under conditions whereby if the base at the interrogation positions is perfectly complementary with the bases at the detection positions, extension of the first probes occurs through the second target domains to form second hybridization complexes, contacting the second hybridization complexes with a ligase to ligate the extended first probes to the second probes to form amplification templates.

In addition the invention provides a multiplex reaction method comprising providing a sample comprising at least first and second targets hybridizing the first and second targets with first and second probes, respectively forming first and second hybridization complexes, respectively, immobilizing the first and second hybridization complexes, washing to remove unhybridized nucleic acids, contacting the first and second hybridization complexes with an enzyme, whereby the first and second probes are modified forming modified first and second probes, respectively, whereby the modified

first and second probes are modified to contain first and second interrogation nucleotides that are complementary to first and second detection nucleotides in the first and second targets, respectively, contacting the modified first and second probes with first and second allele specific primers, respectively, whereby the first and second allele specific primers hybridize to the modified first and second probes, respectively, 5' to the first and second interrogation nucleotides, dNTPs, polymerase, whereby the first and second allele specific primers are modified when a target domain of the allele specific primers is perfectly complementary to the modified target probes to form modified first and second allele specific probes, amplifying the modified first and second allele specific probes to form first and second amplicons, and detecting the first and second amplicons.

In addition the invention provides a method comprising providing a plurality of target nucleic acid sequences each comprising from 3' to 5' a first, second and third target domain, the first target domain comprising a detection position, the second target domain being at least one nucleotide, contacting the target nucleic acid sequences with sets of probes for each target sequence, each set comprising: a first probe comprising from 5' to 3', a first domain comprising a first universal priming sequence, and a second domain comprising a sequence substantially complementary to the first target domain of a target sequence, and an interrogation position within the 3' four terminal bases, a second probe comprising a first domain comprising a sequence substantially complementary to the third target domain of a target sequence, to form a set of first hybridization complexes, contacting the first hybridization complexes with at least a first universal primer that hybridize to the first universal priming sequence, an extension enzyme and dNTPs, under conditions whereby if the base at the interrogation positions are perfectly complementary with the bases at the detection positions, extension of the first probes occurs through the second target domains to form second hybridization complexes, contacting the second hybridization complexes with a ligase to ligate the extended first probes to the second probes to form amplification templates.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** depicts a schematic of a preferred embodiment of the invention. The primary steps of the method include annealing oligonucleotides to immobilized target (e.g. genomic) DNA, a chain extension reaction that is terminated by tagged (e.g. biotinylated) ddNTPs (FIG. **1**A), isolation and amplification of the tagged extension products (FIG. **1**B).

FIG. **2** depicts a preferred method of obtaining templates for single base extension reaction analysis. The four major phases are 1) First Extension from target (gDNA), 2) Second Extension, 3) PCR Amplification, and 4) Allele Specific 4-Dye Single Base Extension.

FIG. **3** depicts a preferred method for complexity reduction and allele selectivity. The locus specific primer hybridizes upstream of the interrogation site. It does not have to be directly adjacent to the interrogation site. The locus specific primer also contains an adapter sequence and universal PCR primer hybridization site. The allele specific primers are designed to the opposite strand of DNA (see diagram) and the 3' ends of the primers correspond to the alleles that are interrogated. The 5' ends of the allele specific primers are hybridization sites for universal PCR primers.

Tagged locus specific primers are annealed to the genomic DNA and washed. DNA polymerase (Taq DNA polymerase), dNTPs, ddNTPs and buffer is added to the hybridized primers. The DNA polymerase will extend the locus specific primers that have hybridized and are matched exactly at the 3' end to DNA. In this first primer extension reaction, the primer extended product has captured the locus allele information and also adjacent DNA sequence information. The primer extension products are eluted away from the genomic DNA. The eluted primer extension products are captured onto another set of streptavidin coated beads through the biotin molecule on the locus specific primer. This capture process purifies the primer extension product and reduces the complexity of DNA going into the second hybridization and extension process. The second capture process may improve the multiplexability of this assay through the reduction of complexity.

Allele specific primers for each interrogated locus are added to the captured DNA and a second hybridization and wash is performed (at high stringency). DNA polymerase (Taq DNA polymerase), dNTPs, and buffer are added to the hybridized primers. An extension reaction is carried out. The extended products are eluted and used in a PCR amplification reaction (using the universal PCR primers specific for these oligos U1, U2 and U3). U2 and U3 are labeled with different fluorescent tags. The ratio in the amount of one allele relative to another is determined by the ratio of the fluorescent tags.

FIG. **4** depicts an alternative embodiment of the method outlined in FIG. **3**. An allele specific hybridization approach for allele determination may be used in conjunction with the first hybridization, wash and extension. In this process, the locus specific primer is hybridized, washed and extended as above. The locus specific primer does not contain adapter sequences or universal primer sequences. The allele specific oligonucleotide contains the universal PCR primer sequences. Allele specific oligonucleotides are added to the extended products, hybridized and washed under stringent conditions. Allele specifically hybridized sequences are retained and later eluted for a PCR reaction.

FIG. **5** depicts an alternative embodiment of the method outlined in FIG. **3**. In this embodiment allele specific extension is followed by locus specific extension.

FIG. **6** depicts an alternative embodiment of the method outlined in FIG. **3**. A second level of allele specificity along with locus specificity may be obtained by using allele specific extension primers in the second extension step of FIG. **5**. Using allele specific extension primers (on alternate strands) in both extension steps would protect against any 3' to 5' exonuclease activity acting in the first allele specific extension step. The extension products from this approach would be placed into two separate PCR reactions containing universal PCR primers specific for each allele set. Misextensions due to exonuclease activity in the first or second extension steps would not be amplified.

FIG. **7** depicts a preferred method of solid-phase allele-specific primer extension genotyping. For each locus, two allele specific oligonucleotides are designed with each allele represented by a unique adapter. The 3' end of the allele specific oligonucleotides extend one or more bases beyond the query site. The oligonucleotides are hybridized to the template on solid phase under stringent conditions. The solid phase is washed to remove improperly hybridized oligonucleotides. The resulting complex is then extended by a polymerase in an allele specific manner. That is a mismatch at the query site will prevent efficient extension.

FIG. **8** depicts an alternative method of labeling as compared to FIG. **7**.

FIG. **9** depicts a schematic of universal allele specific oligonucleotides.

FIG. **10** depicts a method using the universal allele specific oligonucleotides described in FIG. **9**. In this case, since extension must occur from 5' to 3', the U4 and U5 sequences are shown at the 3' end of the template, associated with the allele-selective bases.

FIG. **11** depicts a method of removing non-hybridized nucleic acids by nuclease treatment. That is, the complexity of a nucleic acid sample is initially reduced by hybridization capture with gene specific oligonucleotides. Excess nucleic acid sequences are removed by a single stranded nuclease.

FIGS. **12**A and **12**B depicts the ICAN amplification scheme.

FIGS. **13**A and **13**B depicts a preferred multiplex scheme. Two primers hybridize to a target nucleic acid (FIG. **13**A). The primers include target specific portions and universal priming sites. In addition, one of the primers, preferably the upstream primer, includes an allele specific sequence and an adapter sequence that is specific for the particular allele specific sequence. The primers do not hybridize contiguously on the target. Following hybridization the primer is extended with dNTPs and a polymerase. Following primer extension, the upstream and downstream primers are ligated (FIG. **13**B). The ligated product is them amplified with universal primers that hybridize to the universal priming sites on the primers resulting in the formation amplicons. Amplicons are labeled with either labeled primers or labeled dNTPs and detected as an indication of the presence of a particular allele.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to a variety of compositions and methods directed to multiplexed analysis of nucleic acids. In a preferred embodiment the methods are directed to multiplexing of nucleic acid detection, genotyping and amplification reactions. While a large body of literature and methods exist for the use of high density biochips comprising nucleic acids, the preparation of samples containing target sequences to place on the biochips has not been significantly multiplexed to allow true high throughput methodologies. The present invention is directed to the use of a variety of methods that allow the multiplexed amplification of target sequences prior to detection by any of a variety of methods including placement on an array for detection, mass spectrometry, electrophoretic techniques, FACS analysis, and the like.

In general, the method includes a complexity reduction component, a specificity step and an amplification step. Preferably complexity reduction is performed first. This is followed, in some embodiments, by the genotyping reaction, followed by multiplexed amplification. Generally, the specificity step includes an enzymatic reaction such as a genotyping reaction as described below. Alternatively, the multiplexed amplification reaction is done first, i.e. following complexity reduction, followed by a genotyping reaction. In both instances, the resulting amplicons are then detected, by a variety of detection methods including utilizing solid support arrays (both random and ordered), liquid arrays, or using technologies such as FACS sorting or mass spectroscopy.

Accordingly, the present invention relates to the multiplex amplification and detection of target analytes in a sample. As used herein, the phrase "multiplex" or grammatical equivalents refers to the detection, analysis or amplification of more than one target sequence of interest. In one embodiment multiplex refers to at least 100 or 200 different target sequences while at least 500 different target sequences is preferred. More preferred is at least 1000, with more than 5000 or 10,000 particularly preferred and more than 50,000 or 100,000 most preferred. Detection is performed on a variety of platforms as described herein.

Accordingly, the present invention provides methods for the detection of nucleic acid target sequences in a sample. As will be appreciated by those in the art, the sample solution may comprise any number of things, including, but not limited to, bodily fluids (including, but not limited to, blood, urine, serum, lymph, saliva, anal and vaginal secretions, perspiration and semen, of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred); environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples; purified samples, such as purified genomic DNA, RNA, proteins, etc.; raw samples (bacteria, virus, genomic DNA, etc.). As will be appreciated by those in the art, virtually any experimental manipulation may have been done on the sample.

If required, the target sequence is prepared using known techniques. For example, the sample may be treated to lyse the cells, using known lysis buffers, sonication, electroporation, etc., with purification and amplification as outlined below occurring as needed, as will be appreciated by those in the art. In addition, the reactions outlined herein may be accomplished in a variety of ways, as will be appreciated by those in the art. Components of the reaction may be added simultaneously, or sequentially, in any order, with preferred embodiments outlined below. In addition, the reaction may include a variety of other reagents which may be included in the assays. These include reagents like salts, buffers, neutral proteins, e.g. albumin, detergents, etc., which may be used to facilitate optimal hybridization and detection, and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used, depending on the sample preparation methods and purity of the target.

In addition, when nucleic acids are to be detected preferred methods utilize cutting or shearing techniques to cut the nucleic acid sample containing the target sequence into a size that will facilitate handling and hybridization to the target, particularly for genomic DNA samples. This may be accomplished by shearing the nucleic acid through mechanical forces (e.g. sonication) or by cleaving the nucleic acid using restriction endonucleases, or any other methods known in the art.

In addition, in most embodiments, double stranded target nucleic acids are denatured to render them single stranded so as to permit hybridization of the primers and other probes of the invention. A preferred embodiment utilizes a thermal step, generally by raising the temperature of the reaction to about 95° C., although pH changes and other techniques may also be used.

The present invention provides compositions and methods for detecting the presence or absence of target nucleic acid sequences in a sample. By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487 (1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and

Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et al., Nucleic Acids Res. 19:1437 (1991); and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are incorporated by reference). Other analog nucleic acids include those with positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. English 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrahedron Lett. 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins et al., Chem. Soc. Rev. (1995) pp 169-176). Several nucleic acid analogs are described in Rawls, C & E News Jun. 2, 1997 page 35. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of labels, or to increase the stability and half-life of such molecules in physiological environments.

As will be appreciated by those in the art, all of these nucleic acid analogs may find use in the present invention. In addition, mixtures of naturally occurring nucleic acids and analogs can be made.

Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

Particularly preferred are peptide nucleic acids (PNA) which includes peptide nucleic acid analogs. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This results in two advantages. First, the PNA backbone exhibits improved hybridization kinetics. PNAs have larger changes in the melting temperature (Tm) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4° C. drop in Tm for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9° C. This allows for better detection of mismatches. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration.

The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc. A preferred embodiment utilizes isocytosine and isoguanine in nucleic acids designed to be complementary to other probes, rather than target sequences, as this reduces non-specific hybridization, as is generally described in U.S. Pat. No. 5,681,702. As used herein, the term "nucleoside" includes nucleotides as well as nucleoside and nucleotide analogs, and modified nucleosides such as amino modified nucleosides. In addition, "nucleoside" includes non-naturally occurring analog structures. Thus for example the individual units of a peptide nucleic acid, each containing a base, are referred to herein as a nucleoside.

The compositions and methods of the invention are directed to the multi-plexed detection of target sequences. The term "target sequence" or "target nucleic acid" or grammatical equivalents herein means a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA and rRNA, or others. A preferred embodiment utilizes genomic DNA as the primary target sequence.

As is outlined herein, the target sequence may be a target sequence from a sample, or a secondary target such as a product of a reaction such as a detection sequence from an invasive cleavage reaction, a ligated probe from an OLA reaction, an extended probe from a PCR reaction, or PCR amplification product, ("amplicon") etc.

The target sequence may be any length, with the understanding that longer sequences are more specific. As will be appreciated by those in the art, the complementary target sequence may take many forms. For example, it may be contained within a larger nucleic acid sequence, i.e. all or part of a gene or mRNA, a restriction fragment of a plasmid or genomic DNA, among others. As is outlined more fully below, probes are made to hybridize to target sequences to determine the presence or absence of the target sequence in a sample. Generally speaking, this term will be understood by those skilled in the art. The target sequence may also be comprised of different target domains; for example, in "sandwich" type assays as outlined below, a first target domain of the sample target sequence may hybridize to a capture probe or a portion of capture extender probe, a second target domain may hybridize to a portion of an amplifier probe, a label probe, or a different capture or capture extender probe, etc. In addition, the target domains may be adjacent (i.e. contiguous) or separated. For example, when LCR techniques are used, a first primer may hybridize to a first target domain and a second primer may hybridize to a second target domain; either the domains are adjacent, or they may be separated by one or more nucleotides, coupled with the use of a polymerase and dNTPs, as is more fully outlined below.

The terms "first" and "second" are not meant to confer an orientation of the sequences with respect to the 5'-3' orientation of the target sequence. For example, assuming a 5'-3' orientation of the complementary target sequence, the first target domain may be located either 5' to the second domain, or 3' to the second domain.

As outlined herein, in preferred embodiments the target sequence comprises a position for which sequence information is desired, generally referred to herein as the "detection position" or "detection locus". In a preferred embodiment, the detection position is a single nucleotide, although in some embodiments, it may comprise a plurality of nucleotides, either contiguous with each other or separated by one or more nucleotides. By "plurality" as used herein is meant at least two. As used herein, the base which basepairs with a detection position base in a hybrid is termed a "readout position" or an "interrogation position"; thus many of the first or second step probes of the invention comprise an interrogation position.

In some embodiments, as is outlined herein, the target sequence may not be the sample target sequence but instead is

a product of a reaction herein, sometimes referred to herein as a "secondary" or "derivative" target sequence, or an "amplicon".

Accordingly, in a preferred embodiment the present multiplexed detection scheme includes at least one complexity reduction component, at least one specificity component and at least one amplification component. In addition, the method includes detection of the product of the reaction.

The methods of the invention can take on a wide variety of configurations, as are shown in the figures and described in more detail below. Generally these components include a complexity reduction component, a specificity component and an amplification component. The components can be configured in a variety of ways as disclosed below. That is, in one embodiment a complexity reduction step is first performed. This is followed by either the amplification or specificity step. Alternatively, the specificity step is performed first. This can be followed by the complexity reduction or amplification step. Alternatively, amplification is first performed. This is followed by the complexity and specificity steps.

While the above indicates that each of the three components can be performed in any order. One of skill in the art will appreciate that when amplification is performed first, there will likely be some degree of complexity reduction or specificity involved. In addition, when specificity components are performed first, there will be a degree of complexity reduction. In addition, in some embodiments when amplification is first performed, there will be some degree of specificity and complexity reduction. However, as described below, the method generally includes three components.

Probes and Primers

As one of skill in the art appreciates, there are several probes or primers that are used in the present invention. These probes/primers can take on a variety of configurations and may have a variety of structural components described in more detail below. The first step probe may be either an allele specific probe or locus specific probe. By "allele specific" probe or primer is meant a probe or primer that either hybridizes to a target sequence and discriminates between alleles or hybridizes to a target sequence and is modified in an allele specific manner. By "locus specific" probe or primer is meant a probe or primer that hybridizes to a target sequence in a locus specific manner, but does not necessarily discriminate between alleles. A locus specific primer also may be modified, i.e. extended as described below, such that it includes information about a particular allele, but the locus specific primer does not discriminate between alleles.

In many embodiments, the probes or primers comprise one or more universal priming site(s) and/or adapters, both of which are described below.

The size of the primer and probe nucleic acid may vary, as will be appreciated by those in the art with each portion of the probe and the total length of the probe in general varying from 5 to 500 nucleotides in length. Each portion is preferably between 10 and 100 being preferred, between 15 and 50 being particularly preferred, and from 10 to 35 being especially preferred, depending on the use and amplification technique. Thus, for example, the universal priming site(s) of the probes are each preferably about 15-20 nucleotides in length, with 18 being especially preferred. The adapter sequences of the probes are preferably from 15-25 nucleotides in length, with 20 being especially preferred. The target specific portion of the probe is preferably from 15-50 nucleotides in length. In addition, the primer may include an additional amplification priming site. In a preferred embodiment the additional amplification priming site is a T7 RNA polymerase priming site.

In a preferred embodiment, the allele or locus specific probe or probes comprises a target domain substantially complementary to a first domain of the target sequence. In general, probes of the present invention are designed to be complementary to a target sequence (either the target sequence of the sample or to other probe sequences, as is described herein), such that hybridization of the target and the probes of the present invention occurs. This complementarity need not be perfect; there may be any number of base pair mismatches that will interfere with hybridization between the target sequence and the single stranded nucleic acids of the present invention. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. Thus, by "substantially complementary" herein is meant that the probes are sufficiently complementary to the target sequences to hybridize under the selected reaction conditions.

In one embodiment the target specific portion includes a combinatorial mixture of each nucleotide at each position. In addition the primer includes a universal priming sequence and an allele specific position. Preferably the universal priming sequence is specific for the particular nucleotide at the allele specific position. That is, in this embodiment the locus-specific allele selectivity portions of the primer are replaced with a universal targeting domain that includes region where each position is represented by a combinatorial mixture of nucleotides. One of the positions in the universal region (not necessarily the 3' position) is paired with the allele or SNP to be analyzed. The base at this position is associated with an identifier such as a particular adapter in the primer or with a particular universal priming sequence in the primer (FIG. **9**).

In a preferred configuration, each of the four bases is associated with a different sequence, i.e. universal priming sequence, each sequence having similar amplification efficiencies. For amplification, each of the four primers is labeled with a different label. In an alternate embodiment it is possible to substitute a universal, i.e. promiscuous (inosine, for example) base at one or more positions in the universal sequence. The primer finds use in extension reactions and ligation reactions as described herein. In addition the primers find use in linear amplification schemes as depicted in FIG. **10**. It should be noted that one advantage of using the universal targeting domain is that shorter oligonucleotides can be used. Thus, when universal target domains are used, these domains are preferably from about 5 to 15 nucleotides in length with from 7 to 10 being particularly preferred.

In a preferred embodiment, one of the probes further comprises an adapter sequence, (sometimes referred to in the art as "zip codes" or "bar codes"). Adapters facilitate immobilization of probes to allow the use of "universal arrays". That is, arrays (either solid phase or liquid phase arrays) are generated that contain capture probes that are not target specific, but rather specific to individual (preferably) artificial adapter sequences.

Thus, an "adapter sequence" is a nucleic acid that is generally not native to the target sequence, i.e. is exogenous, but is added or attached to the target sequence. It should be noted that in this context, the "target sequence" can include the primary sample target sequence, or can be a derivative target such as a reactant or product of the reactions outlined herein; thus for example, the target sequence can be a PCR product, a first ligation probe or a ligated probe in an OLA reaction, etc. The terms "barcodes", "adapters", "addresses", "tags" and "zipcodes" have all been used to describe artificial sequences that are added to amplicons to allow separation of nucleic acid fragment pools. One preferred form of adapters

are hybridization adapters. In this embodiment adapters are chosen so as to allow hybridization to the complementary capture probes on a surface of an array. Adapters serve as unique identifiers of the probe and thus of the target sequence. In general, sets of adapters and the corresponding capture probes on arrays are developed to minimize cross-hybridization with both each other and other components of the reaction mixtures, including the target sequences and sequences on the larger nucleic acid sequences outside of the target sequences (e.g. to sequences within genomic DNA). Other forms of adapters are mass tags that can be separated using mass spectroscopy, electrophoretic tags that can be separated based on electrophoretic mobility, etc. Some adapter sequences are outlined in U.S. Ser. No. 09/940,185, filed Aug. 27, 2001, hereby incorporated by reference in its entirety. Preferred adapters are those that meet the following criteria. They are not found in a genome, preferably a human genome, and they do not have undesirable structures, such as hairpin loops.

As will be appreciated by those in the art, the attachment, or joining, of the adapter sequence to the target sequence can be done in a variety of ways. In a preferred embodiment, the adapter sequences are added to the primers of the reaction (extension primers, amplification primers, readout probes, genotyping primers, Rolling Circle primers, etc.) during the chemical synthesis of the primers. The adapter then gets added to the reaction product during the reaction; for example, the primer gets extended using a polymerase to form the new target sequence that now contains an adapter sequence. Alternatively, the adapter sequences can be added enzymatically. Furthermore, the adapter can be attached to the target after synthesis; this post-synthesis attachment can be either covalent or non-covalent. In a preferred embodiment the adapter is added to the target sequence or associated with a particular allele during an enzymatic step. That is, to achieve the level of specificity necessary for highly multiplexed reactions, the product of the specificity or allele specific reaction preferably also includes at least one adapter sequence.

In this embodiment, one or more of the specificity primers comprises a first portion comprising the adapter sequence and a second portion comprising the priming sequence. Extending the amplification primer as is well known in the art results in target sequences that comprise the adapter sequences. The adapter sequences are designed to be substantially complementary to capture probes.

In addition, as will be appreciated by those in the art, the adapter can be attached either on the 3' or 5' ends, or in an internal position, depending on the configuration of the system, as generally outlined in the figures.

In one embodiment the use of adapter sequences allow the creation of more "universal" surfaces; that is, one standard array, comprising a finite set of capture probes can be made and used in any application. The end-user can customize the array by designing different soluble target probes, which, as will be appreciated by those in the art, is generally simpler and less costly. In a preferred embodiment, an array of different and usually artificial capture probes are made; that is, the capture probes do not have complementarity to known target sequences. The adapter sequences can then be incorporated in the target probes.

As will be appreciated by those in the art, the length of the adapter sequences will vary, depending on the desired "strength" of binding and the number of different adapters desired. In a preferred embodiment, adapter sequences range from about 6 to about 500 basepairs in length, with from about 8 to about 100 being preferred, and from about 10 to about 25 being particularly preferred.

In a preferred embodiment, the adapter sequence uniquely identifies the target analyte to which the target probe binds. That is, while the adapter sequence need not bind itself to the target analyte, the system allows for identification of the target analyte by detecting the presence of the adapter. Accordingly, following a binding or hybridization assay and washing, the probes including the adapters are amplified. Detection of the adapter then serves as an indication of the presence of the target analyte.

In one embodiment the adapter includes both an identifier region and a region that is complementary to capture probes on a universal array as described above. In this embodiment, the amplicon hybridizes to capture probes on a universal array. Detection of the adapter is accomplished following hybridization with a probe that is complementary to the adapter sequence. Preferably the probe is labeled as described herein.

In general, unique adapter sequences are used for each unique target analyte. That is, the elucidation or detection of a particular adapter sequence allows the identification of the target analyte to which the target probe containing that adapter sequence bound. However, in some cases, it is possible to "reuse" adapter sequences and have more than one target analyte share an adapter sequence.

In a preferred embodiment the adapters contain different sequences or properties that are indicative of a particular target molecule. That is, each adapter uniquely identifies a target sequence. As described above, the adapters are amplified to form amplicons. The adapter is detected as an indication of the presence of the target analyte, i.e. the particular target nucleic acid.

The use of adapters in combination with amplification following a specific binding event allows for highly multiplexed reactions to be performed.

Also, the probes are constructed so as to contain the necessary priming site or sites for the subsequent amplification scheme. In a preferred embodiment the priming sites are universal priming sites. By "universal priming site" or "universal priming sequences" herein is meant a sequence of the probe that will bind a primer for amplification.

In a preferred embodiment, one universal priming sequence or site is used. In this embodiment, a preferred universal priming sequence is the RNA polymerase T7 sequence, that allows the T7 RNA polymerase make RNA copies of the adapter sequence as outlined below. Additional disclosure regarding the use of T7 RNA polymerase is found in U.S. Pat. Nos. 6,291,170, 5,891,636, 5,716,785, 5,545,522, 5,922,553, 6,225,060 and 5,514,545, all of which are expressly incorporated herein by reference.

In a preferred embodiment, for example when amplification methods requiring two primers such as PCR are used, each probe preferably comprises an upstream universal priming site (UUP) and a downstream universal priming site (DUP). Again, "upstream" and "downstream" are not meant to convey a particular 5'-3' orientation, and will depend on the orientation of the system. Preferably, only a single UUP sequence and a single DUP sequence is used in a probe set, although as will be appreciated by those in the art, different assays or different multiplexing analysis may utilize a plurality of universal priming sequences. In some embodiments probe sets may comprise different universal priming sequences. In addition, the universal priming sites are preferably located at the 5' and 3' termini of the target probe (or the ligated probe), as only sequences flanked by priming sequences will be amplified.

In addition, universal priming sequences are generally chosen to be as unique as possible given the particular assays and

host genomes to ensure specificity of the assay. However, as will be appreciated by those in the art, sets of priming sequences/primers may be used; that is, one reaction may utilize 500 target probes with a first priming sequence or set of sequences, and an additional 500 probes with a second sequence or set of sequences.

As will be appreciated by those in the art, when two priming sequences are used, the orientation of the two priming sites is generally different. That is, one PCR primer will directly hybridize to the first priming site, while the other PCR primer will hybridize to the complement of the second priming site. Stated differently, the first priming site is in sense orientation, and the second priming site is in antisense orientation.

As will be appreciated by those in the art, in general, highly multiplexed reactions can be performed, with all of the universal priming sites being the same for all reactions. Alternatively, "sets" of universal priming sites and corresponding probes can be used, either simultaneously or sequentially. The universal priming sites are used to amplify the modified probes to form a plurality of amplicons that are then detected in a variety of ways, as outlined herein. In preferred embodiments, one of the universal priming sites is a T7 site. In some embodiments this priming site serves as a template for the synthesis of RNA.

Accordingly, the present invention provides first target probe sets. By "probe set" herein is meant a plurality of target probes that are used in a particular multiplexed assay. In this context, plurality means at least two, with more than 10 being preferred, depending on the assay, sample and purpose of the test. In one embodiment the probe set includes more than 100, with more than 500 probes being preferred and more than 1000 being particularly preferred. In a particularly preferred embodiment each probe contains at least 5000, with more than 10,000 probes being most preferred.

Accordingly, the present invention provides first target probe sets that each comprise at least a first universal priming site.

In a preferred embodiment, the target probe may also comprise a label sequence, i.e. a sequence that can be used to bind label probes and is substantially complementary to a label probe. This system is sometimes referred to in the art as "sandwich-type" assays. That is, by incorporating a label sequence into the target probe, which is then amplified and present in the amplicons, a label probe comprising primary (or secondary) detection labels can be added to the mixture, either before addition to the array or after. This allows the use of high concentrations of label probes for efficient hybridization. In one embodiment, it is possible to use the same label sequence and label probe for all target probes on an array; alternatively, different target probes can have a different label sequence. Similarly, the use of different label sequences can facilitate quality control; for example, one label sequence (and one color) can be used for one strand of the target, and a different label sequence (with a different color) for the other; only if both colors are present at the same basic level is a positive called.

Thus, the present invention provides target probes that comprise any, all or any combination of universal priming sequences, bioactive agents (e.g. target specific portion(s)), adapter sequence(s), optionally an additional amplification priming sequence such as T7 RNA priming sequence and optionally label sequences. These target probes are then added to the target sequences to form hybridization complexes. As will be appreciated by those in the art, when nucleic acids are the target, the hybridization complexes contain portions that are double stranded (the target-specific sequences of the target probes hybridized to a portion of the target sequence) and portions that are single stranded (the ends of the target probes comprising the universal priming sequences and the adapter sequences, and any unhybridized portion of the target sequence, such as poly(A) tails, as outlined herein).

Complexity Reduction

Complexity reduction is a principal component of the multiplex scheme set forth herein. Generally, complexity reduction is a method for enriching for a particular target or locus. That is, complexity reduction is considered a method that results in removal of non-target nucleic acids from the sample or removal of probes/primers that have not hybridized correctly or at all to a target nucleic acid. In addition, complexity reduction includes removal of probes that have not been modified during a enzymatic step. That is, complexity reduction includes removing non-target nucleic acids, i.e. enriching for target nucleic acids or removing non-hybridized probes or primers prior to an enzymatic step, i.e. either an amplification or specificity step, or both.

There are a variety of ways one can include a complexity reduction step. These include, but are not limited to, selective immobilization of target nucleic acids or probes/primers that are modified in a target specific manner, selective removal of non-target nucleic acids, and selective destruction of non-target nucleic acids. Such destruction includes but is not limited to denaturation, degradation or cleavage of non-target nucleic acids. In addition, complexity reduction can include components such as target selective amplification, although this also includes amplification and components.

In a preferred embodiment complexity reduction is accomplished by selectively immobilizing a primer that has been modified in a target specific manner. That is, either locus specific or allele specific primers are hybridized with a target. The target can be immobilized or in solution. Following hybridization, the primer is extended in a primer extension reaction. Preferably either the primer or NTPs include a purification tag as described herein that allows for removal or purification of the extended product from the reaction mixture. Once extended, generally the modified primer is immobilized on a solid support as described herein. Following immobilization of the modified primer, the support is washed to remove both non-target nucleic acids and primers that were not modified, i.e. extended. The immobilized primers, thus, include information about the target locus including particular allelic information. This results in enrichment of target nucleic acids or removal of non-target nucleic acids.

In a preferred embodiment the complexity reduction component includes selective immobilization of target nucleic acids. That is, target nucleic acids are preferentially immobilized on a solid support rather than non-target nucleic acids.

In this embodiment target DNA is preferably reduced in size initially. This is easily accomplished by methods as known in the art such as, but not limited to, shearing or cleaving with restriction enzymes. The target nucleic acid is contacted with probes that hybridize to the targets. Preferably the hybridization is performed under low stringency conditions such that the probes do not discriminate between alleles of a particular locus. The resulting complexes are then immobilized on a support. In a preferred embodiment the probes are labeled with a purification tag as described herein to allow for immobilization. Following immobilization, the support is washed to remove non-hybridized targets, while leaving targets that are substantially complementary to the probes immobilized on the solid support. After removal of non-hybridized probes, the target nucleic acids can be removed

with a stringent wash. This allows for enrichment of target sequences that are then available for further analysis.

In one embodiment, the target sequence, probe or primer, including modified primer, is attached to a first solid support. By "substrate" or "solid support" or other grammatical equivalents herein is meant any material that is appropriate for or can be modified to be appropriate for the attachment of the target sequences. As will be appreciated by those in the art, the number of possible substrates is very large. Possible substrates include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon™, etc.), polysaccharides, nylon or nitrocellulose, ceramics, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. Magnetic beads and high throughput microtier plates are particularly preferred.

The composition and geometry of the solid support vary with its use. In this particular embodiment, supports comprising microspheres or beads are preferred for the first solid support. By "microspheres" or "beads" or "particles" or grammatical equivalents herein is meant small discrete particles. The composition of the beads will vary, depending on the class of bioactive agent and the method of synthesis. Suitable bead compositions include those used in peptide, nucleic acid and organic moiety synthesis, including, but not limited to, plastics, ceramics, glass, polystyrene, methylstyrene, acrylic polymers, paramagnetic materials, thoria sol, carbon graphited, titanium dioxide, latex or cross-linked dextrans such as Sepharose, cellulose, nylon, cross-linked micelles and teflon, as well as any other materials outlined herein for solid supports may all be used. "*Microsphere Detection Guide*" from Bangs Laboratories, Fishers Ind. is a helpful guide. Preferably, in this embodiment, when complexity reduction is performed, the microspheres are magnetic microspheres or beads.

The beads need not be spherical; irregular particles may be used. In addition, the beads may be porous, thus increasing the surface area of the bead available for assay. The bead sizes range from nanometers, i.e. 100 nm, to millimeters, i.e. 1 mm, with beads from about 0.2 micron to about 200 microns being preferred, and from about 0.5 to about 5 micron being particularly preferred, although in some embodiments smaller beads may be used.

The target sequence, probe or primer is attached to the first solid support in a number of ways. In a preferred embodiment, purification tags are used. By "purification tag" herein is meant a moiety which can be used to purify a strand of nucleic acid, usually via attachment to a solid support as outlined herein. Suitable purification tags include members of binding partner pairs. For example, the tag may be a hapten or antigen, which will bind its binding partner. In a preferred embodiment, the binding partner can be attached to a solid support as depicted herein and in the figures. For example, suitable binding partner pairs include, but are not limited to: antigens (such as proteins (including peptides)) and antibodies (including fragments thereof (FAbs, etc.)); proteins and small molecules, including biotin/streptavidin; enzymes and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partners. Nucleic acid—nucleic acid binding proteins pairs are also useful. In general, the smaller of the pair is attached to the NTP for incorporation into the primer. Preferred binding partner pairs include, but are not limited to, biotin (or imino-biotin) and streptavidin, digeoxinin and Abs, and Prolinx™ reagents (see www.prolinxinc.com/ie4/home.hmtl).

In a preferred embodiment, the binding partner pair comprises biotin or imino-biotin and streptavidin. Imino-biotin is particularly preferred as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95° C.).

Additional techniques include, but are not limited to, enzymatic attachment, chemical attachment, photochemistry or thermal attachment and absorption.

In a preferred embodiment, as outlined herein, enzymatic techniques are used to attach the target nucleic acid, probe or primer to the support. For example, terminal transferase end-labeling techniques can be used as outlined above; see Hermanson, Bioconjugate Techniques, San Diego, Academic Press, pp 640-643. In this embodiment, a nucleotide labeled with a secondary label (e.g. a binding ligand, such as biotin) is added to a terminus of the target nucleic acid; supports coated or containing the binding partner (e.g. streptavidin) can thus be used to immobilize the target nucleic acid. Alternatively, the terminal transferase can be used to add nucleotides with special chemical functionalities that can be specifically coupled to a support. Preferred embodiments utilize the addition of biotinylated nucleotides followed by capture on streptavidin coated magnetic beads. Similarly, random-primed labeling or nick-translation labeling (supra, pp. 640-643) can also be used. In some embodiments the probe or primer are synthesized with biotinylated nucleotides or biotinylated after synthesis by methods as described herein.

In a preferred embodiment, chemical labeling (supra, pp. 6444-671) can be used. In this embodiment, bisulfite-catalyzed transamination, sulfonation of cytosine residues, bromine activation of T, C and G bases, periodate oxidation of RNA or carbodiimide activation of 5' phosphates can be done.

In a preferred embodiment, photochemistry or heat-activated labeling is done (supra, p 162-166). Thus for example, aryl azides and nitrenes preferably label adenosines, and to a less extent C and T (Aslam et al., Bioconjugation: Protein Coupling Techniques for Biomedical Sciences; New York, Grove's Dictionaries, 833 pp.). Psoralen or angelicin compounds can also be used (Aslam, p 492, supra). The preferential modification of guanine can be accomplished via intercalation of platinum complexes (Aslam, supra).

In a preferred embodiment, the target nucleic acid can be absorbed on positively charged surfaces, such as an amine coated solid phase. The target nucleic acid can be cross-linked to the surface after physical absorption for increased retention (e.g. PEI coating and glutaraldehyde cross-linking; Aslam, supra, p. 485).

In a preferred embodiment, direct chemical attached or photocrosslinking can be done to attach the target nucleic acid to the solid phase, by using direct chemical groups on the solid phase substrate. For example, carbodiimide activation of 5' phosphates, attachment to exocyclic amines on DNA bases, and psoralen can be attached to the solid phase for crosslinking to the DNA. Other methods of tagging and immobilizing nucleic acids are described in U.S. Ser. No. 09/931,285, filed Aug. 16, 2001, which is expressly incorporated herein by reference.

Once attached to the first solid support, the target sequence, probe or primers are amenable to analysis as described herein.

In some embodiments when degradation is the preferred method of performing complexity reduction, the ddTNPs or dNTPs that are added during the reaction confer protection from degradation (whether chemical or enzymatic). Thus, after the assay, the degradation components are added, and

unreacted primers are degraded, leaving only the reacted primers. Labeled protecting groups are particularly preferred; for example, 3'-substituted-2'-dNTPs can contain anthranylic derivatives that are fluorescent (with alkali or enzymatic treatment for removal of the protecting group).

In a preferred embodiment, the secondary label is a nuclease inhibitor, such as thiol NTPs. In this embodiment, the chain-terminating NTPs are chosen to render extended primers resistant to nucleases, such as 3'-exonucleases. Addition of an exonuclease will digest the non-extended primers leaving only the extended primers to bind to the capture probes on the array. This may also be done with OLA, wherein the ligated probe will be protected but the unprotected ligation probe will be digested.

In this embodiment, suitable 3'-exonucleases include, but are not limited to, exo I, exo III, exo VII, and 3'-5' exophosphodiesterases. That is, treatment with single stranded nucleases (either endonucleases or exonucleases) will effectively remove excess nucleic acid sequences that are non-complementary tot eh locus specific primer or extension product (see FIG. **11**). Nuclease treatment can be performed either prior to or after separation, i.e. immobilization and washing, of purified nucleic acid targets.

Alternatively, an 3' exonuclease may be added to a mixture of 3' labeled biotin/streptavidin; only the unreacted oligonucleotides will be degraded. Following exonuclease treatment, the exonuclease and the streptavidin can be degraded using a protease such as proteinase K. The surviving nucleic acids (i.e. those that were biotinylated) are then hybridized to the array.

In a preferred embodiment the non-hybridized nucleic acids are removed by washing. In this embodiment the hybridization complexes are immobilized on a solid support and washed under conditions sufficient to remove non-hybridized nucleic acids, i.e. non-hybridized probes and sample nucleic acids. In a particularly preferred embodiment immobilized complexes are washed under conditions sufficient to remove imperfectly hybridized complexes. That is, hybridization complexes that contain mismatches are also removed in the wash steps.

A variety of hybridization or washing conditions may be used in the present invention, including high, moderate and low stringency conditions; see for example Maniatis et al., Molecular Cloning: A Laboratory Manual, 2*d* Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, et al, hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C. lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g. 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of helix destabilizing agents such as formamide. The hybridization or washing conditions may also vary when a non-ionic backbone, i.e. PNA is used, as is known in the art. In addition, cross-linking agents may be added after target binding to cross-link, i.e. covalently attach, the two strands of the hybridization complex.

In one embodiment the hybridization complexes are immobilized by binding of a purification tag to the solid support. That is, a purification tag is incorporated into the hybridization complexes. Purification tags are described herein and can be incorporated into hybridization complexes in a variety of ways. In one embodiment the locus specific probes contain purification tags as described herein. That is, the probe is synthesized with a purification tag, i.e. biotinylated nucleotides, or a purification tag is added to the probe. Thus, upon hybridization with target nucleic acids, immobilization of the hybridization complexes is accomplished by a purification tag. The purification tag associates with the solid support.

Purification tags are described herein. In a preferred embodiment the purification tag is biotin. That is, preferably the first probe is labeled with biotin. The labeled hybridization complex, therefore, binds to streptavidin coated solid support. Solid supports also are described herein. In a preferred embodiment the solid support is streptavidin coated magnetic beads.

The purification tag also can be incorporated into the locus specific primer following a primer extension reaction as described more fully below. Briefly, following hybridization of locus specific primers with target nucleic acids, a polymerase extension reaction is performed. In this embodiment tagged nucleotides, i.e. biotinylated nucleotides, are incorporated into the primer as a result of the extension reaction. That is, once the target sequence and the first probe sequence have hybridized, the method of this embodiment further comprises the addition of a polymerase and at least one nucleotide (dNTP) labeled with a purification tag. Suitable DNA polymerases include, but are not limited to, the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase. In this embodiment, it also is important to anneal under high stringency conditions so that only correctly hybridized probes and target nucleic acids are extended.

In addition, the purification tag can be incorporated into the target nucleic acid. In this embodiment, the target nucleic acid is labeled with a purification tag and immobilized to the solid support as described above. Preferably the tag is biotin.

Once formed, the tagged extension product is immobilized on the solid support as described above. Once immobilized, the complexes are washed so as to remove unhybridized nucleic acids.

Thus, a complexity reduction includes a locus specific selection of target nucleic acids. Non-specific or non-target nucleic acids are removed.

Once unhybridized probes and non-target nucleic acids have been removed, the probes, primers or hybridization complexes are generally subjected to an extension reaction. As outlined herein, the probes, primers or hybridization complexes can be immobilized or in solution after the optional complexity reduction step. Using the hybridized locus specific or allele specific probe as a primer, extension enzyme such as a polymerase and dNTPs are added to the assay mixture for extension of the primer. The resulting extended primer thus includes sequence information of the target nucleic acid, including the sequence of the specific allele to be detected. Thus, the extended primer serves as the template in

subsequent specificity steps to identify the nucleotide at the detection position, i.e. the particular allele to be detected.

By "extension enzyme" herein is meant an enzyme that will extend a sequence by the addition of NTPs. As is well known in the art, there are a wide variety of suitable extension enzymes, of which polymerases (both RNA and DNA, depending on the composition of the target sequence and precircle probe) are preferred. Preferred polymerases are those that lack strand displacement activity, such that they will be capable of adding only the necessary bases at the end of the probe, without further extending the probe to include nucleotides that are complementary to a targeting domain and thus preventing circularization. Suitable polymerases include, but are not limited to, both DNA and RNA polymerases, including the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase, Phi29 DNA polymerase and various RNA polymerases such as from Thermus sp., or Q beta replicase from bacteriophage, also SP6, T3, T4 and T7 RNA polymerases can be used, among others.

Even more preferred polymerases are those that are essentially devoid of a 5' to 3' exonuclease activity, so as to assure that the probe will not be extended past the 5' end of the probe. Exemplary enzymes lacking 5' to 3' exonuclease activity include the Klenow fragment of the DNA Polymerase and the Stoffel fragment of DNAPTaq Polymerase. For example, the Stoffel fragment of Taq DNA polymerase lacks 5' to 3' exonuclease activity due to genetic manipulations, which result in the production of a truncated protein lacking the N-terminal 289 amino acids. (See e.g., Lawyer et al., J. Biol. Chem., 264:6427-6437 [1989]; and Lawyer et al., PCR Meth. Appl., 2:275-287 [1993]). Analogous mutant polymerases have been generated for polymerases derived from T. maritima, Tsps17, TZ05, Tth and Taf.

Even more preferred polymerases are those that lack a 3' to 5' exonuclease activity, which is commonly referred to as a proof-reading activity, and which removes bases which are mismatched at the 3' end of a primer-template duplex. Although the presence of 3' to 5' exonuclease activity provides increased fidelity in the strand synthesized, the 3' to 5' exonuclease activity found in thermostable DNA polymerases such as Tma (including mutant forms of Tma that lack 5' to 3' exonuclease activity) also degrades single-stranded DNA such as the primers used in the PCR, single-stranded templates and single-stranded PCR products. The integrity of the 3' end of an oligonucleotide primer used in a primer extension process is critical as it is from this terminus that extension of the nascent strand begins. Degradation of the 3' end leads to a shortened oligonucleotide which in turn results in a loss of specificity in the priming reaction (i.e., the shorter the primer the more likely it becomes that spurious or non-specific priming will occur).

Yet even more preferred polymerases are thermostable polymerases. For the purposes of this invention, a heat resistant enzyme is defined as any enzyme that retains most of its activity after one hour at 40° C. under optimal conditions. Examples of thermostable polymerase which lack both 5' to 3'exonuclease and 3' to 5' exonuclease include Stoffel fragment of Taq DNA polymerase. This polymerase lacks the 5' to 3' exonuclease activity due to genetic manipulation and no 3' to 5' activity is present as Taq polymerase is naturally lacking in 3' to 5' exonuclease activity. Tth DNA polymerase is derived form Thermus thermophilus, and is available form Epicentre Technologies, Molecular Biology Resource Inc., or Perkin-Elmer Corp. Other useful DNA polymerases which lack 3' exonuclease activity include a Vent[R](exo-), available from New England Biolabs, Inc., (purified from strains of *E. coli* that carry a DNA polymerase gene from the archaebacterium Thermococcus litoralis), and Hot Tub DNA polymerase derived from Thermus flavus and available from Amersham Corporation.

Other preferred enzymes which are thermostable and deprived of 5' to 3' exonuclease activity and of 3' to 5' exonuclease activity include AmpliTaq Gold. Other DNA polymerases, which are at least substantially equivalent may be used like other N-terminally truncated Thermus aquaticus (Taq) DNA polymerase I the polymerase named KlenTaq I and KlenTaq LA are quite suitable for that purpose. Of course, any other polymerase having these characteristics can also be used according to the invention.

The conditions for performing the addition of one or more nucleotides at the 3' end of the probe will depend on the particular enzyme used, and will generally follow the conditions recommended by the manufacturer of the enzymes used.

In addition, it will be appreciated that more than one complexity reduction step can be performed. That is, following a first complexity reduction step, either the remaining target nucleic acid or the extended locus or allele specific primer, when applicable, are subjected to a subsequent complexity reduction step as described above. That is, an additional locus specific or allele specific primer is hybridized to the target nucleic acid, which can be either the original target nucleic acid or the extended primer, and unhybridized target nucleic acids are removed. This can be repeated as many times as necessary to achieve the required level of enrichment of target nucleic acid.

While the above has been described in the context of complexity reduction, it is appreciated that some level of specificity also is included in these steps. That is, as a result of hybridizing target nucleic acids with locus specific probes, specificity also in accomplished. This is particularly apparent when allele specific probes are used initially.

Specificity Component

Generally following at least one complexity reduction step a specificity step is included in the method of the invention. By "specificity component" is meant a step that discriminates between target nucleic acids, preferably at the level of the allele. That is, the specificity component is an allele specific step (e.g. genotyping or SNP analysis). While some level of specificity can be accomplished by simply hybridizing allele specific probes to the template (i.e. the product of the complexity reduction step above), in a preferred embodiment the specificity step includes an enzymatic step. That is, the fidelity of an enzymatic step improves specificity for allele discrimination. Preferred enzymes include DNA polymerases, RNA polymerases and ligases as described in more detail herein.

Polymerases are described above. Many ligases are known and are suitable for use in the invention, e.g. Lehman, Science, 186: 790-797 (1974); Engler et al, DNA Ligases, pages 3-30 in Boyer, editor, The Enzymes, Vol. 15B (Academic Press, New York, 1982); and the like. Preferred ligases include T4 DNA ligase, T7 DNA ligase, *E. coli* DNA ligase, Taq ligase, Pfu ligase, and Tth ligase. Protocols for their use are well known, e.g. Sambrook et al (cited above); Barany, PCR Methods an Applications, 1: 5-16 (1991); Marsh et al, Strategies, 5: 73-76 (1992); and the like. Generally, ligases require that a 5' phosphate group be present for ligation to the 3' hydroxyl of an abutting strand. Preferred ligases include thermostable or (thermophilic) ligases, such as pfu ligase, Tth ligase, Taq ligase and Ampligase TM DNA ligase (Epicentre Technologies, Madison, Wis.). Ampligase has a low blunt end ligation activity.

The preferred ligase is one which has the least mismatch ligation. The specificity of ligase can be increased by substituting the more specific NAD+-dependent ligases such as *E. coli* ligase and (thermostable) Taq ligase for the less specific T4 DNA ligase. The use of NAD analogues in the ligation reaction further increases specificity of the ligation reaction. See, U.S. Pat. No. 5,508,179 to Wallace et al.

In one embodiment the specificity component is performed with immobilized targets. That is, the products of the complexity reduction step are immobilized on a solid support as outlined herein and described in U.S. Ser. No. 09/931,285, filed Aug. 16, 2001, which is expressly incorporated herein by reference. As discussed herein the target of specificity reaction is referred to as a "specificity target". That is, the product of the complexity reduction step is the specificity target.

In one embodiment the support is the same support as in the initial complexity reduction step. In this embodiment the target nucleic acid is removed from the solid support prior to the specificity assay. The target nucleic acid can be removed by any method that denatures the hybridization complex resulting in release of the target nucleic acid. As one of skill in the art appreciates, in this embodiment the target nucleic acid is not covalently bound to the solid support. That is, it is the target probe that is stably attached to the support. That is, while the attachment of the probe is not necessarily covalent, it is stable enough to withstand denaturation of the hybridization complex and removal of the nonattached target nucleic acid.

In an alternative embodiment the specificity target is in solution. That is, following a complexity reduction step, the hybridization complex between the immobilized target nucleic acid and target probe, which has generally been modified (see above), is denatured and the modified target probe is eluted from the hybridization complex. In a preferred embodiment the specificity target is analyzed in solution. In an alternative embodiment the solution phase specificity target is immobilized on a subsequent solid support.

These specificity assays, i.e. genotyping techniques, fall into five general categories: (1) techniques that rely on traditional hybridization methods that utilize the variation of stringency conditions (temperature, buffer conditions, etc.) to distinguish nucleotides at the detection position; (2) extension techniques that add a base ("the base") to basepair with the nucleotide at the detection position; (3) ligation techniques, that rely on the specificity of ligase enzymes (or, in some cases, on the specificity of chemical techniques), such that ligation reactions occur preferentially if perfect complementarity exists at the detection position; (4) cleavage techniques, that also rely on enzymatic or chemical specificity such that cleavage occurs preferentially if perfect complementarity exists; and (5) techniques that combine these methods. See generally WO 00/63437, incorporated by reference in its entirety.

a) Competitive Hybridization

In a preferred embodiment, the use of competitive hybridization is performed to elucidate either the identity of the nucleotide(s) at the detection position or the presence of a mismatch. For example, sequencing by hybridization has been described (Drmanac et al., Genomics 4:114 (1989); Koster et al., Nature Biotechnology 14:1123 (1996); U.S. Pat. Nos. 5,525,464; 5,202,231 and 5,695,940, among others, all of which are hereby expressly incorporated by reference in their entirety).

It should be noted in this context that "mismatch" is a relative term and meant to indicate a difference in the identity of a base at a particular position, termed the "detection position" herein, between two sequences. In general, sequences that differ from wild type sequences are referred to as mismatches. However, particularly in the case of SNPs, what constitutes "wild type" may be difficult to determine as multiple alleles can be relatively frequently observed in the population, and thus "mismatch" in this context requires the artificial adoption of one sequence as a standard. Thus, for the purposes of this invention, sequences are referred to herein as "match" and "mismatch". Thus, the present invention may be used to detect substitutions, insertions or deletions as compared to a wild-type sequence.

In a preferred embodiment, a plurality of probes (sometimes referred to herein as "readout probes") are used to identify the base at the detection position. In this embodiment, each different readout probe comprises a different detection label (which, as outlined below, can be either a primary label or a secondary label) and a different base at the position that will hybridize to the detection position of the target sequence (herein referred to as the readout position) such that differential hybridization will occur. That is, all other parameters being equal, a perfectly complementary readout probe (a "match probe") will in general be more stable and have a slower off rate than a probe comprising a mismatch (a "mismatch probe") at any particular temperature. Accordingly, by using different readout probes, each with a different base at the readout position and each with a different label, the identification of the base at the detection position is elucidated.

Accordingly, in some embodiments a detectable label is incorporated into the readout probe. In a preferred embodiment, a set of readout probes are used, each comprising a different base at the readout position. In some embodiments, each readout probe comprises a different label, that is distinguishable from the others. For example, a first label may be used for probes comprising adenosine at the readout position, a second label may be used for probes comprising guanine at the readout position, etc. In a preferred embodiment, the length and sequence of each readout probe is identical except for the readout position, although this need not be true in all embodiments.

The number of readout probes used will vary depending on the end use of the assay. For example, many SNPs are biallelic, and thus two readout probes, each comprising an interrogation base that will basepair with one of the detection position bases. For sequencing, for example, for the discovery of SNPs, a set of four readout probes are used, although SNPs may also be discovered with fewer readout parameters.

As will be appreciated by those in the art and additionally outlined below, this system can take on a number of different configurations, including a solution phase assay and a solid phase assay.

Solution Phase Assay

In some embodiments a solution phase assay is performed followed by attaching the target sequence to a solid support such as an array. After the competitive hybridization has occurred, the target sequence is added to the support, which may take on several configurations, outlined below.

Solid Phase Assay

In a preferred embodiment, the competition reaction is done on a solid support, such as an array. This system may take on several configurations.

In a preferred embodiment, a sandwich assay of sorts is used. In this embodiment, the bead, when bead arrays are used, comprises a capture probe that will hybridize to a first target domain of a target sequence, and the readout probe will hybridize to a second target domain. In this embodiment, the first target domain may be either unique to the target, or may be an exogeneous adapter sequence added to the target

sequence as outlined below, for example through the use of PCR reactions. Similarly, a sandwich assay is performed that utilizes a capture extender probe, as described below, to attach the target sequence to the array.

Alternatively, the capture probe itself can be the readout probe; that is, a plurality of microspheres are used, each comprising a capture probe that has a different base at the readout position. In general, the target sequence then hybridizes preferentially to the capture probe most closely matched. In this embodiment, either the target sequence itself is labeled (for example, it may be the product of an amplification reaction) or a label probe may bind to the target sequence at a domain remote from the detection position. In this embodiment, since it is the location on the array that serves to identify the base at the detection position, different labels are not required.

Stringency Variation

In a preferred embodiment, sensitivity to variations in stringency parameters are used to determine either the identity of the nucleotide(s) at the detection position or the presence of a mismatch. As a preliminary matter, the use of different stringency conditions such as variations in temperature and buffer composition to determine the presence or absence of mismatches in double stranded hybrids comprising a single stranded target sequence and a probe is well known.

With particular regard to temperature, as is known in the art, differences in the number of hydrogen bonds as a function of basepairing between perfect matches and mismatches can be exploited as a result of their different Tms (the temperature at which 50% of the hybrid is denatured). Accordingly, a hybrid comprising perfect complementarity will melt at a higher temperature than one comprising at least one mismatch, all other parameters being equal. (It should be noted that for the purposes of the discussion herein, all other parameters (i.e. length of the hybrid, nature of the backbone (i.e. naturally occurring or nucleic acid analog), the assay solution composition and the composition of the bases, including G-C content are kept constant). However, as will be appreciated by those in the art, these factors may be varied as well, and then taken into account.)

In general, as outlined herein, high stringency conditions are those that result in perfect matches remaining in hybridization complexes, while imperfect matches melt off. Similarly, low stringency conditions are those that allow the formation of hybridization complexes with both perfect and imperfect matches. High stringency conditions are known in the art; see for example Maniatis et al., Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, et al., both of which are hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g. 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In another embodiment, less stringent hybridization conditions are used; for example, moderate or low stringency conditions may be used, as are known in the art; see Maniatis and Ausubel, supra, and Tijssen, supra.

As will be appreciated by those in the art, mismatch detection using temperature may proceed in a variety of ways, and is similar to the use of readout probes as outlined above. Again, as outlined above, a plurality of readout probes may be used in a sandwich format; in this embodiment, all the probes may bind at permissive, low temperatures (temperatures below the Tm of the mismatch); however, repeating the assay at a higher temperature (above the Tm of the mismatch) only the perfectly matched probe may bind. Thus, this system may be run with readout probes with different detectable labels, as outlined above. Alternatively, a single probe may be used to query whether a particular base is present.

Alternatively, as described above, the capture probe may serve as the readout probe; in this embodiment, a single label may be used on the target; at temperatures above the Tm of the mismatch, only signals from perfect matches will be seen, as the mismatch target will melt off.

Similarly, variations in buffer composition may be used to elucidate the presence or absence of a mismatch at the detection position. Suitable conditions include, but are not limited to, formamide concentration. Thus, for example, "low" or "permissive" stringency conditions include formamide concentrations of 0 to 10%, while "high" or "stringent" conditions utilize formamide concentrations of ≧40%. Low stringency conditions include NaCl concentrations of ≧1 M, and high stringency conditions include concentrations of ≦0.3 M. Furthermore, low stringency conditions include $MgCl_2$ concentrations of ≧10 mM, moderate stringency as 1-10 mM, and high stringency conditions include concentrations of ≦1 mM.

In this embodiment, as for temperature, a plurality of readout probes may be used, with different bases in the readout position (and optionally different labels). Running the assays under the permissive conditions and repeating under stringent conditions will allow the elucidation of the base at the detection position.

In one embodiment, the probes used as readout probes are "Molecular Beacon" probes as are generally described in Whitcombe et al., Nature Biotechnology 17:804 (1999), hereby incorporated by reference. As is known in the art, Molecular Beacon probes form "hairpin" type structures, with a fluorescent label on one end and a quencher on the other. In the absence of the target sequence, the ends of the hairpin hybridize, causing quenching of the label. In the presence of a target sequence, the hairpin structure is lost in favor of target sequence binding, resulting in a loss of quenching and thus an increase in signal.

In one embodiment, the Molecular Beacon probes can be the capture probes as outlined herein for readout probes. For example, different beads comprising labeled Molecular Beacon probes (and different bases at the readout position) are made optionally they comprise different labels. Alternatively, since Molecular Beacon probes can have spectrally resolvable signals, all four probes (if a set of four different bases with is used) differently labeled are attached to a single bead.

b) Extension Assays

In this embodiment the specificity target is immobilized on a solid support. In a preferred embodiment, extension genotyping is done. In this embodiment, any number of techniques are used to add a nucleotide to the readout position of a probe hybridized to the target sequence adjacent to the detection position. By relying on enzymatic specificity, preferentially a perfectly complementary base is added. All of these methods rely on the enzymatic incorporation of nucleotides at the detection position. This may be done using chain terminating dNTPs, such that only a single base is incorporated (e.g. single base extension methods), or under conditions that only a single type of nucleotide is added followed by identification of the added nucleotide (extension and pyrosequencing techniques).

Single Base Extension

In a preferred embodiment, single base extension (SBE; sometimes referred to as "minisequencing") is used to determine the identity of the base at the detection position. SBE utilizes an extension primer that may have at least one adapter sequence that hybridizes to the target nucleic acid immediately adjacent to the detection position, to form a hybridization complex. A polymerase (generally a DNA polymerase) is used to extend the 3' end of the primer with a nucleotide or nucleotide analog. In some embodiments the nucleotide or nucleotide analog is labeled with a detection label as described herein. Based on the fidelity of the enzyme, a nucleotide is only incorporated into the readout position of the growing nucleic acid strand if it is perfectly complementary to the base in the target strand at the detection position. The nucleotide may be derivatized such that no further extensions can occur, so only a single nucleotide is added. Once the labeled nucleotide is added, detection of the label proceeds as outlined herein. Again, amplification in this case is accomplished through cycling or repeated rounds of reaction/elution, although in some embodiments amplification is not necessary. Alternatively, in some embodiments, amplification is performed prior to the extension reaction. Alternatively, amplification is performed following the extension reaction.

The reaction is initiated by introducing the hybridization complex comprising the specificity target on the support to a solution comprising a first nucleotide. In some embodiments, the nucleotides comprise a detectable label, which may be either a primary or a secondary label. In addition, the nucleotides may be nucleotide analogs, depending on the configuration of the system. For example, if the dNTPs are added in sequential reactions, such that only a single type of dNTP can be added, the nucleotides need not be chain terminating. In addition, in this embodiment, the dNTPs may all comprise the same type of label.

Alternatively, if the reaction comprises more than one dNTP, the dNTPs should be chain terminating, that is, they have a blocking or protecting group at the 3' position such that no further dNTPs may be added by the enzyme. As will be appreciated by those in the art, any number of nucleotide analogs may be used, as long as a polymerase enzyme will still incorporate the nucleotide at the readout position. Preferred embodiments utilize dideoxy-triphosphate nucleotides (ddNTPs) and halogenated dNTPs. Generally, a set of nucleotides comprising ddATP, ddCTP, ddGTP and ddTTP is used, each with a different detectable label, although as outlined herein, this may not be required. Alternative preferred embodiments use acyclo nucleotides (NEN). These chain terminating nucleotide analogs are particularly good substrates for Deep vent (exo$^-$) and thermosequenase.

In addition, as will be appreciated by those in the art, the single base extension reactions of the present invention allow the precise incorporation of modified bases into a growing nucleic acid strand. Thus, any number of modified nucleotides may be incorporated for any number of reasons, including probing structure-function relationships (e.g. DNA:DNA or DNA:protein interactions), cleaving the nucleic acid, crosslinking the nucleic acid, incorporate mismatches, etc.

As will be appreciated by those in the art, the configuration of the genotyping SBE system can take on several forms.

Multi-Base Extension

In a preferred embodiment genotyping is accomplished by primer extension that does not use chain terminating nucleotides. As such, this genotyping is considered multi-base extension. The method includes providing an interrogator oligonucleotide designed to detect one allele of a given SNP. The number of oligonucleotides is determined by the number of distinct SNP alleles being probed. For instance, if one were probing 1000 SNPs, each with two alleles, 2000 oligonucleotides would be necessary. The interrogators are complementary to a stretch of DNA containing the SNP, with the terminal base of each interrogator corresponding to the SNP position, or with the SNP-specific position is within the last 1, 2 3 or 4 nucleotides of the interrogator. For example, when a SNP has an A and C allele, interrogators ending in T and G are provided and in some embodiments may be immobilized on separate elements (beads) to detect the two. Although both the match and the mismatch will hybridize to a given allele, only the match can act as a primer for a DNA polymerase extension reaction. Accordingly, following hybridization of the probes with the target DNA, a polymerase reaction is performed. This results in the extension of the hybrids with a DNA polymerase in the presence of labeled dNTPs. The labeled dNTPs are selectively incorporated into the extension product that results from the probe that is complementary to the SNP position.

In one embodiment, address oligonucleotides (adapters) are incorporated into the interrogator oligonucleotides. As such, in one embodiment one performs the hybridization and extension steps in fluid phase in the absence of beads. Each allele contains a unique adapter. After hybridization/extension the products are hybridized to an array of complementary address sequences for signal detection and analysis.

Solution Phase Assay

As for the OLA reaction described below, the reaction may be done in solution, and then the newly synthesized strands, with the base-specific detectable labels, can be detected. For example, they can be directly hybridized to capture probes that are complementary to the extension primers, and the presence of the label is then detected. As will be appreciated by those in the art, a preferred embodiment utilizes four different detectable labels, i.e. one for each base, such that upon hybridization to the capture probe on the array, the identification of the base can be done isothermally.

In a preferred embodiment, adapter sequences can be used in a solution format. In this embodiment, a single label can be used with a set of four separate primer extension reactions. In this embodiment, the extension reaction is done in solution; each reaction comprises a different dNTP with the label or labeled ddNTP when chain termination is desired. For each locus genotyped, a set of four different extension primers are used, each with a portion that will hybridize to the target sequence, a different readout base and each with a different adapter sequence of 15-40 bases, as is more fully outlined below. After the primer extension reaction is complete, the four separate reactions are pooled and hybridized to an array comprising complementary probes to the adapter sequences.

A genotype is derived by comparing the probe intensities of the four different hybridized adapter sequences corresponding to a given locus.

In addition, since unextended primers do not comprise labels, the unextended primers need not be removed. However, they may be, if desired, as outlined below; for example, if a large excess of primers are used, there may not be sufficient signal from the extended primers competing for binding to the surface.

Alternatively, one of skill in the art could use a single label and temperature to determine the identity of the base; that is, the readout position of the extension primer hybridizes to a position on the capture probe. However, since the three mismatches will have lower Tms than the perfect match, the use of temperature could elucidate the identity of the detection position base.

Solid Phase Assay

Alternatively, the reaction may be done on a surface by capturing the target sequence and then running the SBE reaction, in a sandwich type format. In this embodiment, the capture probe hybridizes to a first domain of the target sequence (which can be endogeneous or an exogeneous adapter sequence added during an amplification reaction), and the extension primer hybridizes to a second target domain immediately adjacent to the detection position. The addition of the enzyme and the required NTPs results in the addition of the interrogation base. In this embodiment, each NTP must have a unique label. Alternatively, each NTP reaction may be done sequentially on a different array. As is known by one of skill in the art, ddNTP and dNTP are the preferred substrates when DNA polymerase is the added enzyme; NTP is the preferred substrate when RNA polymerase is the added enzyme.

Furthermore, capture extender probes can be used to attach the target sequence to the bead. In this embodiment, the hybridization complex comprises the capture probe, the target sequence and the adapter sequence.

Similarly, the capture probe itself can be used as the extension probe, with its terminus being directly adjacent to the detection position. Upon the addition of the target sequence and the SBE reagents, the modified primer is formed comprising a detectable label, and then detected. Again, as for the solution based reaction, each NTP must have a unique label, the reactions must proceed sequentially, or different arrays must be used. Again, as is known by one of skill in the art, ddNTP and dNTP are the preferred substrates when DNA polymerase is the added enzyme; NTP is the preferred substrate when RNA polymerase is the added enzyme.

In addition, as outlined herein, the target sequence may be directly attached to the array; the extension primer hybridizes to it and the reaction proceeds.

Variations on this include, where the capture probe and the extension probe adjacently hybridize to the target sequence. Either before or after extension of the extension probe, a ligation step may be used to attach the capture and extension probes together for stability. These are further described below as combination assays.

As will be appreciated by those in the art, the determination of the base at the detection position can proceed in several ways. In a preferred embodiment, the reaction is run with all four nucleotides (assuming all four nucleotides are required), each with a different label, as is generally outlined herein. Alternatively, a single label is used, by using four reactions. In a preferred embodiment, universal primers or adapters specific for the nucleotide at a detection position are used and detected as outlined below.

Removal of Unextended Primers

In a preferred embodiment, for both SBE as well as a number of other reactions outlined herein, it is desirable to remove the unextended or unreacted primers from the assay mixture, and particularly from the array, as unextended primers will compete with the extended (labeled) primers in binding to capture probes, thereby diminishing the signal. The concentration of the unextended primers relative to the extended primer may be relatively high, since a large excess of primer is usually required to generate efficient primer annealing. Accordingly, a number of different techniques may be used to facilitate the removal of unextended primers. As outlined above, these generally include methods based on removal of unreacted primers by binding to a solid support, protecting the reacted primers and degrading the unextended ones, and separating the unreacted and reacted primers.

Separation Systems

The use of secondary label systems (and even some primary label systems) can be used to separate unreacted and reacted probes; for example, the addition of streptavidin to a nucleic acid greatly increases its size, as well as changes its physical properties, to allow more efficient separation techniques. For example, the mixtures can be size fractionated by exclusion chromatography, affinity chromatography, filtration or differential precipitation.

Non-Terminated Extension

In a preferred embodiment, methods of adding a single base are used that do not rely on chain termination. That is, similar to SBE, enzymatic reactions that utilize dNTPs and polymerases can be used; however, rather than use chain terminating dNTPs, regular dNTPs are used. This method relies on a time-resolved basis of detection; only one type of base is added during the reaction.

Pyrosequencing

Pyrosequencing is an extension and sequencing method that can be used to add one or more nucleotides to the detection position(s); it is very similar to SBE except that chain terminating NTPs need not be used (although they may be). Pyrosequencing relies on the detection of a reaction product, PPi, produced during the addition of an NTP to a growing oligonucleotide chain, rather than on a label attached to the nucleotide. One molecule of PPi is produced per dNTP added to the extension primer. That is, by running sequential reactions with each of the nucleotides, and monitoring the reaction products, the identity of the added base is determined.

The release of pyrophosphate (PPi) during the DNA polymerase reaction can be quantitatively measured by many different methods and a number of enzymatic methods have been described; see Reeves et al., Anal. Biochem. 28:282 (1969); Guillory et al., Anal. Biochem. 39:170 (1971); Johnson et al., Anal. Biochem. 15:273 (1968); Cook et al., Anal. Biochem. 91:557 (1978); Drake et al., Anal. Biochem. 94:117 (1979); WO93/23564; WO 98/28440; WO98/13523; Nyren et al., Anal. Biochem. 151:504 (1985); all of which are incorporated by reference. The latter method allows continuous monitoring of PPi and has been termed ELIDA (Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay). A preferred embodiment utilizes any method which can result in the generation of an optical signal, with preferred embodiments utilizing the generation of a chemiluminescent or fluorescent signal.

A preferred method monitors the creation of PPi by the conversion of PPi to ATP by the enzyme sulfurylase, and the subsequent production of visible light by firefly luciferase (see Ronaghi et al., Science 281:363 (1998), incorporated by reference). In this method, the four deoxynucleotides (dATP, dGTP, dCTP and dTTP; collectively dNTPs) are added step-

wise to a partial duplex comprising a sequencing primer hybridized to a single stranded DNA template and incubated with DNA polymerase, ATP sulfurylase, luciferase, and optionally a nucleotide-degrading enzyme such as apyrase. A dNTP is only incorporated into the growing DNA strand if it is complementary to the base in the template strand. The synthesis of DNA is accompanied by the release of PPi equal in molarity to the incorporated dNTP. The PPi is converted to ATP and the light generated by the luciferase is directly proportional to the amount of ATP. In some cases the unincorporated dNTPs and the produced ATP are degraded between each cycle by the nucleotide degrading enzyme.

Accordingly, a preferred embodiment of the methods of the invention is as follows. A substrate comprising the target sequences and extension primers, forming hybridization complexes, is dipped or contacted with a reaction volume (chamber or well) comprising a single type of dNTP, an extension enzyme, and the reagents and enzymes necessary to detect PPi. If the dNTP is complementary to the base of the target portion of the target sequence adjacent to the extension primer, the dNTP is added, releasing PPi and generating detectable light, which is detected as generally described in U.S. Ser. Nos. 09/151,877 and 09/189,543, and PCT US98/09163, all of which are hereby incorporated by reference. If the dNTP is not complementary, no detectable signal results. The substrate is then contacted with a second reaction volume (chamber) comprising a different dNTP and the additional components of the assay. This process is repeated if the identity of a base at a second detection position is desirable.

In a preferred embodiment, washing steps may be done in between the dNTP reactions, as required. These washing steps may optionally comprise a nucleotide-degrading enzyme, to remove any unreacted dNTP and decreasing the background signal, as is described in WO 98/28440, incorporated herein by reference.

As will be appreciated by those in the art, the system can be configured in a variety of ways, including both a linear progression or a circular one; for example, four arrays may be used that each can dip into one of four reaction chambers arrayed in a circular pattern. Each cycle of sequencing and reading is followed by a 90 degree rotation, so that each substrate then dips into the next reaction well.

As for simple extension and SBE, the pyrosequencing systems may be configured in a variety of ways; for example, the target sequence may be immobilized in a variety of ways, including direct attachment of the target sequence; the use of a capture probe with a separate extension probe; the use of a capture extender probe, a capture probe and a separate extension probe; the use of adapter sequences in the target sequence with capture and extension probes; and the use of a capture probe that also serves as the extension probe.

One additional benefit of pyrosequencing for genotyping purposes is that since the reaction does not rely on the incorporation of labels into a growing chain, the unreacted extension primers need not be removed.

In addition, pyrosequencing can be used as a "switch" to activate a detectable enzymatic reaction, thus providing an amplification of sorts. The by-product of the polymerase reaction, PPi, is converted to ATP during pyrosequencing reactions. In standard pyrosequencing that utilizes a luciferase/luciferin assay, the detection sensitivity is limited because only a single photon is generated per nucleotide incorporation event. However, in a preferred embodiment, if PPi, or a simple enzymatic derivative such as Pi or ATP is used to "activate" an enzyme or protein, the detection sensitivity is increased. A number of different proteins are either "on" or "off" depending on their phosphorylation status. In this was, PPi (or ATP) acts a "switch" to turn on or off a stream of detection molecules, similar to the way a transistor controls a large flow of electricity by using a small current or potential to gat the process. That is, the generation of PPi results in an enzymatic cascade that results in a detectable event; the PPi generation results in a "switch". For example, ATP may be used to phosphorylate a peroxidase enzyme, which when phosphorylated becomes "active" likes horse radish peroxidase (HRP). This HRP activity is then detected using standard hydrogen peroxide/luminol HRP detection systems. There are a large number of enzymes and proteins regulated by phosphorylation. What is important is that the activating or switch enzyme that utilizes Pi, PPi or ATP as the substrate discriminates the activating species from the original dNTP used in the extension reaction.

Allelic PCR

In a preferred embodiment, the method used to detect the base at the detection position is allelic PCR, referred to herein as "aPCR". As described in Newton et al., Nucl. Acid Res. 17:2503 (1989), hereby expressly incorporated by reference, allelic PCR allows single base discrimination based on the fact that the PCR reaction does not proceed well if the terminal 3'-nucleotide is mismatched, assuming the DNA polymerase being used lacks a 3'-exonuclease proofreading activity. Accordingly, the identification of the base proceeds by using allelic PCR primers (sometimes referred to herein as aPCR primers) that have readout positions at their 3' ends. Thus the target sequence comprises a first domain comprising at its 5' end a detection position.

In general, aPCR may be briefly described as follows. A double stranded target nucleic acid is denatured, generally by raising the temperature, and then cooled in the presence of an excess of a aPCR primer, which then hybridizes to the first target strand. If the readout position of the aPCR primer basepairs correctly with the detection position of the target sequence, a DNA polymerase (again, that lacks 3'-exonuclease activity) then acts to extend the primer with dNTPs, resulting in the synthesis of a new strand forming a hybridization complex. The sample is then heated again, to disassociate the hybridization complex, and the process is repeated. By using a second PCR primer for the complementary target strand, rapid and exponential amplification occurs. Thus aPCR steps are denaturation, annealing and extension. The particulars of aPCR are well known, and include the use of a thermostable polymerase such as Taq I polymerase and thermal cycling.

Accordingly, the aPCR reaction requires at least one aPCR primer, a polymerase, and a set of dNTPs. As outlined herein, the primers may comprise the label, or one or more of the dNTPs may comprise a label.

Furthermore, the aPCR reaction may be run as a competition assay of sorts. For example, for biallelic SNPs, a first aPCR primer comprising a first base at the readout position and a first label, and a second aPCR primer comprising a different base at the readout position and a second label, may be used. The PCR primer for the other strand is the same. The examination of the ratio of the two colors can serve to identify the base at the detection position.

Allelic Primer Extension

In this embodiment allele specific primers when hybridized with their complementary target sequence serve as template for primer extension with a DNA polymerase. In some respects the method is similar to aPCR as described herein with the exception that only one primer need hybridize with the target sequence prior to amplification. That is, in contrast

with PCR amplification that requires two primers, only one primer is necessary for amplification according to the method.

In a preferred embodiment, the primer is immobilized. In a preferred embodiment the primer is immobilized to microspheres or beads as described herein.

In general, as is more fully outlined below, the capture probes on the beads of the array are designed to be substantially complementary to the extended part of the primer; that is, unextended primers will not bind to the capture probes.

Ligation Techniques for Genotyping

In this embodiment, the readout of the base at the detection position proceeds using a ligase. In this embodiment, it is the specificity of the ligase which is the basis of the genotyping; that is, ligases generally require that the 5' and 3' ends of the ligation probes have perfect complementarity to the target for ligation to occur. Thus, in a preferred embodiment, the identity of the base at the detection position proceeds utilizing OLA as described above. The method can be run at least two different ways; in a first embodiment, only one strand of a target sequence is used as a template for ligation; alternatively, both strands may be used; the latter is generally referred to as Ligation Chain Reaction or LCR.

This method is based on the fact that two probes can be preferentially ligated together, if they are hybridized to a target strand and if perfect complementarity exists at the two bases being ligated together. Thus, in this embodiment, the target sequence comprises a contiguous first target domain comprising the detection position and a second target domain adjacent to the detection position. That is, the detection position is "between" the rest of the first target domain and the second target domain, or the detection position is one nucleotide from the 3' terminus of one of the ligation probes. A first ligation probe is hybridized to the first target domain and a second ligation probe is hybridized to the second target domain. If the first ligation probe has a base perfectly complementary to the detection position base, and the adjacent base on the second probe has perfect complementarity to its position, a ligation structure is formed such that the two probes can be ligated together to form a ligated probe. If this complementarity does not exist, no ligation structure is formed and the probes are not ligated together to an appreciable degree. This may be done using heat cycling, to allow the ligated probe to be denatured off the target sequence such that it may serve as a template for further reactions. In addition, as is more fully outlined below, this method may also be done using ligation probes that are separated by one or more nucleotides, if dNTPs and a polymerase are added (this is sometimes referred to as "Genetic Bit" analysis).

In a preferred embodiment, LCR is done for two strands of a double-stranded target sequence. The target sequence is denatured, and two sets of probes are added: one set as outlined above for one strand of the target, and a separate set (i.e. third and fourth primer probe nucleic acids) for the other strand of the target. In a preferred embodiment, the first and third probes will hybridize, and the second and fourth probes will hybridize, such that amplification can occur. That is, when the first and second probes have been attached, the ligated probe can now be used as a template, in addition to the second target sequence, for the attachment of the third and fourth probes. Similarly, the ligated third and fourth probes will serve as a template for the attachment of the first and second probes, in addition to the first target strand. In this way, an exponential, rather than just a linear, amplification can occur.

As will be appreciated by those in the art, the ligation product can be detected in a variety of ways. Preferably, detection is accomplished by removing the unligated labeled probe from the reaction before application to a capture probe. In one embodiment, the unligated probes are removed by digesting 3' non-protected oligonucleotides with a 3' exonuclease, such as, exonuclease I. The ligation products are protected from exo I digestion by including, for example, the use of a number of sequential phosphorothioate residues at their 3' terminus (for example at least four), thereby, rendering them resistant to exonuclease digestion. The unligated detection oligonucleotides are not protected and are digested.

As for most or all of the methods described herein, the assay can take on a solution-based form or a solid-phase form.

Solution Based OLA

In a preferred embodiment, the ligation reaction is run in solution. In this embodiment, only one of the primers carries a detectable label, e.g. the first ligation probe, and the capture probe on the bead is substantially complementary to the other probe, e.g. the second ligation probe. In this way, unextended labeled ligation primers will not interfere with the assay. This substantially reduces or eliminates false signal generated by the optically-labeled 3' primers.

In addition, a solution-based OLA assay that utilizes adapter sequences may be done. In this embodiment, rather than have the target sequence comprise the adapter sequences, one of the ligation probes comprises the adapter sequence. This facilitates the creation of "universal arrays". For example, the first ligation probe has an adapter sequence that is used to attach the ligated probe to the array.

Again, as outlined above for SBE, unreacted ligation primers may be removed from the mixture as needed. For example, the first ligation probe may comprise the label (either a primary or secondary label) and the second may be blocked at its 3' end with an exonuclease blocking moiety; after ligation and the introduction of the nuclease, the labeled ligation probe will be digested, leaving the ligation product and the second probe; however, since the second probe is unlabeled, it is effectively silent in the assay. Similarly, the second probe may comprise a binding partner used to pull out the ligated probes, leaving unligated labeled ligation probes behind. The binding pair is then disassociated for subsequent amplification or detection.

Solid Phase Based OLA

Alternatively, the target nucleic acid is immobilized on a solid-phase surface. The OLA assay is performed and unligated oligonucleotides are removed by washing under appropriate stringency to remove unligated oligonucleotides and thus the label. For example, the capture probe can comprise one of the ligation probes.

Again, as outlined above, the detection of the OLA reaction can also occur directly, in the case where one or both of the primers comprises at least one detectable label, or indirectly, using sandwich assays, through the use of additional probes; that is, the ligated probes can serve as target sequences, and detection may utilize amplification probes, capture probes, capture extender probes, label probes, and label extender probes, etc. Alternatively, the OLA product is amplified. In a preferred embodiment the amplicons comprise labels.

In some embodiments target nucleic acids include both DNA and RNA. In a preferred embodiment RNA is mRNA. In some embodiments when RNA is the target nucleic acid, it is desirable to perform a reverse transcription assay prior to OLA as described herein. The reverse transcription assay results in the formation of cDNA. This method is particularly advantageous in determining either gene expression levels or genotyping, or both. That is, the cDNA is representative of the level of mRNA. Accordingly, gene expression analysis is performed. In addition, the cDNA also serves as a template

for OLA which allows for genotyping. Thus, the use of both DNA and/or RNA allows for increased multiplexing of samples on an array.

Solid Phase Oligonucleotide Ligation Assay (SPOLA)

In a preferred embodiment, a novel method of OLA is used, termed herein "solid phase oligonucleotide assay", or "SPOLA". In this embodiment, the ligation probes are both attached to the same site on the surface of the array (e.g. when microsphere arrays are used, to the same bead), one at its 5' end (the "upstream probe") and one at its 3' end (the "downstream probe"). This may be done as is will be appreciated by those in the art. At least one of the probes is attached via a cleavable linker, that upon cleavage, forms a reactive or detectable (fluorophore) moiety. If ligation occurs, the reactive moiety remains associated with the surface; but if no ligation occurs, due to a mismatch, the reactive moiety is free in solution to diffuse away from the surface of the array. The reactive moiety is then used to add a detectable label.

Generally, as will be appreciated by those in the art, cleavage of the cleavable linker should result in asymmetrical products; i.e. one of the "ends" should be reactive, and the other should not, with the configuration of the system such that the reactive moiety remains associated with the surface if ligation occurred. Thus, for example, amino acids or succinate esters can be cleaved either enzymatically (via peptidases (aminopeptidase and carboxypeptidase) or proteases) or chemically (acid/base hydrolysis) to produce an amine and a carboxyl group. One of these groups can then be used to add a detectable label, as will be appreciated by those in the art and discussed herein.

Padlock Probe Ligation

In a preferred embodiment, the ligation probes are specialized probes called "padlock probes". Nilsson et al, 1994, Science 265:2085, hereby incorporated by reference. These probes have a first ligation domain that is identical to a first ligation probe, in that it hybridizes to a first target sequence domain, and a second ligation domain, identical to the second ligation probe, that hybridizes to an adjacent target sequence domain. Again, as for OLA, the detection position can be either at the 3' end of the first ligation domain or at the 5' end of the second ligation domain. However, the two ligation domains are connected by a linker, frequently nucleic acid. The configuration of the system is such that upon ligation of the first and second ligation domains of the padlock probe, the probe forms a circular probe, and forms a complex with the target sequence wherein the target sequence is "inserted" into the loop of the circle.

In this embodiment, the unligated probes may be removed through degradation (for example, through a nuclease), as there are no "free ends" in the ligated probe.

Cleavage Techniques for Genotyping

In a preferred embodiment, the specificity for genotyping is provided by a cleavage enzyme. There are a variety of enzymes known to cleave at specific sites, either based on sequence specificity, such as restriction endonucleases, or using structural specificity, such as is done through the use of invasive cleavage technology.

Endonuclease Techniques

In a preferred embodiment, enzymes that rely on sequence specificity are used. In general, these systems rely on the cleavage of double stranded sequence containing a specific sequence recognized by a nuclease, preferably an endonuclease including resolvases.

These systems may work in a variety of ways. In one embodiment, a labeled readout probe (generally attached to a bead of the array) is used; the binding of the target sequence forms a double stranded sequence that a restriction endonuclease can then recognize and cleave, if the correct sequence is present. The cleavage results in the loss of the label, and thus a loss of signal.

Alternatively, as will be appreciated by those in the art, a labelled target sequence may be used as well; for example, a labelled primer may be used in the PCR amplification of the target, such that the label is incorporated in such a manner as to be cleaved off by the enzyme.

Alternatively, the readout probe (or, again, the target sequence) may comprise both a fluorescent label and a quencher, as is known in the art. In this embodiment, the label and the quencher are attached to different nucleosides, yet are close enough that the quencher molecule results in little or no signal being present. Upon the introduction of the enzyme, the quencher is cleaved off, leaving the label, and allowing signaling by the label.

In addition, as will be appreciated by those in the art, these systems can be both solution-based assays or solid-phase assays, as outlined herein.

Furthermore, there are some systems that do not require cleavage for detection; for example, some nucleic acid binding proteins will bind to specific sequences and can thus serve as a secondary label. For example, some transcription factors will bind in a highly sequence dependent manner, and can distinguish between two SNPs. Having bound to the hybridization complex, a detectable binding partner can be added for detection. In addition, mismatch binding proteins based on mutated transcription factors can be used.

In addition, as will be appreciated by those in the art, this type of approach works with other cleavage methods as well, for example the use of invasive cleavage methods, as outlined below.

Invasive Cleavage

In a preferred embodiment, the determination of the identity of the base at the detection position of the target sequence proceeds using invasive cleavage technology. As outlined above for amplification, invasive cleavage techniques rely on the use of structure-specific nucleases, where the structure can be formed as a result of the presence or absence of a mismatch. Generally, invasive cleavage technology may be described as follows. A target nucleic acid is recognized by two distinct probes. A first probe, generally referred to herein as an "invader" probe, is substantially complementary to a first portion of the target nucleic acid. A second probe, generally referred to herein as a "signal probe", is partially complementary to the target nucleic acid; the 3' end of the signal oligonucleotide is substantially complementary to the target sequence while the 5' end is non-complementary and preferably forms a single-stranded "tail" or "arm". The non-complementary end of the second probe preferably comprises a "generic" or "unique" sequence, frequently referred to herein as a "detection sequence", that is used to indicate the presence or absence of the target nucleic acid, as described below. The detection sequence of the second probe preferably comprises at least one detectable label. Alternative methods have the detection sequence functioning as a target sequence for a capture probe, and thus rely on sandwich configurations using label probes.

Hybridization of the first and second oligonucleotides near or adjacent to one another on the target nucleic acid forms a number of structures. In a preferred embodiment, a forked cleavage structure forms and is a substrate of a nuclease which cleaves the detection sequence from the signal oligonucleotide. The site of cleavage is controlled by the distance or overlap between the 3' end of the invader oligonucleotide and the downstream fork of the signal oligonucleotide. There-

fore, neither oligonucleotide is subject to cleavage when misaligned or when unattached to target nucleic acid.

As above, the invasive cleavage assay is preferably performed on an array format. In a preferred embodiment, the signal probe has a detectable label, attached 5' from the site of nuclease cleavage (e.g. within the detection sequence) and a capture tag, as described herein for removal of the unreacted products (e.g. biotin or other hapten) 3' from the site of nuclease cleavage. After the assay is carried out, the uncleaved probe and the 3' portion of the cleaved signal probe (e.g. the detection sequence) may be extracted, for example, by binding to streptavidin beads or by crosslinking through the capture tag to produce aggregates or by antibody to an attached hapten. By "capture tag" herein is a meant one of a pair of binding partners as described above, such as antigen/antibody pairs, digoxygenenin, dinitrophenol, etc.

The cleaved 5' region, e.g. the detection sequence, of the signal probe, comprises a label and is detected and optionally quantitated. In one embodiment, the cleaved 5' region is hybridized to a probe on an array (capture probe) and optically detected. As described below, many different signal probes can be analyzed in parallel by hybridization to their complementary probes in an array. In a preferred embodiment, combination techniques are used to obtain higher specificity and reduce the detection of contaminating uncleaved signal probe or incorrectly cleaved product, an enzymatic recognition step is introduced in the array capture procedure. For example, as more fully outlined below, the cleaved signal probe binds to a capture probe to produce a double-stranded nucleic acid in the array. In this embodiment, the 3' end of the cleaved signal probe is adjacent to the 5' end of one strand of the capture probe, thereby, forming a substrate for DNA ligase (Broude et al. 1991. PNAS 91: 3072-3076). Only correctly cleaved product is ligated to the capture probe. Other incorrectly hybridized and non-cleaved signal probes are removed, for example, by heat denaturation, high stringency washes, and other methods that disrupt base pairing.

Accordingly, the present invention provides methods of determining the identity of a base at the detection position of a target sequence. In this embodiment, the target sequence comprises, 5' to 3', a first target domain comprising an overlap domain comprising at least a nucleotide in the detection position, and a second target domain contiguous with the detection position. A first probe (the "invader probe") is hybridized to the first target domain of the target sequence. A second probe (the "signal probe"), comprising a first portion that hybridizes to the second target domain of the target sequence and a second portion that does not hybridize to the target sequence, is hybridized to the second target domain. If the second probe comprises a base that is perfectly complementary to the detection position a cleavage structure is formed. The addition of a cleavage enzyme, such as is described in U.S. Pat. Nos. 5,846,717; 5,614,402; 5,719,029; 5,541,311 and 5,843,669, all of which are expressly incorporated by reference, results in the cleavage of the detection sequence from the signaling probe. This then can be used as a target sequence in an assay complex.

In addition, as for a variety of the techniques outlined herein, unreacted probes (i.e. signaling probes, in the case of invasive cleavage), may be removed using any number of techniques. For example, the use of a binding partner coupled to a solid support comprising the other member of the binding pair can be done. Similarly, after cleavage of the primary signal probe, the newly created cleavage products can be selectively labeled at the 3' or 5' ends using enzymatic or chemical methods.

Again, as outlined above, the detection of the invasive cleavage reaction can occur directly, in the case where the detection sequence comprises at least one label, or indirectly, using sandwich assays, through the use of additional probes; that is, the detection sequences can serve as target sequences, and detection may utilize amplification probes, capture probes, capture extender probes, label probes, and label extender probes, etc.

In addition, as for most of the techniques outlined herein, these techniques may be done for the two strands of a double-stranded target sequence. The target sequence is denatured, and two sets of probes are added: one set as outlined above for one strand of the target, and a separate set for the other strand of the target.

Thus, the invasive cleavage reaction requires, in no particular order, an invader probe, a signaling probe, and a cleavage enzyme.

As for other methods outlined herein, the invasive cleavage reaction may be done as a solution based assay or a solid phase assay.

Solution-Based Invasive Cleavage

The invasive cleavage reaction may be done in solution, followed by addition of one of the components to an array, with optional (but preferable) removal of unreacted probes. For example, the reaction is carried out in solution, using a capture tag (i.e. a member of a binding partner pair) that is separated from the label on the detection sequence with the cleavage site. After cleavage (dependent on the base at the detection position), the signaling probe is cleaved. The capture tag is used to remove the uncleaved probes (for example, using magnetic particles comprising the other member of the binding pair), and the remaining solution is added to the array. The detection sequence can be directly attached to the capture probe. In this embodiment, the detection sequence can effectively act as an adapter sequence. In alternate embodiments, the detection sequence is unlabelled and an additional label probe is used; as outlined below, this can be ligated to the hybridization complex.

Solid-Phase Based Assays

The invasive cleavage reaction can also be done as a solid-phase assay. The target sequence can be attached to the array using a capture probe (in addition, although not shown, the target sequence may be directly attached to the array). In a preferred embodiment, the signaling probe comprises both a fluorophore label (attached to the portion of the signaling probe that hybridizes to the target) and a quencher (generally on the detection sequence), with a cleavage site in between. Thus, in the absence of cleavage, very little signal is seen due to the quenching reaction. After cleavage, however, the detection sequence is removed, along with the quencher, leaving the unquenched fluorophore. Similarly, the invasive probe may be attached to the array.

In a preferred embodiment, the invasive cleavage reaction is configured to utilize a fluorophore-quencher reaction. A signaling probe comprising both a fluorophore and a quencher is attached to the bead. The fluorophore is contained on the portion of the signaling probe that hybridizes to the target sequence, and the quencher is contained on a portion of the signaling probe that is on the other side of the cleavage site (termed the "detection sequence" herein). In a preferred embodiment, it is the 3' end of the signaling probe that is attached to the bead (although as will be appreciated by those in the art, the system can be configured in a variety of different ways, including methods that would result in a loss of signal upon cleavage). Thus, the quencher molecule is located 5' to the cleavage site. Upon assembly of an assay complex, comprising the target sequence, an invader probe, and a signaling

probe, and the introduction of the cleavage enzyme, the cleavage of the complex results in the disassociation of the quencher from the complex, resulting in an increase in fluorescence.

In this embodiment, suitable fluorophore-quencher pairs are as known in the art. For example, suitable quencher molecules comprise Dabcyl.

Redundant Genotyping

In a preferred embodiment, the invention provides a method of increasing the confidence of genotyping results. The method includes performing genotyping more than once on a particular target sequence. That is, a sample or target analyte is genotyped at least twice. Preferably, the sample is genotyped with different techniques such as Invader™ and OLA as described herein. If the results of the individual genotyping assays agree, then confidence that the genotyping results are correct is increased.

Amplification Reactions

In this embodiment, the invention provides compositions and methods for amplification and/or detection (and optionally quantification) of products of nucleic acid amplification reactions. Suitable amplification methods include both target amplification and signal amplification. Target amplification involves the amplification (i.e. replication) of the target sequence to be detected, resulting in a significant increase in the number of target molecules. Target amplification strategies include but are not limited to the polymerase chain reaction (PCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA).

Alternatively, rather than amplify the target, alternate techniques use the target as a template to replicate a signaling probe, allowing a small number of target molecules to result in a large number of signaling probes, that then can be detected. Signal amplification strategies include the ligase chain reaction (LCR), cycling probe technology (CPT), invasive cleavage techniques such as Invader™ technology, Q-Beta replicase (QβR) technology, and the use of "amplification probes" such as "branched DNA" that result in multiple label probes binding to a single target sequence.

All of these methods require a primer nucleic acid (including nucleic acid analogs) that is hybridized to a target sequence to form a hybridization complex, and an enzyme is added that in some way modifies the primer to form a modified primer. For example, PCR generally requires two primers, dNTPs and a DNA polymerase; LCR requires two primers that adjacently hybridize to the target sequence and a ligase; CPT requires one cleavable primer and a cleaving enzyme; invasive cleavage requires two primers and a cleavage enzyme; etc. Thus, in general, a target nucleic acid is added to a reaction mixture that comprises the necessary amplification components, and a modified primer is formed.

In general, the modified primer comprises a detectable label, such as a fluorescent label, which is either incorporated by the enzyme or present on the original primer. As required, the unreacted primers are removed, in a variety of ways, as will be appreciated by those in the art and outlined herein. The hybridization complex is then disassociated, and the modified primer is detected and optionally quantitated by a microsphere array. In some cases, the newly modified primer serves as a target sequence for a secondary reaction, which then produces a number of amplified strands, which can be detected as outlined herein.

Accordingly, the reaction starts with the addition of a primer nucleic acid to the target sequence which forms a hybridization complex. Once the hybridization complex between the primer and the target sequence has been formed, an enzyme, sometimes termed an "amplification enzyme", is used to modify the primer. As for all the methods outlined herein, the enzymes may be added at any point during the assay, either prior to, during, or after the addition of the primers. The identity of the enzyme will depend on the amplification technique used, as is more fully outlined below. Similarly, the modification will depend on the amplification technique, as outlined below.

Once the enzyme has modified the primer to form a modified primer, the hybridization complex is disassociated. In one aspect, dissociation is by modification of the assay conditions. In another aspect, the modified primer no longer hybridizes to the target nucleic acid and dissociates. Either one or both of these aspects can be employed in signal and target amplification reactions as described below. Generally, the amplification steps are repeated for a period of time to allow a number of cycles, depending on the number of copies of the original target sequence and the sensitivity of detection, with cycles ranging from 1 to thousands, with from 10 to 100 cycles being preferred and from 20 to 50 cycles being especially preferred. When linear strand displacement amplification is used cycle numbers can reach thousands to millions.

After a suitable time of amplification, unreacted primers are removed, in a variety of ways, as will be appreciated by those in the art and described below, and the hybridization complex is disassociated. In general, the modified primer comprises a detectable label, such as a fluorescent label, which is either incorporated by the enzyme or present on the original primer, and the modified primer is detected by any of the methods as known to the skilled artisan and include but are not limited to the methods described herein

Target Amplification

In a preferred embodiment, the amplification is target amplification. Target amplification involves the amplification (replication) of the target sequence to be detected, such that the number of copies of the target sequence is increased. Suitable target amplification techniques include, but are not limited to, the polymerase chain reaction (PCR), strand displacement amplification (SDA), transcription mediated amplification (TMA) and nucleic acid sequence based amplification (NASBA).

Polymerase Chain Reaction Amplification

In a preferred embodiment, the target amplification technique is PCR. The polymerase chain reaction (PCR) is widely used and described, and involves the use of primer extension combined with thermal cycling to amplify a target sequence; see U.S. Pat. Nos. 4,683,195 and 4,683,202, and PCR Essential Data, J. W. Wiley & sons, Ed. C. R. Newton, 1995, all of which are incorporated by reference. In addition, there are a number of variations of PCR which also find use in the invention, including "quantitative competitive PCR" or "QC-PCR", "arbitrarily primed PCR" or "AP-PCR", "immuno-PCR", "Alu-PCR", "PCR single strand conformational polymorphism" or "PCR-SSCP", "reverse transcriptase PCR" or "RT-PCR", "biotin capture PCR", "vectorette PCR", "panhandle PCR", and "PCR select cDNA subtraction", "allele-specific PCR", among others. In some embodiments, PCR is not preferred.

In general, PCR may be briefly described as follows. A double stranded target nucleic acid is denatured, generally by raising the temperature, and then cooled in the presence of an excess of a PCR primer, which then hybridizes to the first target strand. A DNA polymerase then acts to extend the primer with dNTPs, resulting in the synthesis of a new strand forming a hybridization complex. The sample is then heated again, to disassociate the hybridization complex, and the process is repeated. By using a second PCR primer for the complementary target strand, rapid and exponential amplifi-

cation occurs. Thus PCR steps are denaturation, annealing and extension. The particulars of PCR are well known, and include the use of a thermostable polymerase such as Taq I polymerase and thermal cycling.

Accordingly, the PCR reaction requires at least one PCR primer, a polymerase, and a set of dNTPs. As outlined herein, the primers may comprise the label, or one or more of the dNTPs may comprise a label.

In one embodiment asymmetric PCR is performed. In this embodiment, unequal concentrations of primers are included in the amplification reaction. The concentrations are designed such that one primer is in excess or is saturating, while the other primer is limiting or is at a sub-saturating concentration.

In one embodiment, PCR primers for amplification of a plurality of target nucleic acids are immobilized on a single bead. That is, at least first and second PCR primer pairs are immobilized to a bead or microsphere. The microsphere is contacted with a sample and PCR performed as described herein. Detection of the amplified product or products is accomplished by any of the detection methods described herein, but in a preferred embodiment, detection proceeds by hybridization with allele specific oligonucleotides. That is, upon amplification of the target nucleotides, the immobilized PCR product is hybridized with oligonucleotides that are complementary to the amplified product.

In a preferred embodiment the allele specific oligonucleotides contain discrete labels. That is, the oligonucleotides contain distinguishable labels. As a result of hybridization between the allele specific oligonucleotides and the amplified product(s), detection of a particular label provides an indication of the presence of a particular target nucleic acid in the sample.

In one embodiment, the PCR primers are designed to amplify different genomic markers. That is, markers such as translocations or other chromosomal abnormalities are targeted for amplification. In an additional embodiment, the primers are designed to amplify genomic regions containing single nucleotide polymorphisms (SNPs). As such, the resulting hybridization with allele specific oligonucleotides provides an indication of the marker or SNP. In one embodiment, a plurality of markers or SNPs is detected on each bead. That is, at least two markers or SNPs are detected on each bead.

In general, as is more fully outlined below, the capture probes on the beads of the array are designed to be substantially complementary to the extended part of the primer; that is, unextended primers will not bind to the capture probes. Alternatively, as further described below, unreacted probes may be removed prior to addition to the array.

In a preferred embodiment the amplification reaction as a multiplex amplification reaction as described herein. In one embodiment the amplification reaction uses a plurality of PCR primers to amplify a plurality of target sequences. In this embodiment plurality of target sequences are simultaneously amplified with the plurality of amplification primer pairs.

An alternative embodiment the multiplex PCR reaction uses universal primers as described herein. That is, universal PCR primers hybridized to universal priming sites on the target sequence and thereby amplify a plurality of target sequences. This embodiment is potentially preferred because it requires only a limited number of PCR primers. That is, as few as one primer pairs can amplify a plurality of target sequences.

Strand Displacement Amplification (SDA)

In a preferred embodiment, the target amplification technique is SDA. Strand displacement amplification (SDA) is generally described in Walker et al., in Molecular Methods for Virus Detection, Academic Press, Inc., 1995, and U.S. Pat. Nos. 5,455,166 and 5,130,238, all of which are hereby expressly incorporated by reference in their entirety.

In general, SDA may be described as follows. A single stranded target nucleic acid, usually a DNA target sequence, is contacted with an SDA primer. An "SDA primer" generally has a length of 25-100 nucleotides, with SDA primers of approximately 35 nucleotides being preferred. An SDA primer is substantially complementary to a region at the 3' end of the target sequence, and the primer has a sequence at its 5' end (outside of the region that is complementary to the target) that is a recognition sequence for a restriction endonuclease, sometimes referred to herein as a "nicking enzyme" or a "nicking endonuclease", as outlined below. The SDA primer then hybridizes to the target sequence. The SDA reaction mixture also contains a polymerase (an "SDA polymerase", as outlined below) and a mixture of all four deoxynucleoside-triphosphates (also called deoxynucleotides or dNTPs, i.e. dATP, dTTP, dCTP and dGTP), at least one species of which is a substituted or modified dNTP; thus, the SDA primer is modified, i.e. extended, to form a modified primer, sometimes referred to herein as a "newly synthesized strand". The substituted dNTP is modified such that it will inhibit cleavage in the strand containing the substituted dNTP but will not inhibit cleavage on the other strand. Examples of suitable substituted dNTPs include, but are not limited, 2'deoxyadenosine 5'-O-(1-thiotriphosphate), 5-methyldeoxycytidine 5'-triphosphate, 2'-deoxyuridine 5'-triphosphate, and 7-deaza-2'-deoxyguanosine 5'-triphosphate. In addition, the substitution of the dNTP may occur after incorporation into a newly synthesized strand; for example, a methylase may be used to add methyl groups to the synthesized strand. In addition, if all the nucleotides are substituted, the polymerase may have 5'→3' exonuclease activity. However, if less than all the nucleotides are substituted, the polymerase preferably lacks 5'→3' exonuclease activity.

As will be appreciated by those in the art, the recognition site/endonuclease pair can be any of a wide variety of known combinations. The endonuclease is chosen to cleave a strand either at the recognition site, or either 3' or 5' to it, without cleaving the complementary sequence, either because the enzyme only cleaves one strand or because of the incorporation of the substituted nucleotides. Suitable recognition site/endonuclease pairs are well known in the art; suitable endonucleases include, but are not limited to, HincII, HindII, AvaI, Fnu4HI, TthIII, NcII, BstXI, BamHI, etc. A chart depicting suitable enzymes, and their corresponding recognition sites and the modified dNTP to use is found in U.S. Pat. No. 5,455,166, hereby expressly incorporated by reference.

Once nicked, a polymerase (an "SDA polymerase") is used to extend the newly nicked strand, 5'→3', thereby creating another newly synthesized strand. The polymerase chosen should be able to intiate 5'→3' polymerization at a nick site, should also displace the polymerized strand downstream from the nick, and should lack 5'→3' exonuclease activity (this may be additionally accomplished by the addition of a blocking agent). Thus, suitable polymerases in SDA include, but are not limited to, the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase.

Accordingly, the SDA reaction requires, in no particular order, an SDA primer, an SDA polymerase, a nicking endonuclease, and dNTPs, at least one species of which is modified. Again, as outlined above for PCR, preferred embodiments utilize capture probes complementary to the newly synthesized portion of the primer, rather than the primer region, to allow unextended primers to be removed.

In general, SDA does not require thermocycling. The temperature of the reaction is generally set to be high enough to prevent non-specific hybridization but low enough to allow specific hybridization; this is generally from about 37° C. to about 42° C., depending on the enzymes.

In a preferred embodiment, as for most of the amplification techniques described herein, a second amplification reaction can be done using the complementary target sequence, resulting in a substantial increase in amplification during a set period of time. That is, a second primer nucleic acid is hybridized to a second target sequence, that is substantially complementary to the first target sequence, to form a second hybridization complex. The addition of the enzyme, followed by disassociation of the second hybridization complex, results in the generation of a number of newly synthesized second strands.

Nucleic Acid Sequence Based Amplification (NASBA) and Transcription Mediated Amplification (TMA)

In a preferred embodiment, the target amplification technique is nucleic acid sequence based amplification (NASBA). NASBA is generally described in U.S. Pat. No. 5,409,818; Sooknanan et al., Nucleic Acid Sequence-Based Amplification, Ch. 12 (pp. 261-285) of Molecular Methods for Virus Detection, Academic Press, 1995; and "Profiting from Gene-based Diagnostics", CTB International Publishing Inc., N.J., 1996, all of which are incorporated by reference. NASBA is very similar to both TMA and QBR. Transcription mediated amplification (TMA) is generally described in U.S. Pat. Nos. 5,399,491, 5,888,779, 5,705,365, 5,710,029, all of which are incorporated by reference. The main difference between NASBA and TMA is that NASBA utilizes the addition of RNAse H to effect RNA degradation, and TMA relies on inherent RNAse H activity of the reverse transcriptase.

In general, these techniques may be described as follows. A single stranded target nucleic acid, usually an RNA target sequence (sometimes referred to herein as "the first target sequence" or "the first template"), is contacted with a first primer, generally referred to herein as a "NASBA primer" (although "TMA primer" is also suitable). Starting with a DNA target sequence is described below. These primers generally have a length of 25-100 nucleotides, with NASBA primers of approximately 50-75 nucleotides being preferred. The first primer is preferably a DNA primer that has at its 3' end a sequence that is substantially complementary to the 3' end of the first template. The first primer also has an RNA polymerase promoter at its 5' end (or its complement (antisense), depending on the configuration of the system). The first primer is then hybridized to the first template to form a first hybridization complex. The reaction mixture also includes a reverse transcriptase enzyme (an "NASBA reverse transcriptase") and a mixture of the four dNTPs, such that the first NASBA primer is modified, i.e. extended, to form a modified first primer, comprising a hybridization complex of RNA (the first template) and DNA (the newly synthesized strand).

By "reverse transcriptase" or "RNA-directed DNA polymerase" herein is meant an enzyme capable of synthesizing DNA from a DNA primer and an RNA template. Suitable RNA-directed DNA polymerases include, but are not limited to, avian myloblastosis virus reverse transcriptase ("AMV RT") and the Moloney murine leukemia virus RT. When the amplification reaction is TMA, the reverse transcriptase enzyme further comprises a RNA degrading activity as outlined below.

In addition to the components listed above, the NASBA reaction also includes an RNA degrading enzyme, also sometimes referred to herein as a ribonuclease, that will hydrolyze RNA of an RNA:DNA hybrid without hydrolyzing single- or double-stranded RNA or DNA. Suitable ribonucleases include, but are not limited to, RNase H from *E. coli* and calf thymus.

The ribonuclease activity degrades the first RNA template in the hybridization complex, resulting in a disassociation of the hybridization complex leaving a first single stranded newly synthesized DNA strand, sometimes referred to herein as "the second template".

In addition, the NASBA reaction also includes a second NASBA primer, generally comprising DNA (although as for all the probes herein, including primers, nucleic acid analogs may also be used). This second NASBA primer has a sequence at its 3' end that is substantially complementary to the 3' end of the second template, and also contains an antisense sequence for a functional promoter and the antisense sequence of a transcription initiation site. Thus, this primer sequence, when used as a template for synthesis of the third DNA template, contains sufficient information to allow specific and efficient binding of an RNA polymerase and initiation of transcription at the desired site. Preferred embodiments utilizes the antisense promoter and transcription initiation site are that of the T7 RNA polymerase, although other RNA polymerase promoters and initiation sites can be used as well, as outlined below.

The second primer hybridizes to the second template, and a DNA polymerase, also termed a "DNA-directed DNA polymerase", also present in the reaction, synthesizes a third template (a second newly synthesized DNA strand), resulting in second hybridization complex comprising two newly synthesized DNA strands.

Finally, the inclusion of an RNA polymerase and the required four ribonucleoside triphosphates (ribonucleotides or NTPs) results in the synthesis of an RNA strand (a third newly synthesized strand that is essentially the same as the first template). The RNA polymerase, sometimes referred to herein as a "DNA-directed RNA polymerase", recognizes the promoter and specifically initiates RNA synthesis at the initiation site. In addition, the RNA polymerase preferably synthesizes several copies of RNA per DNA duplex. Preferred RNA polymerases include, but are not limited to, T7 RNA polymerase, and other bacteriophage RNA polymerases including those of phage T3, phage ϕII, Salmonella phage sp6, or Pseudomonase phage gh-1.

In some embodiments, TMA and NASBA are used with starting DNA target sequences. In this embodiment, it is necessary to utilize the first primer comprising the RNA polymerase promoter and a DNA polymerase enzyme to generate a double stranded DNA hybrid with the newly synthesized strand comprising the promoter sequence. The hybrid is then denatured and the second primer added.

Accordingly, the NASBA reaction requires, in no particular order, a first NASBA primer, a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, a DNA polymerase, an RNA degrading enzyme, NTPs and dNTPs, in addition to the detection components outlined below.

These components result in a single starting RNA template generating a single DNA duplex; however, since this DNA duplex results in the creation of multiple RNA strands, which can then be used to initiate the reaction again, amplification proceeds rapidly.

Accordingly, the TMA reaction requires, in no particular order, a first TMA primer, a second TMA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse tran-

scriptase with RNA degrading activity, a DNA polymerase, NTPs and dNTPs, in addition to the detection components outlined below.

These components result in a single starting RNA template generating a single DNA duplex; however, since this DNA duplex results in the creation of multiple RNA strands, which can then be used to initiate the reaction again, amplification proceeds rapidly.

As outlined herein, the detection of the newly synthesized strands can proceed in several ways. Direct detection can be done when the newly synthesized strands comprise detectable labels, either by incorporation into the primers or by incorporation of modified labelled nucleotides into the growing strand. Alternatively, as is more fully outlined below, indirect detection of unlabelled strands (which now serve as "targets" in the detection mode) can occur using a variety of sandwich assay configurations. As will be appreciated by those in the art, any of the newly synthesized strands can serve as the "target" for form an assay complex on a surface with a capture probe. In NASBA and TMA, it is preferable to utilize the newly formed RNA strands as the target, as this is where significant amplification occurs.

In this way, a number of secondary target molecules are made. As is more fully outlined below, these reactions (that is, the products of these reactions) can be detected in a number of ways.

Rolling-Circle Amplification (RCA)

In a preferred embodiment the signal amplification technique is RCA. Rolling-circle amplification is generally described in Baner et al. (1998) *Nuc. Acids Res*. 26:5073-5078; Barany, F. (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193; and Lizardi et al. (1998) *Nat. Genet*. 19:225-232, all of which are incorporated by reference in their entirety.

In general, RCA may be described in two ways. First, as is outlined in more detail below, a single probe is hybridized with a target nucleic acid. Each terminus of the probe hybridizes adjacently on the target nucleic acid and the OLA assay as described above occurs. Alternatively, two probes are hybridized with the target nucleic acid and the OLA assay as described above occurs. When ligated, the probe is circularized while hybridized to the target nucleic acid, or a circular primer is added to the ligated target nucleic acid complex. Addition of a polymerase results in extension of the circular probe. However, since the probe has no terminus, the polymerase continues to extend the probe repeatedly. Thus results in amplification of the circular probe.

A second alternative approach involves OLA followed by RCA. In this embodiment, an immobilized primer is contacted with a target nucleic acid. Complementary sequences will hybridize with each other resulting in an immobilized duplex. A second primer is contacted with the target nucleic acid. The second primer hybridizes to the target nucleic acid adjacent to the first primer. An OLA assay is performed as described above. Ligation only occurs if the primer are complementary to the target nucleic acid. When a mismatch occurs, particularly at one of the nucleotides to be ligated, ligation will not occur. Following ligation of the oligonucleotides, the ligated, immobilized, oligonucleotide is then hybridized with an RCA probe. This is a circular probe that is designed to specifically hybridize with the ligated oligonucleotide and will only hybridize with an oligonucleotide that has undergone ligation. RCA is then performed as is outlined in more detail below.

Accordingly, in an preferred embodiment, a single oligonucleotide is used both for OLA and as the circular template for RCA (referred to herein as a "padlock probe" or a "RCA probe"). That is, each terminus of the oligonucleotide contains sequence complementary to the target nucleic acid and functions as an OLA primer as described above. That is, the first end of the RCA probe is substantially complementary to a first target domain, and the second end of the RCA probe is substantially complementary to a second target domain, adjacent to the first domain. Hybridization of the oligonucleotide to the target nucleic acid results in the formation of a hybridization complex. Ligation of the "primers" (which are the discrete ends of a single oligonucleotide) results in the formation of a modified hybridization complex containing a circular probe i.e. an RCA template complex. That is, the oligonucleotide is circularized while still hybridized with the target nucleic acid. This serves as a circular template for RCA. Addition of a polymerase to the RCA template complex results in the formation of an amplified product nucleic acid. Following RCA, the amplified product nucleic acid is detected. This can be accomplished in a variety of ways; for example, the polymerase may incorporate labeled nucleotides, or alternatively, a label probe is used that is substantially complementary to a portion of the RCA probe and comprises at least one label is used.

The polymerase can be any polymerase, but is preferably one lacking 3' exonuclease activity (3' exo$^-$). Examples of suitable polymerase include but are not limited to exonuclease minus DNA Polymerase I large (Klenow) Fragment, Phi29 DNA polymerase, Taq DNA Polymerase and the like. In addition, in some embodiments, a polymerase that will replicate single-stranded DNA (i.e. without a primer forming a double stranded section) can be used.

In a preferred embodiment, the RCA probe contains an adapter sequence as outlined herein, with adapter capture probes on the array, for example on a microsphere when microsphere arrays are being used. Alternatively, unique portions of the RCA probes, for example all or part of the sequence corresponding to the target sequence, can be used to bind to a capture probe.

In a preferred embodiment, the padlock probe contains a restriction site. The restriction endonuclease site allows for cleavage of the long concatamers that are typically the result of RCA into smaller individual units that hybridize either more efficiently or faster to surface bound capture probes. Thus, following RCA, the product nucleic acid is contacted with the appropriate restriction endonuclease. This results in cleavage of the product nucleic acid into smaller fragments. The fragments are then hybridized with the capture probe that is immobilized resulting in a concentration of product fragments onto the microsphere. Again, as outlined herein, these fragments can be detected in one of two ways: either labelled nucleotides are incorporated during the replication step, or an additional label probe is added.

Thus, in a preferred embodiment, the padlock probe comprises a label sequence; i.e. a sequence that can be used to bind label probes and is substantially complementary to a label probe. In one embodiment, it is possible to use the same label sequence and label probe for all padlock probes on an array; alternatively, each padlock probe can have a different label sequence.

The padlock probe also contains a priming site for priming the RCA reaction. That is, each padlock probe comprises a sequence to which a primer nucleic acid hybridizes forming a template for the polymerase. The primer can be found in any portion of the circular probe. In a preferred embodiment, the primer is located at a discrete site in the probe. In this embodiment, the primer site in each distinct padlock probe is identical, although this is not required. Advantages of using primer sites with identical sequences include the ability to use only a single primer oligonucleotide to prime the RCA assay

with a plurality of different hybridization complexes. That is, the padlock probe hybridizes uniquely to the target nucleic acid to which it is designed. A single primer hybridizes to all of the unique hybridization complexes forming a priming site for the polymerase. RCA then proceeds from an identical locus within each unique padlock probe of the hybridization complexes.

In an alternative embodiment, the primer site can overlap, encompass, or reside within any of the above-described elements of the padlock probe. That is, the primer can be found, for example, overlapping or within the restriction site or the identifier sequence. In this embodiment, it is necessary that the primer nucleic acid is designed to base pair with the chosen primer site.

Thus, the padlock probe of the invention contains at each terminus, sequences corresponding to OLA primers. The intervening sequence of the padlock probe contain in no particular order, an adapter sequence and a restriction endonuclease site. In addition, the padlock probe contains a RCA priming site.

Thus, in a preferred embodiment the OLA/RCA is performed in solution followed by restriction endonuclease cleavage of the RCA product. The cleaved product is then applied to an array comprising beads, each bead comprising a probe complementary to the adapter sequence located in the padlock probe. The amplified adapter sequence correlates with a particular target nucleic acid. Thus the incorporation of an endonuclease site allows the generation of short, easily hybridizable sequences. Furthermore, the unique adapter sequence in each rolling circle padlock probe sequence allows diverse sets of nucleic acid sequences to be analyzed in parallel on an array, since each sequence is resolved on the basis of hybridization specificity.

In an alternative OLA/RCA method, one of the OLA primers is immobilized on the microsphere; the second primer is added in solution. Both primers hybridize with the target nucleic acid forming a hybridization complex as described above for the OLA assay.

As described herein, the microsphere is distributed on an array. In a preferred embodiment, a plurality of microspheres each with a unique OLA primer is distributed on the array.

Following the OLA assay, and either before, after or concurrently with distribution of the beads on the array, a segment of circular DNA is hybridized to the bead-based ligated oligonucleotide forming a modified hybridization complex. Addition of an appropriate polymerase (3' exo$^-$), as is known in the art, and corresponding reaction buffer to the array leads to amplification of the circular DNA. Since there is no terminus to the circular DNA, the polymerase continues to travel around the circular template generating extension product until it detaches from the template. Thus, a polymerase with high processivity can create several hundred or thousand copies of the circular template with all the copies linked in one contiguous strand.

Again, these copies are subsequently detected by one of two methods; either hybridizing a labeled oligo complementary to the circular target or via the incorporation of labeled nucleotides in the amplification reaction. The label is detected using conventional label detection methods as described herein.

In one embodiment, when the circular DNA contains sequences complementary to the ligated oligonucleotide it is preferable to remove the target DNA prior to contacting the ligated oligonucleotide with the circular DNA (See FIG. 7). This is done by denaturing the double-stranded DNA by methods known in the art. In an alternative embodiment, the double stranded DNA is not denatured prior to contacting the circular DNA.

In an alternative embodiment, when the circular DNA contains sequences complementary to the target nucleic acid, it is preferable that the circular DNA is complementary at a site distinct from the site bound to the ligated oligonucleotide. In this embodiment it is preferred that the duplex between the ligated oligonucleotide and target nucleic acid is not denatured or disrupted prior to the addition of the circular DNA so that the target DNA remains immobilized to the bead.

Hybridization and washing conditions are well known in the art; various degrees of stringency can be used. In some embodiments it is not necessary to use stringent hybridization or washing conditions as only microspheres containing the ligated probes will effectively hybridize with the circular DNA; microspheres bound to DNA that did not undergo ligation (those without the appropriate target nucleic acid) will not hybridize as strongly with the circular DNA as those primers that were ligated. Thus, hybridization and/or washing conditions are used that discriminate between binding of the circular DNA to the ligated primer and the unligated primer.

Alternatively, when the circular probe is designed to hybridize to the target nucleic acid at a site distinct from the site bound to the ligated oligonucleotide, hybridization and washing conditions are used to remove or dissociate the target nucleic acid from unligated oligonucleotides while target nucleic acid hybridizing with the ligated oligonucleotides will remain bound to the beads. In this embodiment, the circular probe only hybridizes to the target nucleic acid when the target nucleic acid is hybridized with a ligated oligonucleotide that is immobilized on a bead.

As is well known in the art, an appropriate polymerase (3' exo$^-$) is added to the array. The polymerase extends the sequence of a single-stranded DNA using double-stranded DNA as a primer site. In one embodiment, the circular DNA that has hybridized with the appropriate OLA reaction product serves as the primer for the polymerase. In the presence of an appropriate reaction buffer as is known in the art, the polymerase will extend the sequence of the primer using the single-stranded circular DNA as a template. As there is no terminus of the circular DNA, the polymerase will continue to extend the sequence of the circular DNA. In an alternative embodiment, the RCA probe comprises a discrete primer site located within the circular probe. Hybridization of primer nucleic acids to this primer site forms the polymerase template allowing RCA to proceed.

In a preferred embodiment, the polymerase creates more than 100 copies of the circular DNA. In more preferred embodiments the polymerase creates more than 1000 copies of the circular DNA; while in a most preferred embodiment the polymerase creates more than 10,000 copies or more than 50,000 copies of the template.

The amplified circular DNA sequence is then detected by methods known in the art and as described herein. Detection is accomplished by hybridizing with a labeled probe. The probe is labeled directly or indirectly. Alternatively, labeled nucleotides are incorporated into the amplified circular DNA product. The nucleotides can be labeled directly, or indirectly as is further described herein.

The RCA as described herein finds use in allowing highly specific and highly sensitive detection of nucleic acid target sequences. In particular, the method finds use in improving the multiplexing ability of DNA arrays and eliminating costly sample or target preparation. As an example, a substantial savings in cost can be realized by directly analyzing genomic DNA on an array, rather than employing an intermediate PCR

amplification step. The method finds use in examining genomic DNA and other samples including mRNA.

In addition the RCA finds use in allowing rolling circle amplification products to be easily detected by hybridization to probes in a solid-phase format (e.g. an array of beads). An additional advantage of the RCA is that it provides the capability of multiplex analysis so that large numbers of sequences can be analyzed in parallel. By combining the sensitivity of RCA and parallel detection on arrays, many sequences can be analyzed directly from genomic DNA.

In an alternative embodiment, the OLA assay includes employing a standard solution phase OLA assay using adapter sequences to capture the OLA product. In this case, the allele specific oligonucleotides also contain a sequence that is complementary to a circular RCA primer that is indicative of the respective allele. That is, the OLA primer designed to hybridize to one allele contains a specific sequence for hybridization to a specific RCA primer. Likewise, the OLA primer designed to hybridize to the second allele contains a specific sequence for hybridization to a second specific RCA primer. Following OLA and capture of the OLA product, both RCA primers are hybridized with the OLA product, but only the RCA primer that is complementary to the respective RCA primer site will hybridize with that site. An RCA assay is performed and the product detected as described herein. The RCA product is an indication of the presence of a particular allele.

In one embodiment RCA is used to amplify cDNA. As is known in the art, cDNA is obtained by reverse transcription of mRNA. The resulting cDNA, therefore is a representation of the mRNA population in a given sample. Accordingly, it is desirable to examine cDNA to gain insight into the relative level of mRNA of a sample. However, frequently there exists a need to amplify the cDNA in order to obtain sufficient quantities for various analyses. Previously, amplification strategies involved exponential techniques such as PCR. A potential problem with exponential amplification is that it occasionally results in distorted mRNA profiles. Given the desire to examine mRNA populations, which provide an indication of the expression level of different gene products, there is a desire to develop amplification techniques that provide a more accurate indication of the mRNA levels in a sample.

Accordingly, the present invention provides a method of amplifying cDNA using the RCA as described herein. In a preferred embodiment, the method includes circularizing the cDNA and amplifying the circularized substrate with a DNA polymerase. In a preferred embodiment the cDNA is circularized by hybridization with a "guide linker". By "guide linker" is meant an oligonucleotide that is complementary to the 5' and 3' termini of the cDNA molecule. Generally, the 5' terminus of a cDNA molecule contains a poly-T track. In addition, the 3' terminus of cDNA frequently contains multiple C nucleotides. Generally three or four C nucleotides are added to the 3' terminus of the cDNA. Without being bound by theory, it is thought that these Cs are a result of non-template mediated addition of the C nucleotides to the 3' terminus by the DNA Polymerase. Accordingly, in a preferred embodiment the guide linker contains a plurality of A nucleotides at one terminus and a plurality of G nucleotides at the other terminus. That is, it contains at its 5' terminus a plurality of G nucleotides and at its 3' terminus a plurality of A nucleotides. A preferred guide linker contains the sequence GGGAAAA, although it could contain more or fewer Gs or As at each of the respective termini.

Upon hybridization of the guide linker with the cDNA, the circular cDNA is covalently closed following incubation with ligase. That is, incubation with ligase results in covalent attachment of the 5'T and 3'C of the cDNA). The circular cDNA/guide linker complex is then contacted with a DNA polymerase that extends the circular template as described herein. The cDNA/guide linker complex serves as a template for the polymerase. This results in linear amplification of the cDNA and results in a population of cDNA that is representative of the mRNA levels of a sample. That is, the amplified cDNA provides an indication of the gene expression level of a sample. In addition, the amplified products represent full length cDNAs as a result of selection with a guide linker that contains a poly-T tract and a poly-G tract.

As described herein, in some embodiments labeled nucleotides are incorporated into the amplified cDNA product. This results in linear amplification of the signal.

The amplified cDNA product finds use in a variety of assays including gene expression analysis. The amplified products find use as probes that can be applied to an array as described herein.

Cycling Probe Technology (CPT

Cycling probe technology (CPT) is a nucleic acid detection system based on signal or probe amplification rather than target amplification, such as is done in polymerase chain reactions (PCR). Cycling probe technology relies on a molar excess of labeled probe which contains a scissile linkage of RNA. Upon hybridization of the probe to the target, the resulting hybrid contains a portion of RNA:DNA. This area of RNA:DNA duplex is recognized by RNAseH and the RNA is excised, resulting in cleavage of the probe. The probe now consists of two smaller sequences which may be released, thus leaving the target intact for repeated rounds of the reaction. The unreacted probe is removed and the label is then detected. CPT is generally described in U.S. Pat. Nos. 5,011,769, 5,403,711, 5,660,988, and 4,876,187, and PCT published applications WO 95/05480, WO 95/1416, and WO 95/00667, all of which are specifically incorporated herein by reference.

Oligonucleotide Ligation Assay

The oligonucleotide ligation assay (OLA; sometimes referred to as the ligation chain reaction (LCR)) involve the ligation of at least two smaller probes into a single long probe, using the target sequence as the template for the ligase. See generally U.S. Pat. Nos. 5,185,243, 5,679,524 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; and WO 89/09835, all of which are incorporated by reference.

Invader™

Invader™ technology is based on structure-specific polymerases that cleave nucleic acids in a site-specific manner. Two probes are used: an "invader" probe and a "signaling" probe, that adjacently hybridize to a target sequence with a non-complementary overlap. The enzyme cleaves at the overlap due to its recognition of the "tail", and releases the "tail" with a label. This can then be detected. The Invader™ technology is described in U.S. Pat. Nos. 5,846,717; 5,614,402; 5,719,028; 5,541,311; and 5,843,669, all of which are hereby incorporated by reference.

ICAN Amplification

ICAN methodology is a preferred amplification method that includes hybridizing chimeric-primers composed of RNA (3' end) and DNA (5' end) and providing a DNA polymerase with strand displacement activity (BcaBEST™ DNA polymerase from Takara Shuzo Co., Ltd), which extends the primer forming a double stranded intermediate. Subsequently, a ribonuclease cleaves the junction of the DNA-RNA hybrid (RNase H). Subsequently, an additional chimeric primer hybridizes with the extension product or original target and displaces one strand of the double stranded interme-

diate. This cycle is repeated thereby amplifying the target. Amplification is outlined in FIG. **12**. In a preferred embodiment ICAN method can be used to amplify specific regions of DNA at a constant temperature of 50 to 65° C. That is, the amplification is isothermal.

SPIA™

In a preferred embodiment, a linear amplification scheme known as ESPIA, or SPIA is applied. This amplification technique is disclosed in WO 01/20035 A2 and U.S. Ser. No. 6,251,639, which are incorporated by reference herein. Generally, the method includes hybridizing chimeric RNA/DNA amplification primers to the probes or target. Preferably the DNA portion of the probe is 3' to the RNA. Optionally the method includes hybridizing a polynucleotide comprising a termination polynucleotide sequence to a region of the template that is 5' with respect to hybridization of the composite primer to the template. Following hybridization of the primer to the template, the primer is extended with DNA polymerase. Subsequently, the RNA is cleaved from the composite primer with an enzyme that cleaves RNA from an RNA/DNA hybrid. Subsequently, an additional RNA/DNA chimeric primer is hybridized to the template such that the first extended primer is displaced from the target probe. The extension reaction is repeated, whereby multiple copies of the probe sequence are generated.

Amplicon Enrichment

In this alternate method, following amplification, as described above, the amplicons are hybridized to a solid-phase containing immobilized targets, i.e. genomic DNA or oligonucleotides corresponding to targeted SNPs. Preferably the amplification primers include universal primers, as described herein. In a preferred embodiment hybridization is performed at high temperatures such that only the desired PCR products (those that include or span the particular allele) are retained, while non-specific products or primer-dimers, which have a reduced Tm are removed by washing. That is, the notable difference between the Tms of specific products, which are preferably form 65 to 85° C., more preferably form 70 to 80° C., and the Tms of the non-specific products, which is around from about 45-60° C., provides a separation window for controlling or discriminating between the two populations during hybridization and washing.

The immobilized target can be any nucleic acid as described herein. Preferably the immobilized target is genomic DNA or oligonucleotides corresponding to particular SNPs. Alternatively, it could be pooled genomic DNA from a variety of sources or individually amplified products.

Once the non-specific products have been removed, the retained PCR products may be detected. Alternatively, they may be additionally amplified. Alternatively, they may be used in any genotyping assays as are known in the art and described herein.

Label

By "detection label" or "detectable label" herein is meant a moiety that allows detection. This may be a primary label or a secondary label. Accordingly, detection labels may be primary labels (i.e. directly detectable) or secondary labels (indirectly detectable).

In a preferred embodiment, the detection label is a primary label. A primary label is one that can be directly detected, such as a fluorophore. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) magnetic, electrical, thermal labels; and c) colored or luminescent dyes. Labels can also include enzymes (horseradish peroxidase, etc.) and magnetic particles. Preferred labels include chromophores or phosphors but are preferably fluorescent dyes. Suitable dyes for use in the invention include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, quantum dots (also referred to as "nanocrystals": see U.S. Ser. No. 09/315,584, hereby incorporated by reference), pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, Cy dyes (Cy3, Cy5, etc.), alexa dyes, phycoerythin, bodipy, and others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference.

In a preferred embodiment, a secondary detectable label is used. A secondary label is one that is indirectly detected; for example, a secondary label can bind or react with a primary label for detection, can act on an additional product to generate a primary label (e.g. enzymes), or may allow the separation of the compound comprising the secondary label from unlabeled materials, etc. Secondary labels find particular use in systems requiring separation of labeled and unlabeled probes, such as SBE, OLA, invasive cleavage reactions, etc; in addition, these techniques may be used with many of the other techniques described herein. Secondary labels include, but are not limited to, one of a binding partner pair; chemically modifiable moieties; nuclease inhibitors, enzymes such as horseradish peroxidase, alkaline phosphatases, lucifierases, etc.

In a preferred embodiment, the secondary label is a binding partner pair. For example, the label may be a hapten or antigen, which will bind its binding partner. In a preferred embodiment, the binding partner can be attached to a solid support to allow separation of extended and non-extended primers. For example, suitable binding partner pairs include, but are not limited to: antigens (such as proteins (including peptides)) and antibodies (including fragments thereof (FAbs, etc.)); proteins and small molecules, including biotin/streptavidin; enzymes and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partners. Nucleic acid nucleic acid binding proteins pairs are also useful. In general, the smaller of the pair is attached to the NTP for incorporation into the primer. Preferred binding partner pairs include, but are not limited to, biotin (or imino-biotin) and streptavidin, digeoxinin and Abs, and Prolinx™ reagents (see www.prolinx-inc.com/ie4/home.hmtl).

In a preferred embodiment, the binding partner pair comprises biotin or imino-biotin and streptavidin. Imino-biotin is particularly preferred as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95° C.).

In a preferred embodiment, the binding partner pair comprises a primary detection label (for example, attached to the NTP and therefore to the extended primer) and an antibody that will specifically bind to the primary detection label. By "specifically bind" herein is meant that the partners bind with specificity sufficient to differentiate between the pair and other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the assay, including wash steps to remove non-specific binding. In some embodiments, the dissociation constants of the pair will be less than about 10^{-4}-10^{-6} M^{-1}, with less than about 10^{-5} to 10^{-9} M^{-1} being preferred and less than about 10^{-7}-10^{-9} M^{-1} being particularly preferred.

In a preferred embodiment, the secondary label is a chemically modifiable moiety. In this embodiment, labels comprising reactive functional groups are incorporated into the nucleic acid. The functional group can then be subsequently

labeled with a primary label. Suitable functional groups include, but are not limited to, amino groups, carboxy groups, maleimide groups, oxo groups and thiol groups, with amino groups and thiol groups being particularly preferred. For example, primary labels containing amino groups can be attached to secondary labels comprising amino groups, for example using linkers as are known in the art; for example, homo- or hetero-bifunctional linkers as are well known (see 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages 155-200, incorporated herein by reference).

However, in this embodiment, the label is a secondary label, a purification tag, that can be used to capture the sequence comprising the tag onto a second solid support surface.

The addition of the polymerase and the labeled dNTP are done under conditions to allow the formation of a modified first probe. The modified first probe is then added to a second solid support using the purification tag as outlined herein.

Once immobilized, several reagents are adding to the modified probe. In a preferred embodiment, first and second universal probes are added, with a polymerase and dNTPs, such that the modified probe is amplified to form amplicons, which can then be detected on arrays as outlined below. While the figures are generally directed to PCR systems, other amplification systems can be used, as are generally outlined in Ser. No. 09/517,945, filed Mar. 3, 2000, Ser. No. 60/161,148, filed Oct. 22, 1999, Ser. No. 60/135,051, filed May 20, 1999, Ser. No. 60/244,119, filed Oct. 26, 2000, Ser. No. 09/556,463, filed Apr. 21, 2000, and Ser. No. 09/553,993, filed Apr. 20, 2000, all of which are expressly incorporated herein by reference.

Combination Techniques

Other preferred configurations of the system are shown in the figures.

In one embodiment the target nucleic acid is first immobilized. This is followed by a specificity step, i.e. allele specific extension (see FIG. **1**) and amplification. That is, following immobilization of the target nucleic acids, the target nucleic acids are contacted with allele specific probes under stringent annealing conditions. Non-hybridized probes are removed by a stringent wash. Subsequently the hybridized probes or primers are contacted with an enzyme such as a polymerase in the presence of labeled ddNTP (see FIG. **1**) forming a modified primer. Preferably the label is a purification tag as described herein. The ddNTP is only incorporated into the primer that is perfectly complementary to the target nucleic acid. The modified primer is then eluted from the immobilized target nucleic acid, and contacted with amplification primers to form amplicons. In one embodiment the eluted primer is purified by binding to a binding partner for the affinity tag. Then the purified and modified primer is contacted with amplification primers for amplification, forming amplicons. The amplicons are then detected as an indication of the presence of the particular target nucleic acid.

In a preferred embodiment, as shown in FIG. **1**, the allele specific primer also includes an adapter sequence and priming sequences. That is, the primer includes from 5' to 3', and upstream amplification priming site, an adapter sequence, a downstream amplification priming site, and an allele specific sequence Priming sequences hybridize with amplification primers; the adapter sequence mediates attachment of the amplicons to a support for subsequent detection of amplicons. In preferred embodiments, as described herein, the priming sequences are universal priming sequences. This allows for highly multiplexed amplification. In a preferred embodiment at least one of the universal priming sequences is specific for a particular allele.

As shown in the figures, allele detection can proceed on a number of levels. In one embodiment adapters are distinct for the particular allele. Thus, following amplification of the adapter sequences, detection of the adapter provides identification of the particular allele to be detected.

Alternatively, allele detection proceeds as a result of allele specific amplification. As shown in FIG. **1**, at least one of the priming sequences on the primers for each allele is specific for a particular allele. Thus, following the specificity assay, one of the alleles will be identified. Following addition of the respective amplification primers, only one set of the primers will hybridize with the priming sequences. Thus, only one of the sets of primers will generate an amplicon. In a preferred embodiment, each of the sets of primers is labeled with distinct label. Because only one of the sets will be amplified, detection of a label provides an indication of the primer that was amplified. This, in turn identifies the nucleotide at the detection position.

In an alternative embodiment the target nucleic acid is first contacted with a first target specific probe under stringent annealing conditions and a first extension reaction is performed with either dNTPs or ddNTPS forming a first extension product (see FIG. **2**). The first target specific probe in this embodiment is either a locus specific probe or an allele specific probe. This step reduces the complexity of the sample. Subsequently the first extension product is contacted with a second probe that has the same sequence as a portion of the target sequence, i.e. the second probe is complementary to the extension product, and again can be either an allele specific probe or a locus specific probe. Following hybridization of the second probe, a second extension reaction is performed.

In a preferred embodiment the primers for the first and second extension reaction also include amplification priming sites. Preferably the amplification priming sites are universal priming sites as described herein. Accordingly, the resulting extension product is amplified (the amplification component of the multiplexing scheme). The resulting double stranded product is then denatured and either of the strands is used as a template for a single base extension (SBE) reaction as described in more detail below (the specificity component). In the SBE reaction, chain terminating nucleotides such as ddTNPs are used as substrates for the polymerase and are incorporated into a target probe that is hybridized to the single stranded amplicon template adjacent to the interrogation position. Preferably the ddNTPs are labeled as described below. Preferably, the ddNTPs are discretely labeled such that they can be discriminated in the detection step.

In an alternative embodiment a first biotinylated or otherwise tagged probe is hybridized with a target nucleic acid and a first extension reaction is performed. The primer or probe is either an allele specific or locus specific probe. The extended product is then purified from the mixture by the tag. Again, this serves as the complexity reduction step. Subsequently, a second primer is hybridized to the first extension product and a second extension reaction is performed, preferably in an allele specific manner, i.e. with discriminatory probes that are specific for each allele. This represents the specificity step. Preferably, both of the primers used in the extension reactions contain universal priming sites. Thus, universal primers can be added for universal amplification of the extension products (the amplification component) (see FIG. **3**). In a preferred embodiment, each allele specific primer includes a distinct amplification priming site. Thus, following allele discrimination, only one of the primers can be used for amplification,

resulting in allele specific amplification. Preferably the amplification primers contain discrete labels, which again allows for detection of which particular primers served as amplification templates. This, again, identifies the particular allele to be detected. In an additional preferred embodiment, at least one of the primers includes an adapter sequence as outlined below.

In an alternative embodiment tagged, i.e. biotinylated, primers are hybridized with a target nucleic acid. Preferably the hybridization complex is immobilized. Either the target or the primer can be the immobilized component. After annealing, the immobilized complexes are washed to remove unbound nucleic acids. This is followed by an extension reaction. This is the complexity reduction component of the assay. Subsequently, the extended probe is removed via the purification tag. The purified probe is then hybridized with allele specific probes (the specificity component). The hybridized probes are then amplified (the amplification component) (see FIG. **4**).

In a preferred embodiment the allele specific probe contains universal priming sites and an adapter sequence. Preferably the universal priming sites are specific for a particular allele. That is, one of the universal priming sites may be common to all alleles, but the second universal priming site is specific for a particular allele. Following hybridization the allele specific primer, the complexes are washed to remove unbound or mismatched primers. Thus, this configuration allows for allele specific amplification. Amplicons are detected as an indication of the presence of a particular allele.

In an alternative embodiment, the specificity component occurs first, In this embodiment allele specific probes are hybridized with the target nucleic acid; an extension assay is performed whereby only the perfectly complementary probe is extended. That is, only the probe that is perfectly complementary to the probe at the interrogation position serves as a substrate for extension reaction. Preferably the extension reaction includes tagged, i.e. biotinylated, dNTPs such that the extension product is tagged. The extension product is then purified from the reaction mixture. Subsequently, a second allele specific primer is hybridized to the extension product. This step also serves as a second specificity step. In this embodiment the specificity steps also serve as complexity reduction components in that they enrich for target nucleic acids. Following the addition of the second allele specific primer and extension, the extension product is amplified, preferably with universal primers (see FIG. **5**).

as discussed previously, it is preferably for the at least one allele specific primer to contain an allele specific priming site, preferably an allele specific universal priming site. Again, this configuration allows for multiplexed allele specific amplification using universal primers.

In an alternative embodiment, the target nucleic acid is first immobilized and hybridized with allele specific primers. Preferably the allele specific primers also include an adapter sequence that is indicative of the particular allele. Allele specific extension is then performed whereby only the primer that is perfectly complementary to the detection position of the target nucleic acid will serve as a template for primer extension. That is, mismatched primers will not be extended. Of note, the allele specific position of the primer need not be the 3' terminal nucleotide of the primer (see FIGS. **7** and **8**). That is, the primer may extend beyond the detection position of the target nucleic acid. In this embodiment it is preferable to include labeled dNTPs or ddNTPs or both such that the extension product is labeled and can be detected.

In a preferred embodiment both dNTPs and ddNTPs are included in the extension reaction mixture. In this embodiment only one label is needed, and the amount of label can be determined and altered by varying the relative concentration of labeled and unlabeled dTNPs and ddNTPs. That is, in one embodiment labeled ddNTPs are included in the extension mix at a dilution such that each termination will result in placement of single label on each strand. Thus, this method allows for quantification of targets. Alternatively, if a higher signal is needed, a mixture of labeled dNTPs can be used along with chain terminating nucleotides at a lower concentration. The result is the incorporation of multiple labels per extension product. Preferably the primers also include adapters which facilitate immobilization of the extension products for detection.

In an additional preferred configuration, target nucleic acids are hybridized with tagged locus specific primers. Preferably the primer includes a locus specific portion and a universal priming site. Of note, as is generally true for locus specific primers, they need not be immediately adjacent to the detection position. Upon hybridization, the hybridization complexes are immobilized, preferably by binding moiety that specifically binds the tag on the locus specific primer. The immobilized complexes are then washed to remove unlabeled nucleic acids; the remaining hybridization complexes are then subject to an extension reaction. Following extension of the locus specific primer, a nucleotide complementary to the nucleotide at the detection position will be incorporated into the extension product. In some embodiments it is desirable to limit the size of the extension because this reduces the complexity of subsequent annealing steps. This may be accomplished by including both dNTPs and ddNTPs in the reaction mixture.

Following the first extension, a second locus or allele specific primer is hybridized to the immobilized extension product and a second extension reaction occurs. Preferably the second extension primer includes a target specific portion and a universal priming site. After extension, universal amplification primers can be added to the reaction and the extension products amplified. The amplicons can then be used for detection of the particular allele. This can be accomplished by competitive hybridization, as described herein. Alternatively, it can be accomplished by an additional extension reaction. When the extension reaction is performed, preferably a primer that contains an adapter sequence and a target specific portion is hybridized with the amplicons. Preferably the target specific portion hybridizes up to a position that is adjacent to the detection position, i.e. the particular allele to be detected. Polymerase and labeled ddNTPs are then added and the extension reaction proceeds, whereby incorporation of a particular label is indicative of the nucleotide that is incorporated into the extension primer. This nucleotide is complementary to the nucleotide at the detection position. Thus, analyzing or detecting which nucleotide is incorporated into the primer provides an indication of the nucleotide at the allele position. The extended primer is detected by methods that include but are not limited to the methods described herein.

In another embodiment, the genotyping specificity is conferred by the extension reaction. In this embodiment, two probes (sometimes referred to herein as "primers") are hybridized non-contiguously to a target sequence comprising, from 3' to 5', a first second and third target domain. Preferably the target is immobilized. That is, in a preferred embodiment, the target sequence is genomic DNA and is attached to a solid support as is generally described in U.S. Ser. No. 09/931,285, hereby expressly incorporated by reference in its entirety. In this embodiment, magnetic beads, tubes

or microtiter plates are particularly preferred solid supports, although other solid supports as described below can also be used.

The first probe hybridized to the first domain, contains a first universal priming sequence and contains, at the 3' end (within the terminal six bases), an interrogation position. Subsequently, the unhybridized primers are removed. This is followed by providing an extension enzyme such as a polymerase, and NTPs (which includes both dNTPs, NTPs and analogs, as outlined below). If the interrogation position is perfectly complementary to the detection position of the target sequence, the extension enzyme will extend through the second target domain to form an extended first probe, ending at the beginning of the third domain, to which the second probe is hybridized. A second probe is complementary to the third target domain, and upon addition of a ligase, the extended first probe will ligate to the second probe. The addition of a primer allows amplification to form amplicons. If the second probe comprises an antisense second primer, exponential amplification may occur, such as in PCR. Similarly, one or other of the probes may comprise an adapter or address sequence, which facilitates detection. For example, the adapter may serve to allow hybridization to a "universal array". Alternatively, the adapter may serve as a mobility modifier for electrophoresis or mass spectrometry analysis, or as a label sequence for the attachment of labels or beads for flow cytometry analysis.

In another embodiment, the reaction is similar except that it is the ligation reaction that provides the detection position/interrogation specificity. In this embodiment, it is the second probe that comprises a 5' interrogation position. The extended first probe will not be ligated to the second probe if there is a mismatch between the interrogation position and the target sequence. As above, the addition of a primer allows amplification to form amplicons. If the second probe comprises an antisense second primer, exponential amplification may occur, such as in PCR. Similarly, one or other of the probes may comprise an adapter or address sequence, which facilitates detection. For example, the adapter may serve to allow hybridization to a "universal array". Alternatively, the adapter may serve as a mobility modifier for electrophoresis or mass spectrometry analysis, or as a label sequence for the attachment of labels or beads for flow cytometry analysis.

Once prepared, and attached to a solid support as required, the target sequence is used in genotyping reactions. It should be noted that while the discussion below focuses on certain assays, in general, for each reaction, each of these techniques may be used in a solution based assay, wherein the reaction is done in solution and a reaction product is bound to the array for subsequent detection, or in solid phase assays, where the reaction occurs on the surface and is detected, either on the same surface or a different one.

The assay continues with the addition of a first probe. The first probe comprises, a 5' first domain comprising a first universal priming sequence. The universal priming sites are used to amplify the modified probes to form a plurality of amplicons that are then detected in a variety of ways, as outlined herein. In preferred embodiments, one of the universal priming sites is a T7 site, such that RNA is ultimately made to form the amplicon. Alternatively, as more fully outlined below, two universal priming sequences are used, one on the second probe generally in antisense orientation, such that PCR reactions or other exponential amplification reactions can be done. Alternatively, a single universal primer can be used for amplification. Linear amplification can be performed using the SPIA assay, T7 amplification, linear TMA and the like, as described herein.

The first probe further comprises, 3' to the priming sequence, a second domain comprising a sequence substantially complementary to the first target domain of the target sequence. Again, the second target domain comprises n nucleotides, wherein n is an integer of at least 1, and preferably from 1 to 100s, with from 1 to 10 being preferred and from 1, 2, 3, 4 and 5 being particularly preferred. What is important is that the first and third target domains are non-contiguous, e.g. not adjacent.

In a preferred embodiment, the first probe, further comprises, 3' to the second domain, an interrogation position within the 3' six terminal bases. As used herein, the base which basepairs with a detection position base in a hybrid is termed a "readout position" or an "interrogation position"; thus one or the other of the first or second probes of the invention comprise an interrogation position, as outlined herein. In some cases, when two SNP positions or detection positions are being elucidated, both the first and the second probes may comprise interrogation positions.

When the first probe comprises the interrogation position, it falls within the six 3' terminal nucleotides, with within three, and preferably two, and most preferably it is the 3' terminal nucleotide. Alternatively, the first probe does not contain the interrogation position; rather the second probe does. This depends on whether the extension enzyme or the ligation enzyme is to confer the specificity required for the genotyping reaction.

In addition to the first probes of the invention, the compositions of the invention further comprise a second probe for each target sequence. The second probes each comprise a first domain comprising a sequence substantially complementary to the third target domain of a target sequence as outlined herein.

In some embodiments, the second probes comprise a second universal priming site. As outlined herein, the first and second probes can comprise two universal primers, one in each orientation, for use in PCR reactions or other amplification reactions utilizing two primers. That is, as is known in the art, the orientation of primers is such to allow exponential amplification, such that the first universal priming sequence is in the "sense" orientation and the second universal priming sequence is in the "antisense" orientation.

In a preferred embodiment, it is the second probe that comprises the interrogation position. In this embodiment, the second probe comprises a 5' interrogation nucleotide, although in some instances, depending on the ligase, the interrogation nucleotide may be within 1-3 bases of the 5' terminus. However, it is preferred that the interrogation base be the 5' base.

In a preferred embodiment, either the first or second probe further comprises an adapter sequence, (sometimes referred to in the art as "zip codes") to allow the use of "universal arrays". That is, arrays are generated that contain capture probes that are not target specific, but rather specific to individual artificial adapter sequences.

It should be noted that when two universal priming sequences and an adapter is used, the orientation of the construct should be such that the adapter gets amplified; that is, the two universal priming sequences are generally at the termini of the amplification template, described below.

The first and second probes are added to the target sequences to form a first hybridization complexes. The first hybridization complexes are contacted with a first universal primer that hybridizes to the first universal priming sequence, an extension enzyme and dNTPs.

If it is the first probe that comprises the interrogation nucleotide, of the base at the interrogation position is perfectly

complementary with the base at the detection position, extension of the first primer occurs through the second target domain, stopping at the 5' of the second probe, to form extended first probes that are hybridized to the target sequence, forming second hybridization complexes. If, however, the base at the interrogation position is not perfectly complementary with the base at the detection position, extension of the first probe will not occur, and no subsequent amplification or detection will occur.

Extension of the enzyme will also occur if it is the second probe that comprises the interrogation position.

Once extended, the extended first probe is adjacent to the 5' end of the second probe. In the case where the interrogation position was in the first probe, the two ends of the probes (the 3' end of the first probe and the 5' end of the second probe) are respectively perfectly complementary to the target sequence at these positions, and the two probes can be ligated together with a suitable ligase to form amplification templates.

The conditions for carrying out the ligation will depend on the particular ligase used and will generally follow the manufacturer's recommendations.

If, however, it is the second probe that carries the interrogation position at its 5' end, the base at the interrogation position must be perfectly complementary to the detection position in the target sequence to allow ligation. In the absence of perfect complementarity, no significant ligation will occur between the extended first probe and the second probe.

It should be noted that the enzymes may be added sequentially or simultaneously. If the target sequences are attached to a solid support, washing steps may also be incorporated if required.

The ligation of the extended first probe and the second probe results in an amplification template comprising at least one, and preferably two, universal primers and an optional adapter. Amplification can then be done, in a wide variety of ways. As will be appreciated by those in the art, there are a wide variety of suitable amplification techniques requiring either one or two primers, as is generally outlined in U.S. Ser. No. 09/517,945, hereby expressly incorporated by reference.

Detection Systems

All of the methods and compositions herein are drawn to methods of detecting, quantifying and/or determining the base at the detection position of a target nucleic acid, generally by having differential reactions occur depending on the presence or absence of a mismatch. The reaction products are generally detected on arrays as is outlined herein, although a number of different detection methods may be used.

As is more fully outlined below, preferred systems of the invention work as follows. An amplicon is attached (via hybridization) to an array site. This attachment is generally a direct hybridization between a adapter on the amplicon and a corresponding capture probe, although in some instances, the system can rely on indirect "sandwich" complexes using capture extender probes as are known in the art. In a preferred embodiment, the target sequence (e.g. the amplicon) itself comprises the labels. Alternatively, a label probe is added, that will hybridize to a label sequence on the amplicon, forming an assay complex. The capture probes of the array are substantially (and preferably perfectly) complementary to the adapter sequences.

The terms length determination, separation-by-length assay, and separation-by-length assay medium are taken collectively to mean a process and its related apparatus that achieves separation of DNA fragments on the basis of length, size, mass, or any other physical property. This includes generally, liquid chromatography, electrophoresis and direct mass spectrometry; more particularly, high performance liquid chromatography (HPLC) and capillary electrophoresis or gel electrophoresis, and MALDI-TOF MS respectively.

Other detection assays or formats include classical configurations such as the "dot-blot". This method of hybridization gained wide-spread use, and many versions were developed (see M. L. M. Anderson and B. D. Young, in Nucleic Acid Hybridization—A Practical Approach, B. D. Hames and S. J. Higgins, Eds., IRL Press, Washington D.C., Chapter 4, pp. 73-111, 1985). The "dot blot" hybridization has been further developed for multiple analysis of genomic mutations (D. Nanibhushan and D. Rabin, in EPA 0228075, Jul. 8, 1987) and for the detection of overlapping clones and the construction of genomic maps (G. A. Evans, in U.S. Pat. No. 5,219, 726, Jun. 15, 1993).

Another format, the so-called "sandwich" hybridization, involves attaching oligonucleotide probes covalently to a solid support and using them to capture and detect multiple nucleic acid targets. (M. Ranki et al., Gene, 21, pp. 77-85, 1983; A. M. Palva, T. M. Ranki, and H. E. Soderlund, in UK Patent Application GB 2156074A, Oct. 2, 1985; T. M. Ranki and H. E. Soderlund in U.S. Pat. No. 4,563,419, Jan. 7, 1986; A. D. B. Malcolm and J. A. Langdale, in PCT WO 86/03782, Jul. 3, 1986; Y. Stabinsky, in U.S. Pat. No. 4,751,177, Jan. 14, 1988; T. H. Adams et al., in PCT WO 90/01564, Feb. 22, 1990; R. B. Wallace et al. 6 Nucleic Acid Res. 11, p. 3543, 1979; and B. J. Connor et al., 80 Proc. Natl. Acad. Sci. USA pp. 278-282, 1983). Multiplex versions of these formats are called "reverse dot blots".

In another approach of matrix hybridization, Beattie et al., in The 1992 San Diego Conference: Genetic Recognition, November, 1992, used a microrobotic system to deposit micro-droplets containing specific DNA sequences into individual microfabricated sample wells on a glass substrate. The hybridization in each sample well is detected by interrogating miniature electrode test fixtures, which surround each individual microwell with an alternating current (AC) electric field.

One preferred aspect of the present invention is that it results in high-throughput screening capabilities. In the assays described herein, from a few up to millions of different tags identifying, e.g., SNPs, can be identified simultaneously. For example, using simple dot-blot hybridization methods, membranes with thousands of immobilized probes can be generated for screening against tags. The solid-phase techniques described below can be adapted to having literally millions of different immobilized nucleic acids per square inch. Similarly, very large sets of amplified DNAs, e.g,. tags, can be immobilized on membranes for simultaneous screening against one or more sequence.

In one embodiment, the identity of the amplification products are determined by detecting the molecular weights of the amplification product or a fragment thereof, such as by chromatography or mass spectroscopy.

For instance, the gross molecular weight of an amplification product or a discrete fragment thereof can be detected. As set forth above, each member of a probe library (i.e., all of the probes in the reaction) has a unique molecular weight label based on the particular sequence of the tag. For instance, mass spectrometry can provide high detection sensitivity and accuracy of mass measurements that can discern between probes which, while identical in length, differ in sequence by only base. Thus, complex libraries can be constructed by calculating the overall molecular weight of each amplification product to be detected by varying the G/C/A/T content in the tag sequence. In certain preferred embodiments, the nucleic acid sequence which is being detected includes, as its only vari-

able sequence, the tag sequence and not the template homology regions. Such fragments can be generated, for example, by including restriction sites that flank the tag sequence, or choosing the PCR primers such that only the tag sequence is the only variable region of the covalently closed circular product which is included in the amplification products. That being said, in those embodiments where the amplification product which is being detected also includes the template homology region(s), the calculation and design of the tag sequences will need to include the variability in the THRs as well in order to produce products having a unique molecular weight so as to be discernable from one another by mass spectroscopy or other detection means as may be chosen.

Those skilled in the art will recognize that very simple algorithms can be used to calculate the molecular weights for each member of a library by varying the sequence of the tag, taking into account if necessary the sequences of the template homology regions. The molecular weight complexity of the tag can be increased by allowing the probes to vary in length as well sequence.

In certain instances, the library can be deconvoluted by chromatographic techniques prior to detection by mass spectroscopy. For example, prior to introducing a sample into the spectrometer, the mixture can first be at least semi-purified. Separation procedures based on size (e.g. gel-filtration), solubility (e.g. isoelectric precipitation) or electric charge (e.g. electrophoresis, isoelectric focusing, ion exchange chromatography) may be used to separate a mixture of amplimers. A preferred separation procedure is high performance liquid chromatography (HPLC).

In certain embodiments, the amplification product can include an integrated mass label for multiplex sequencing. Multiplexing by mass modification in this case is obtained by mass-modifying the nucleic acid primer, e.g., at the level of the sugar or base moiety. Such embodiments are most practical when amplification products are to be mixed for detection after the amplification step rather than before.

Suitable mass spectrometry techniques for use in the present invention include DNA analyses of the present invention include collision-induced dissociation (CID) fragmentation analysis (e.g., CID in conjunction with a MS/MS configuration, see Schram, K. (1990) "Mass Spectrometry of Nucleic Acid Components," in Biomedical Applications of Mass Spectrometry 34:203-287; and Crain P. (1990) Mass Spectrometry Reviews 9:505-554); fast atomic bombardment (FAB mass spectrometry) and plasma desorption (PD mass spectrometry), see Koster et al. (1987) Biomedical Environmental Mass Spectrometry 14:111-116; and electrospray/ionspray (ES) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (see Fenn et al. (1984) J. Phys. Chem. 88:4451-4459, Smith et al. (1990) Anal. Chem. 62:882-889, and Ardrey, B. (1992) Spectroscopy Europe 4:10-18). MALDI mass spectrometry is particularly well suited to such analyses when a time-of-flight (TOF) configuration is used as a mass analyzer (MALDI-TOF). See International Publication No. WO 97/33000, published Sep. 12, 1997, see also Huth-Fehre et al. (1992) Rapid Communications in Mass Spectrometry 6:209-213, and Williams et al. (1990) Rapid Communications in Mass Spectrometry 4:348-351.

Suitable mass spectrometry techniques for use in the mass tag analyses of the present invention include collision-induced dissociation (CID) fragmentation analysis (e.g., CID in conjunction with a MS/MS configuration, see Schram, K. (1990) "Mass Spectrometry of Nucleic Acid Components," in Biomedical Applications of Mass Spectrometry 34:203-287; and Crain P. (1990) Mass Spectrometry Reviews 9:505-554); fast atomic bombardment (FAB mass spectrometry) and plasma desorption (PD mass spectrometry), see Koster et al. (1987 Biomedical Environmental Mass Spectrometry 14:111-116; and electrospray/ionspray (ES) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (see Fenn et al. (1984) J. Phys. Chem. 88:4451-4459, Smith et al. (1990) Anal. Chem. 62:882-889, and Ardrey, B. (1992) Spectroscopy Europe 4:10-18). MALDI mass spectrometry is particularly well suited to such analyses when a time-of-flight(TOF) configuration is used as a mass analyzer (MALDI-TOF). See International Publication No. WO 97/33000, published Sep. 12, 1997, see also Huth-Fehre et al. (1992) Rapid Communications in Mass Spectrometry 6:209-213, and Williams et al. (1990) Rapid Communications in Mass Spectrometry 4:348-351.

In this regard, a number of mass tags suitable for use with nucleic acids are known (see U.S. Pat. No. 5,003,059 to Brennan and U.S. Pat. No. 5,547,835 to Koster), including mass tags which are cleavable from the nucleic acid (see International Publication No. WO 97/27331).

In another embodiment, the hybridization tags are detected on a micro-formatted multiplex or matrix devices (e.g., DNA chips) (see M. Barinaga, 253 Science, pp. 1489, 1991; W. Bains, 10 Bio/Technology, pp. 757-758, 1992). These methods usually attach specific DNA sequences to very small specific areas of a solid support, such as micro-wells of a DNA chip. In one variant, the invention is adapted to solid phase arrays for the rapid and specific detection of multiple polymorphic nucleotides, e.g., SNPs. Typically, an oligonucleotide is linked to a solid support and a tag nucleic acid is hybridized to the oligonucleotide. Either the oligonucleotide, or the tag, or both, can be labeled, typically with a fluorophore. Where the tag is labeled, hybridization is detected by detecting bound fluorescence. Where the oligonucleotide is labeled, hybridization is typically detected by quenching of the label. Where both the oligonucleotide and the tag are labeled, detection of hybridization is typically performed by monitoring a color shift resulting from proximity of the two bound labels. A variety of labeling strategies, labels, and the like, particularly for fluorescent based applications are described, supra.

In one embodiment, an array of oligonucleotides are synthesized on a solid support. Exemplary solid supports include glass, plastics, polymers, metals, metalloids, ceramics, organics, etc. Using chip masking technologies and photoprotective chemistry it is possible to generate ordered arrays of nucleic acid probes. These arrays, which are known, e.g., as "DNA chips," or as very large scale immobilized polymer arrays ("VLSIPS TM" arrays) can include millions of defined probe regions on a substrate having an area of about 1 cm2 to several cm2, thereby incorporating sets of from a few to millions of probes.

The construction and use of solid phase nucleic acid arrays to detect target nucleic acids is well described in the literature. See, Fodor et al. (1991) Science, 251: 767-777; Sheldon et al. (1993) Clinical Chemistry 39(4): 718-719; Kozal et al. (1996) Nature Medicine 2(7): 753-759 and Hubbell U.S. Pat. No. 5,571,639. See also, Pinkel et al. PCT/US95/16155 (WO 96/17958). In brief, a combinatorial strategy allows for the synthesis of arrays containing a large number of probes using a minimal number of synthetic steps. For instance, it is possible to synthesize and attach all possible DNA 8 mer oligonucleotides (48, or 65,536 possible combinations) using only 32 chemical synthetic steps. In general, VLSIPS TM procedures provide a method of producing 4n different oligonucleotide probes on an array using only 4 n synthetic steps.

Light-directed combinatorial synthesis of oligonucleotide arrays on a glass surface is performed with automated phosphoramidite chemistry and chip masking techniques similar to photoresist technologies in the computer chip industry. Typically, a glass surface is derivatized with a silane reagent containing a functional group, e.g., a hydroxyl or amine group blocked by a photolabile protecting group. Photolysis through a photolithogaphic mask is used selectively to expose functional groups which are then ready to react with incoming 5'-photoprotected nucleoside phosphoramidites. The phosphoramidites react only with those sites which are illuminated (and thus exposed by removal of the photolabile blocking group). Thus, the phosphoramidites only add to those areas selectively exposed from the preceding step. These steps are repeated until the desired array of sequences have been synthesized on the solid surface.

A 96 well automated multiplex oligonucleotide synthesizer (A.M.O.S.) has also been developed and is capable of making thousands of oligonucleotides (Lashkari et al. (1995) PNAS 93: 7912). Existing light-directed synthesis technology can generate high-density arrays containing over 65,000 oligonucleotides (Lipshutz et al. (1995) BioTech. 19: 442.

Combinatorial synthesis of different oligonucleotide analogues at different locations on the array is determined by the pattern of illumination during synthesis and the order of addition of coupling reagents. Monitoring of hybridization of target nucleic acids to the array is typically performed with fluorescence microscopes or laser scanning microscopes. In addition to being able to design, build and use probe arrays using available techniques, one of skill is also able to order custom-made arrays and array-reading devices from manufacturers specializing in array manufacture. For example, Affymetrix Corp., in Santa Clara, Calif. manufactures DNA VLSIP TM arrays.

It will be appreciated that oligonucleotide design is influenced by the intended application. For example, where several oligonucleotide-tag interactions are to be detected in a single assay, e.g., on a single DNA chip, it is desirable to have similar melting temperatures for all of the probes. Accordingly, the length of the probes are adjusted so that the melting temperatures for all of the probes on the array are closely similar (it will be appreciated that different lengths for different probes may be needed to achieve a particular T[m] where different probes have different GC contents). Although melting temperature is a primary consideration in probe design, other factors are optionally used to further adjust probe construction, such as selecting against primer self-complementarity and the like. The "active" nature of the devices provide independent electronic control over all aspects of the hybridization reaction (or any other affinity reaction) occurring at each specific microlocation. These devices provide a new mechanism for affecting hybridization reactions which is called electronic stringency control (ESC). For DNA hybridization reactions which require different stringency conditions, ESC overcomes the inherent limitation of conventional array technologies. The active devices of this invention can electronically produce "different stringency conditions" at each microlocation. Thus, all hybridizations can be carried out optimally in the same bulk solution. These arrays are described in U.S. Pat. No. 6,051,380 by Sosnowski et al.

Accordingly, the present invention provides array compositions comprising at least a first substrate with a surface comprising individual sites. By "array" or "biochip" herein is meant a plurality of nucleic acids in an array format; the size of the array will depend on the composition and end use of the array. Nucleic acids arrays are known in the art, and can be classified in a number of ways; both ordered arrays (e.g. the ability to resolve chemistries at discrete sites), and random arrays (e.g. bead arrays) are included. Ordered arrays include, but are not limited to, those made using photolithography techniques (Affymetrix GeneChip™), spotting techniques (Synteni and others), printing techniques (Hewlett Packard and Rosetta), electrode arrays, three dimensional "gel pad" arrays, etc. Liquid arrays may also be used, i.e. three-dimensional array methods such as flow cytometry. When flow cytometry is the detection method, amplicons are immobilized to a support such as a microsphere as described herein. The microspheres are applied to a flow cytometer and the amplicons are detected optically as described herein.

In a preferred embodiment, when beads are used, the beads are distributed in or on an additional support or substrate is generally flat (planar), although as will be appreciated by those in the art, other configurations of substrates may be used as well; for example, three dimensional configurations can be used, for example by embedding the beads in a porous block of plastic that allows sample access to the beads and using a confocal microscope for detection. Similarly, the beads may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume. Preferred substrates include optical fiber bundles as discussed below, and flat planar substrates such as glass, polystyrene and other plastics and acrylics. In a preferred embodiment such substrates include multi-well plates as are known in the art. In a preferred embodiment magnetic force is used to immobilized magnetic beads on the solid support.

A preferred embodiment utilizes microspheres on a variety of array substrates including fiber optic bundles, as are outlined in PCTs US98/21193, PCT US99/14387 and PCT US98/05025; WO98/50782; and U.S. Ser. Nos. 09/287,573, 09/151,877, 09/256,943, 09/316,154, 60/119,323, 09/315,584; all of which are expressly incorporated by reference. While much of the discussion below is directed to the use of microsphere arrays on fiber optic bundles, any array format of nucleic acids on solid supports may be utilized.

Arrays containing from about 2 different bioactive agents (e.g. different beads, when beads are used) to many millions can be made, with very large arrays being possible. Generally, the array will comprise from two to as many as a billion or more, depending on the size of the beads and the array substrate, as well as the end use of the array, thus very high density, high density, moderate density, low density and very low density arrays may be made. Preferred ranges for very high density arrays are from about 10,000,000 to about 2,000,000,000, with from about 100,000,000 to about 1,000,000,000 being preferred (all numbers being in square cm). High density arrays range about 100,000 to about 10,000,000, with from about 1,000,000 to about 5,000,000 being particularly preferred. Moderate density arrays range from about 10,000 to about 100,000 being particularly preferred, and from about 20,000 to about 50,000 being especially preferred. Low density arrays are generally less than 10,000, with from about 1,000 to about 5,000 being preferred. Very low density arrays are less than 1,000, with from about 10 to about 1000 being preferred, and from about 100 to about 500 being particularly preferred. In some embodiments, the compositions of the invention may not be in array format; that is, for some embodiments, compositions comprising a single bioactive agent may be made as well. In addition, in some arrays, multiple array substrates may be used, either of different or identical compositions. Thus for example, large arrays may comprise a plurality of smaller array substrates.

In addition, one advantage of the present compositions is that particularly through the use of fiber optic technology, extremely high density arrays can be made. Thus for example,

because beads of 200 μm or less (with beads of 200 nm possible) can be used, and very small fibers are known, it is possible to have as many as 40,000 or more (in some instances, 1 million) different elements (e.g. fibers and beads) in a 1 mm^2 fiber optic bundle, with densities of greater than 25,000,000 individual beads and fibers (again, in some instances as many as 50-100 million) per 0.5 cm^2 obtainable (4 million per square cm for 5μ center-to-center and 100 million per square cm for 1μ center-to-center).

By "array substrate" or "array solid support" or other grammatical equivalents herein is meant any material that can be modified to contain discrete individual sites appropriate for the attachment or association of beads and is amenable to at least one detection method. As will be appreciated by those in the art, the number of possible array substrates is very large. Possible array substrates include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In general, the array substrates allow optical detection and do not themselves appreciably fluoresce.

Generally the array substrate is flat (planar), although as will be appreciated by those in the art, other configurations of array substrates may be used as well; for example, three dimensional configurations can be used, for example by embedding the beads in a porous block of plastic that allows sample access to the beads and using a confocal microscope for detection. Similarly, the beads may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume. Preferred array substrates include optical fiber bundles as discussed below, and flat planar array substrates such as paper, glass, polystyrene and other plastics and acrylics.

In a preferred embodiment, the array substrate is an optical fiber bundle or array, as is generally described in U.S. Ser. Nos. 08/944,850 and 08/519,062, PCT US98/05025, and PCT US98/09163, all of which are expressly incorporated herein by reference. Preferred embodiments utilize preformed unitary fiber optic arrays. By "preformed unitary fiber optic array" herein is meant an array of discrete individual fiber optic strands that are co-axially disposed and joined along their lengths. The fiber strands are generally individually clad. However, one thing that distinguished a preformed unitary array from other fiber optic formats is that the fibers are not individually physically manipulatable; that is one strand generally cannot be physically separated at any point along its length from another fiber strand.

Generally, the arrayed array compositions of the invention can be configured in several ways; see for example U.S. Ser. No. 09/473,904, hereby expressly incorporated by reference. In a preferred embodiment, as is more fully outlined below, a "one component" system is used. That is, a first array substrate comprising a plurality of assay locations (sometimes also referred to herein as "assay wells"), such as a microtiter plate, is configured such that each assay location contains an individual array. That is, the assay location and the array location are the same. For example, the plastic material of the microtiter plate can be formed to contain a plurality of "bead wells" in the bottom of each of the assay wells. Beads containing the capture probes of the invention can then be loaded into the bead wells in each assay location as is more fully described below.

Alternatively, a "two component" system can be used. In this embodiment, the individual arrays are formed on a second array substrate, which then can be fitted or "dipped" into the first microtiter plate substrate. A preferred embodiment utilizes fiber optic bundles as the individual arrays, generally with "bead wells" etched into one surface of each individual fiber, such that the beads containing the capture probes are loaded onto the end of the fiber optic bundle. The composite array thus comprises a number of individual arrays that are configured to fit within the wells of a microtiter plate.

By "composite array" or "combination array" or grammatical equivalents herein is meant a plurality of individual arrays, as outlined above. Generally the number of individual arrays is set by the size of the microtiter plate used; thus, 96 well, 384 well and 1536 well microtiter plates utilize composite arrays comprising 96, 384 and 1536 individual arrays, although as will be appreciated by those in the art, not each microtiter well need contain an individual array. It should be noted that the composite arrays can comprise individual arrays that are identical, similar or different. That is, in some embodiments, it may be desirable to do the same 2,000 assays on 96 different samples; alternatively, doing 192,000 experiments on the same sample (i.e. the same sample in each of the 96 wells) may be desirable. Alternatively, each row or column of the composite array could be the same, for redundancy/quality control. As will be appreciated by those in the art, there are a variety of ways to configure the system. In addition, the random nature of the arrays may mean that the same population of beads may be added to two different array surfaces, resulting in substantially similar but perhaps not identical arrays.

At least one surface of the array substrate is modified to contain discrete, individual sites for later association of microspheres. These sites may comprise physically altered sites, i.e. physical configurations such as wells or small depressions in the array substrate that can retain the beads, such that a microsphere can rest in the well, or the use of other forces (magnetic or compressive), or chemically altered or active sites, such as chemically functionalized sites, electrostatically altered sites, hydrophobically/hydrophilically functionalized sites, spots of adhesive, etc.

The sites may be a pattern, i.e. a regular design or configuration, or randomly distributed. A preferred embodiment utilizes a regular pattern of sites such that the sites may be addressed in the X-Y coordinate plane. "Pattern" in this sense includes a repeating unit cell, preferably one that allows a high density of beads on the array substrate. However, it should be noted that these sites may not be discrete sites. That is, it is possible to use a uniform surface of adhesive or chemical functionalities, for example, that allows the attachment of beads at any position. That is, the surface of the array substrate is modified to allow attachment of the microspheres at individual sites, whether or not those sites are contiguous or non-contiguous with other sites. Thus, the surface of the array substrate may be modified such that discrete sites are formed that can only have a single associated bead, or alternatively, the surface of the array substrate is modified and beads may go down anywhere, but they end up at discrete sites. That is, while beads need not occupy each site on the array, no more than one bead occupies each site.

In a preferred embodiment, the surface of the array substrate is modified to contain wells, i.e. depressions in the surface of the array substrate. This may be done as is generally known in the art using a variety of techniques, including, but not limited to, photolithography, stamping techniques, molding techniques and microetching techniques. As will be

appreciated by those in the art, the technique used will depend on the composition and shape of the array substrate.

In a preferred embodiment, physical alterations are made in a surface of the array substrate to produce the sites. In a preferred embodiment, the array substrate is a fiber optic bundle and the surface of the array substrate is a terminal end of the fiber bundle, as is generally described in Ser. Nos. 08/818,199 and 09/151,877, both of which are hereby expressly incorporated by reference. In this embodiment, wells are made in a terminal or distal end of a fiber optic bundle comprising individual fibers. In this embodiment, the cores of the individual fibers are etched, with respect to the cladding, such that small wells or depressions are formed at one end of the fibers. The required depth of the wells will depend on the size of the beads to be added to the wells.

Generally in this embodiment, the microspheres are non-covalently associated in the wells, although the wells may additionally be chemically functionalized as is generally described below, cross-linking agents may be used, or a physical barrier may be used, i.e. a film or membrane over the beads.

In a preferred embodiment, the surface of the array substrate is modified to contain chemically modified sites, that can be used to attach, either covalently or non-covalently, the microspheres of the invention to the discrete sites or locations on the array substrate. "Chemically modified sites" in this context includes, but is not limited to, the addition of a pattern of chemical functional groups including amino groups, carboxy groups, oxo groups and thiol groups, that can be used to covalently attach microspheres, which generally also contain corresponding reactive functional groups; the addition of a pattern of adhesive that can be used to bind the microspheres (either by prior chemical functionalization for the addition of the adhesive or direct addition of the adhesive); the addition of a pattern of charged groups (similar to the chemical functionalities) for the electrostatic attachment of the microspheres, i.e. when the microspheres comprise charged groups opposite to the sites; the addition of a pattern of chemical functional groups that renders the sites differentially hydrophobic or hydrophilic, such that the addition of similarly hydrophobic or hydrophilic microspheres under suitable experimental conditions will result in association of the microspheres to the sites on the basis of hydroaffinity. For example, the use of hydrophobic sites with hydrophobic beads, in an aqueous system, drives the association of the beads preferentially onto the sites. As outlined above, "pattern" in this sense includes the use of a uniform treatment of the surface to allow attachment of the beads at discrete sites, as well as treatment of the surface resulting in discrete sites. As will be appreciated by those in the art, this may be accomplished in a variety of ways.

In some embodiments, the beads are not associated with an array substrate. That is, the beads are in solution or are not distributed on a patterned substrate.

In a preferred embodiment, the compositions of the invention further comprise a population of microspheres. By "population" herein is meant a plurality of beads as outlined above for arrays. Within the population are separate subpopulations, which can be a single microsphere or multiple identical microspheres. That is, in some embodiments, as is more fully outlined below, the array may contain only a single bead for each capture probe; preferred embodiments utilize a plurality of beads of each type.

By "microspheres" or "beads" or "particles" or grammatical equivalents herein is meant small discrete particles. The composition of the beads will vary, depending on the class of capture probe and the method of synthesis. Suitable bead compositions include those used in peptide, nucleic acid and organic moiety synthesis, including, but not limited to, plastics, ceramics, glass, polystyrene, methylstyrene, acrylic polymers, paramagnetic materials, thoria sol, carbon graphite, titanium dioxide, latex or cross-linked dextrans such as Sepharose, cellulose, nylon, cross-linked micelles and Teflon may all be used. "*Microsphere Detection Guide*" from Bangs Laboratories, Fishers Ind. is a helpful guide.

The beads need not be spherical; irregular particles may be used. In addition, the beads may be porous, thus increasing the surface area of the bead available for either capture probe attachment or tag attachment. The bead sizes range from nanometers, i.e. 100 nm, to millimeters, i.e. 1 mm, with beads from about 0.2 micron to about 200 microns being preferred, and from about 0.5 to about 5 micron being particularly preferred, although in some embodiments smaller beads may be used.

It should be noted that a key component of the invention is the use of an array substrate/bead pairing that allows the association or attachment of the beads at discrete sites on the surface of the array substrate, such that the beads do not move during the course of the assay.

Each microsphere comprises a capture probe, although as will be appreciated by those in the art, there may be some microspheres which do not contain a capture probe, depending on the synthetic methods.

Attachment of the nucleic acids may be done in a variety of ways, as will be appreciated by those in the art, including, but not limited to, chemical or affinity capture (for example, including the incorporation of derivatized nucleotides such as AminoLink or biotinylated nucleotides that can then be used to attach the nucleic acid to a surface, as well as affinity capture by hybridization), cross-linking, and electrostatic attachment, etc. In a preferred embodiment, affinity capture is used to attach the nucleic acids to the beads. For example, nucleic acids can be derivatized, for example with one member of a binding pair, and the beads derivatized with the other member of a binding pair. Suitable binding pairs are as described herein for IBL/DBL pairs. For example, the nucleic acids may be biotinylated (for example using enzymatic incorporate of biotinylated nucleotides, for by photoactivated cross-linking of biotin). Biotinylated nucleic acids can then be captured on streptavidin-coated beads, as is known in the art. Similarly, other hapten-receptor combinations can be used, such as digoxigenin and anti-digoxigenin antibodies. Alternatively, chemical groups can be added in the form of derivatized nucleotides, that can them be used to add the nucleic acid to the surface.

Preferred attachments are covalent, although even relatively weak interactions (i.e. non-covalent) can be sufficient to attach a nucleic acid to a surface, if there are multiple sites of attachment per each nucleic acid. Thus, for example, electrostatic interactions can be used for attachment, for example by having beads carrying the opposite charge to the bioactive agent.

Similarly, affinity capture utilizing hybridization can be used to attach nucleic acids to beads.

Alternatively, chemical crosslinking may be done, for example by photoactivated crosslinking of thymidine to reactive groups, as is known in the art.

In a preferred embodiment, each bead comprises a single type of capture probe, although a plurality of individual capture probes are preferably attached to each bead. Similarly, preferred embodiments utilize more than one microsphere containing a unique capture probe; that is, there is redundancy

built into the system by the use of subpopulations of microspheres, each microsphere in the subpopulation containing the same capture probe.

As will be appreciated by those in the art, the capture probes may either be synthesized directly on the beads, or they may be made and then attached after synthesis. In a preferred embodiment, linkers are used to attach the capture probes to the beads, to allow both good attachment, sufficient flexibility to allow good interaction with the target molecule, and to avoid undesirable binding reactions.

In a preferred embodiment, the capture probes are synthesized directly on the beads. As is known in the art, many classes of chemical compounds are currently synthesized on solid supports, such as peptides, organic moieties, and nucleic acids. It is a relatively straightforward matter to adjust the current synthetic techniques to use beads.

In a preferred embodiment, the capture probes are synthesized first, and then covalently attached to the beads. As will be appreciated by those in the art, this will be done depending on the composition of the capture probes and the beads. The functionalization of solid support surfaces such as certain polymers with chemically reactive groups such as thiols, amines, carboxyls, etc. is generally known in the art. Accordingly, "blank" microspheres may be used that have surface chemistries that facilitate the attachment of the desired functionality by the user. Some examples of these surface chemistries for blank microspheres include, but are not limited to, amino groups including aliphatic and aromatic amines, carboxylic acids, aldehydes, amides, chloromethyl groups, hydrazide, hydroxyl groups, sulfonates and sulfates.

In general, the methods of making the arrays and of decoding the arrays is done to maximize the number of different candidate agents that can be uniquely encoded. The compositions of the invention may be made in a variety of ways. In general, the arrays are made by adding a solution or slurry comprising the beads to a surface containing the sites for attachment of the beads. This may be done in a variety of buffers, including aqueous and organic solvents, and mixtures. The solvent can evaporate, and excess beads are removed.

In a preferred embodiment, when non-covalent methods are used to associate the beads with the array, a novel method of loading the beads onto the array is used. This method comprises exposing the array to a solution of particles (including microspheres and cells) and then applying energy, e.g. agitating or vibrating the mixture. In a preferred embodiment when the array substrate is a fiber optic bundle, the array substrate is tapped into the beads. That is, the energy is tapping. This results in an array comprising more tightly associated particles, as the agitation is done with sufficient energy to cause weakly-associated beads to fall off (or out, in the case of wells). These sites are then available to bind a different bead. In this way, beads that exhibit a high affinity for the sites are selected. Arrays made in this way have two main advantages as compared to a more static loading: first of all, a higher percentage of the sites can be filled easily, and secondly, the arrays thus loaded show a substantial decrease in bead loss during assays. Thus, in a preferred embodiment, these methods are used to generate arrays that have at least about 50% of the sites filled, with at least about 75% being preferred, and at least about 90% being particularly preferred. Similarly, arrays generated in this manner preferably lose less than about 20% of the beads during an assay, with less than about 10% being preferred and less than about 5% being particularly preferred.

Methods of adding, washing and detecting the amplicons on the array are well known.

Thus, the compositions of the present invention may be used in a variety of research, clinical, quality control, or field testing settings.

In a preferred embodiment, the present invention finds use in the quantification of PCR reactions. Thus, the invention provides a method for quantifying the number of one or more specific sequences in a sample of nucleic acids. The method may be similar to any of the methods described above, so long as the product being detected is present in proportions that are directly correlated with the amount of original template sequence. This is the case, e.g., where the method involves a hybridization step to the template DNA, circularization of the probe, extension of the primers and detection of the extension product. In a preferred embodiment, the method further comprises an amplification step, wherein the amplification reaction is a controlled amplification. This is the case, e.g., when using PCR amplification and stopping the PCR reaction during the exponential phase. The amount of amplified product in this situation will be directly proportional to the amount of original sequence in the nucleic acid sample. Thus, in a preferred embodiment, several amplification reactions are conducted in parallel, using a different number of amplification cycles in each of them. This will assure that at least one of the reactions will have been stopped in the exponential phase.

In methods for quantifying the number of a specific sequence in a sample, it may also be desirable in certain situations to include a marker nucleic acid. The marker nucleic acid can be added to the reaction during the hybridization stage or at any stage thereafter and be subject or not to the same reactions. Alternatively, the marker DNA is used merely to determine the amount of amplied product at the end of the amplification step.

The methods for genotyping and those for quantifying can be used simultaneously, so long as the processes are controlled, such that the amount of amplified product is directly correlated to the amount of the original sequence in the sample nucleic acid.

All references cited herein are expressly incorporated by reference.

We claim:

1. A multiplex method for determining whether a sample contains at least 100 different target sequences, comprising:

a) providing a sample which may contain at least 100 different single-stranded target sequences attached to a first solid support;

b) contacting said target sequences with a probe set comprising more than 100 different single-stranded probes, wherein each of said more than 100 different probes comprises:

i) a first universal priming site, wherein each of said more than 100 different probes has identical universal priming sites, and

ii) a target specific domain, such that different double-stranded hybridization complexes are formed, each of the different hybridization complexes comprising one of said more than 100 different single-stranded probes and one of the different single-stranded target sequences from the sample;

c) removing unhybridized probes;

d) contacting said probes of the hybridization complexes with a first enzyme and forming different modified probes;

e) contacting said modified probes with:

i) at least a first primer that hybridizes to said universal priming site;

ii) NTPs; and

iii) an extension enzyme;

wherein said different modified probes are amplified and forming different amplicons;

f) immobilizing said different amplicons to a second solid support, and

g) detecting said different amplicons immobilized to said second solid support, thereby determining whether the sample contains at least 100 different target sequences.

2. The method of claim **1** wherein each of said different target sequences comprises a detection position, each of said probes comprises an interrogation position, and said first enzyme modifies said probes if there is substantial complementarity between the bases at said interrogation position and said detection position.

3. The method of claim **1**, wherein said second solid support comprises an array.

4. The method of claim **3**, wherein said array comprises 1,000 probes per square cm.

5. The method of claim **1** or **3**, wherein each of said different probes in step b) further comprises a unique adapter sequence, whereby the modified probes formed in step d) comprise said unique adapter sequence and said amplicons formed in step e) comprise said unique adapter sequence.

6. The method of claim **5**, wherein said second solid support comprises capture probes that are specific to said unique adapter sequence of said amplicons.

7. The method of claim **1** or **3**, wherein each of said different probes further comprises an adapter sequence that differs from said target specific domain and that is used to uniquely identify the target sequence.

8. The method of claim **7**, wherein said second solid support comprises capture probes that are specific to said adapter sequence.

9. The method of claim **1**, wherein step (b) further comprises hybridizing said target sequences with a plurality of other probes and step (d) further comprises ligating said other probes to said probes.

10. The method of claim **1** or **9**, wherein said target sequences are covalently attached to said first solid support.

11. The method of claim **9**, wherein each of said other probes comprises a second universal priming site.

12. The method of claim **1** or **9**, wherein step (d) further comprises forming extended probes by extending said probes with a polymerase and NTPs having a biotin tag that allows removal of said extended probes.

13. The method of claim **9**, wherein each of said target sequences comprises a detection position, each of said probes comprises an interrogation position, and said first enzyme modifies said probes if there is substantial complementarity between the bases at said interrogation position and said detection position.

14. The method of claim **1**, **9** or **13**, wherein said first solid support comprises a plurality of beads.

15. The method of claim **14**, wherein said second solid support comprises a population of beads comprising capture probes that are specific to said amplicons.

16. The method of claim **15**, wherein said second solid support further comprises a substrate comprising wells, wherein each of said wells has a single bead from said population of beads.

17. The method of claim **1**, **3**, **9** or **11**, wherein said second solid support comprises a population of beads comprising capture probes that are specific to said amplicons.

18. The method of claim **17**, wherein said second solid support further comprises a substrate comprising wells, wherein each of said wells has a single bead from said population of beads.

19. The method of claim **9**, **11** or **13**, wherein step (d) further comprises extending said other probes with a polymerase and forming extended probes and ligating said extended probes to said probes.

20. The method of claim **9**, **11** or **13**, wherein step (d) further comprises extending said probes with a polymerase and forming extended probes, and ligating said extended probes to said other probes.

21. The method of claim **9**, **11** or **13**, wherein said second solid support comprises a population of beads comprising capture probes that are specific to said amplicons.

22. The method of claim **21**, wherein said second solid support further comprises a substrate comprising wells, wherein each of said wells has a single bead from said population of beads.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,955,794 B2
APPLICATION NO. : 10/177727
DATED : June 7, 2011
INVENTOR(S) : Min-Jui Richard Shen et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page, Item (75), line 1, inventor name "Mun-Jui" should read --Min-Jui--.

Title page, Item (60), insert --Continuation-in-part of application No. 09/779,376, filed on Feb. 7, 2001, now abandoned, and a continuation-in-part of application No. 09/931,285, filed on Aug. 16, 2001, now Pat. No. 6,913,884, and a continuation-in-part of application No. 09/915,231, filed on Jul. 24, 2001, now Pat. No. 6,890,741.--.

In the Specifications

Column 1, lines 9-10, cancel the text "and WO 01/57269, filed Feb. 7, 2001".

Signed and Sealed this
Twenty-fifth Day of June, 2013

Teresa Stanek Rea
Acting Director of the United States Patent and Trademark Office

EXHIBIT 12

In The
United States District Court
District of Delaware

CORNELL UNIVERSITY, ET AL.
Plaintiffs

v.

ILLUMINA INC.
Defendant

C.A. 1:10-cv-00433-LPS-MPT

MONIB A. ZIRVI
IN SUPPORT OF CORNELL UNIVERSITY,
ET AL.

DECLARATION

I, Monib A Zirvi, declare that I graduated Summa Cum Laude from Princeton University in 1993 with a Degree in Electrical Engineering with special interest in Computer Engineering and a Certificate in Engineering Biology. I was then selected for the Medical Scientist Training Program at the Tri-Institutional M.D./Ph.D. Program at Cornell University Medical College/Memorial Sloan Kettering Cancer Center/Rockefeller University. In 1994, I met Dr. Francis Barany who is currently a Professor of Microbiology at Weill Medicine Cornell.

In February 1994, Dr. Barany had submitted a Program Project Grant Application to the National Cancer Institute entitled, "New Methods for Cancer Detection." In this grant application, detailed designs and concepts for a novel DNA array called the Universal Zipcode Array was laid out in detail. I became intrigued with the possibility of these new approaches to revolutionize the way DNA, the basic code of Life, was analyzed, sequenced and detected. I therefore decided to obtain his Ph.D. in the Laboratory of Dr. Francis Barany.

I graduated with my Ph.D. in 1999 and published an article on "Ligase-based detection of Mononucleotide Repeat Sequences." I then completed my M.D. Degree in 2000 and eventually completed a Dermatology Residency at the Hospital of the University of Pennsylvania. I have been a practicing Physician for years and I have been awarded a number of Awards including Castle Connolly's Top Doctor Award.

One of the initial research projects that I worked on in the laboratory of Dr. Francis Barany, was the synthesis of special artificial DNA sequences called Zipcodes. These consisted of sets of artificial (non-naturally occurring) sequences which were specifically created to decouple the process of hybridization from the detection of sequences in DNA analyses. Prior to this, the state of the art at the time involved synthetic DNA sequences arranged in arrays with the use of pure hybridization to discriminate between two very similar DNA fragments. This was the approach taken by the leader in the DNA array market at the time, Affymetrix. The designs and specifications and uses of Universal Zipcode Arrays were patented in the '917 and '470 Patent families.

Dr. Francis Barany's Grant Proposal in 1994 was reviewed by an NIH Study Section. One of the reviewers on the study section was the then Chief Technology Officer of Affymetrix, Dr. Stephen Fodor, who

took the ideas he learned about during the confidential review of Dr. Barany's Grant application and started using them at Affymetrix. Other employees at Affymetrix at the time included Dr. Mark Chee,

Dr. Jian-Bin Fan and Dr. Kevin Gunderson, all of whom eventually left Affymetrix to join Illumina Inc. when it was founded by Dr. David Walt of Tufts University. Other early employees at Illumina included Dr. John Stuelpnagel.

It is noteworthy to mention that in this present matter before the Court, that when Cornell requested to depose Dr. Mark Chee and Dr. Jian-Bin Fan, Illumina Inc. claimed that "they could not be located," and thus have not been deposed. There are numerous relevant patent applications and other documents submitted under their names, many of which have important unanswered fact disputes regarding inventorship. It has become apparent that Illumina may have misled the Court and Cornell when they stated they did not know where Jian-Bing Fan was. There have been papers and patents with him as a co-author filed with Illumina's knowledge since the start of this case. In addition, an Illumina employee, Xuyu Cai registered the AnchorDx.com website that Jian-Bing Fan used in 2015 to make it appear that he had fled. The head of Market Development for Illumina in China attended a meeting with Jian-Bing Fan in September 2016 in China. Please see Appendix for all of the relevant information regarding this potential Obstruction of Justice and Witness tampering matter.

As noted, Illumina. Inc. was founded in part by Dr. David Walt of Tufts University. Prior to 1995, Dr. Walt had been conducting research in use of fiberoptic bead-based microsensors for detection of volatile small molecules, but not DNA. In 1995, prior to starting Illumina, Inc., Dr. Walt had spoken with Dr. Francis Barany about Dr. Barany's lab research and these discussions included not only descriptions of the concept of a Universal Zipcode Array, but also potential uses of the Zipcode Array including detection of products of ligation reactions. At that time, PCR (polymerase chain reaction) was a well-known method of replicating many copies of a fragment of DNA. LDR, or Ligase Detection Reaction, involved using a thermostable DNA ligase to connect two nearby oligonucleotides and then detecting whether the ligase successfully joined the two oligonucleotides. Thereafter, Dr. Walt changed the focus of his laboratory research both at Tufts and at Illumina, Inc., and began developing commercial products to detect SNPs (single nucleotide polymorphisms) and eventually sequence DNA. There are relevant facts that the Court needs to be aware of regarding this matter. It is also notable that Dr. David Walt was not presented as a witness by Illumina Inc., even though as Chief Scientific Officer and Founder, he would have firsthand knowledge of materially pertinent facts.

Illumina produces randomly assembled bead arrays. A randomly assembled array is not useful until one knows which sequences are at which X and Y positions in a two dimensional array. As Dr. David Walt stated in testimony under oath in the *Syntrix Biosystems, Inc. v. Illumina Inc.* trial in the United States District Court in the Western District of Washington at Tacoma, once the beads were laid out on a silicon or a glass or plastic slide, one of the substrates described in his patent, the arrangement, "would be two dimensional." To define the location of any individual bead would require knowing only two positions, the X value and the Y value. Illumina Inc. claims that Infinium arrays are pure hybridization arrays. However, prior to use, the sequences at a given location must be determined. That is why in their own product literature, it is stated that the user must make sure the DMAP file (which contains the location data) must be in place prior to the analysis. Their "dynamically linked libraries", or DLL files, used to analyze the data in the DMAP files using the DMAP Decode File Client has labeled the sequence IDs at each X value and Y value, "Zipcode." Therefore, it is clear, that they consider that all of the unique "address" sequences that they use to decode or locate their beads are Zipcodes.

It is evident and clearly known by scientists that any given strand of DNA has a complementary sequence. During the initial experiments done at Cornell to demonstrate that Universal Zipcode Arrays work, known Zipcode sequences of 24 bases in length were synthesized and spotted in duplicate on arrays. This demonstrated that the two or more Zipcode sequences were hybridizing to the correct addresses on the array. This is clearly embodied in Example 3 in the '917 Patent. The data is shown in Figure 28 and quantified in Tables 4A and 4B.

I helped conduct some of these experiments and multiple Zipcode sequences and substrates and surfaces and configurations were tested. All of these initial experiments, used fluorescently labeled DNA using only Zipcode sequences and their complements with NO target DNA sequence. Some of these experiments later included the addition of non-Zipcode sequences to a Zipcode sequence prior to attachment on the array. This was one of the ways tested to increase the efficiency of hybridization by creating a larger template for DNA fragments to hybridize to. PCR/LDR products with Zipcodes were also eventually used to combine target and Zipcode sequences. Further, in the '470 Patent family, there are descriptions of embodiments in which composite Zipcodes containing NO target sequence would be captured on Universal Zipcode arrays.

In February 1999, I wrote software in Visual Basic and Microsoft Excel to generate lists of Zipcode sequences. These sequences were generated to have 24 bases in length and have a narrow melting temperature (Tm) range. In addition, each Zipcode sequence in the set differed from other Zipcode sequences in the set by 25% or more. These Zipcode sets were created using a group of 36 tetramers (4 base sequences) with unique properties. One of the copyrighted and eventually patented sequences (US Patents 7455965, 8492085, 9340834) that was generated using this approach was:

TTGAAAAGCCTACACGACGGCGAA

After my deposition in July 2015, I was asked for potential search terms that were materially relevant. This patented 24 base sequence plays a crucial fact role in this case. With the advent of search engines such as Google, these sequences act as a digital fingerprint to detect if Intellectual Property theft or infringement has occurred. If you Google this sequence, Dr. Barany's paper publishing this and other sequences appears in the search results. Surprisingly, a failed Illumina Patent application authored by Dr. Kevin Gunderson and Dr. Mark Chee entitled, "Probes and Decoder Oligonucleotides" also appears. In fact, this is the first entry in Table 2 in this and other Patent application submissions. It is labeled "Illumacode 1" in the initial provisional submission. This is materially relevant to this case as it not only demonstrates plagiarism and infringement of Intellectual Property rights, but also shows that Illumina considers their "Illumacodes" as Zipcodes corroborating the findings in Illumina's DMAP Decode File Client DLL files. These both constitute clear evidence of willful infringement.

In another relevant case, Dr. John Zebala, a former member of the Barany Laboratory founded Syntrix Biosystems, Inc. This company sued Illumina Inc. for willful infringement and the verdict labeled this as an example of more egregious acts of infringement. *See Syntrix Biosystems, Inc. v. Illumina Inc.,* Cause No. 13-5870BHS (W.D. Wash.). Tellingly, one of the first successful commercial products which Illumina used to generate revenues early on was called the Sentrix **Universal** Array Matrix. In this case, Dr. John Zebala had similarly shared confidential Intellectual Property information with Illumina and its employees and they proceeded to attempt to re-patent very similar ideas in their own names in a rapidly thrown together provisional application in February 2000.

Illumina, Inc. and its employees in the same time frame, also

gained access to proprietary intellectual property generated by Dr. Francis Barany and the other inventors in the '917 and '470 patent families. Using this information, Illumina. Inc. and/or its employees made multiple attempts in 1999 and 2000 and beyond to re-patent the ideas of another in a series of rapidly submitted provisional patent applications. These applications when examined together show a clear pattern of recidivism in terms of hand-drawing figures which very closely mimicked and exactly copied in some cases key figures from Dr. Barany's Program Project Grant, Papers and Patents.

In one particularly egregious instance, Dr. Kevin Gunderson and Dr. Mark Chee submitted on May 20, 1999, a provisional application (60/135123) entitled, "Addressing Arrays using Sequence Specific Adapters." In the very first paragraph of this submission, Cornell's WO 97/31256 and **Universal** arrays are clearly mentioned. It is an obvious attempt to re-patent Zipcodes under the guise of "Adapters." Many of these provisional applications were abandoned after multiple rejection letters by the USPTO and WIPO, many of which specifically cited Barany Laboratory Patents and Publications including WO 97/31256 ('917 Patent Family). Tellingly, references to Barany Laboratory Patents and Papers appear to have been deliberately omitted in subsequent follow up applications by Dr. Gunderson and Dr. Chee while at Illumina Inc.

Finally, the focus of my thesis project involved determining which adjacent sequences (with and without gaps), thermo-stable ligases were capable of successfully ligating. To study this, mononucleotide repeat sequences were chosen as a model system. These are naturally occurring repeats, errors in which sometimes contribute to development or progression of Cancers. It was demonstrated using various different length primers (some with deliberate gaps) and templates of varying lengths, that the ligase preferred perfectly matched primers without any gaps, but that significant signal was also generated from a subset of templates that contained gaps (See Tables in Appendix of my Doctoral Dissertation). This fact is also materially relevant to the *Cornell University, et al. v. Illumina Inc.* case.

I feel compelled to "come forward with specific facts showing that there is a genuine issue for trial." *Matsushita* 475 U.S. at 587 (internal quotation marks omitted.) It is clear that the Universal Zipcode Array concept has revolutionized the way DNA is sequenced and SNPs (single nucleotide polymorphisms) are analyzed. This has led to not only a decrease in the cost of DNA analyses but a rapid increase in the efficiency with which these are done. The Universal

Zipcode Array and other important concepts and assays conceived of and developed and proven in the Barany Laboratory in the mid 1990's, as well as exact copyright sequences generated in 1999 have been instrumental in this revolution.

As stated in *Vehicle IP, LLC v. AT&T Mobility, LLC, et al.* in C.A. No. 09-1007-LPS in the United States District Court Delaware (2016):

"The willfulness inquiry asks whether a party engaged in 'conduct warranting enhanced damages,' under 35 U.S.C § 284, behavior the Supreme Court has described as 'willful, wanton, malicious, bad-faith, deliberate, consciously wrongful, flagrant, or [] characteristic of a pirate.' *Halo Elecs., Inc. v. Pulse Elecs., Inc.*, 136 S. Ct. 1923, 1932 (2016). Prevailing on a claim of either indirect or willful infringement requires a patentee to prove, among other things, that an accused infringer acted with a specific intent to infringe. *See id.* at 1933 (willful infringement); *Global-Tech*, 563 U.S. at 765-66 (2011) (indirect infringement)."

Also stated in *Vehicle IP, LLC v. AT&T Mobility, LLC, et al.* in C.A. No. 09-1007-LPS in the United States District Court Delaware (2016):

"The Court will 'draw all reasonable inferences in favor of the nonmoving party'"

It is humbly hoped that these relevant material facts will be duly considered to help ensure that Justice is served in this Case.

I declare that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Respectfully submitted,

Date: _____ *By:* _____

Dr. Monib A. Zirvi, MD/PhD

EXHIBIT 13

Subpoena sent to Jian-Bing Fan and Illumina on July 23, 2015

IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF DELAWARE

CORNELL UNIVERSITY, CORNELL RESEARCH FOUNDATION, INC., LIFE TECHNOLOGIES CORPORATION AND APPLIED BIOSYSTEMS, LLC,	)	
Plaintiffs,	)	
v.	)	C.A. No. 10-433-LPS-MPT
ILLUMINA, INC.,	)	
Defendant.	)	

NOTICE OF SERVICE

PLEASE TAKE NOTICE that on July 23, 2015, the following documents were served on the persons listed below in the manner indicated:

1. Subpoena Directed to Jian-Bing Fan
2. First Notice of Deposition Pursuant to Federal Rule of Civil Procedure 30(b)(6) Directed to Illumina, Inc.
3. Second Notice of Deposition Pursuant to Federal Rule of Civil Procedure 30(b)(6) Directed to Illumina, Inc.
4. Notice of Deposition of Kevin Gunderson, Ph.D.

Illumina claims employees Mark Chee and Jian Bing Fan were not available during the time of deposition in the Cornell Case?

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	IP-0645-US
	Application Number	
Title of Invention	METHODS AND COMPOSITIONS FOR SINGLE CELL GENOMICS	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Jian-Bing		Fan	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	San Diego	State/Province	CA	Country of Residence i	US

Mailing Address of Inventor:

Address 1	5247 Pearlman Way		
Address 2			
City	San Diego	State/Province	CA
Postal Code	92130	Country i	US

Yet, on June 17th, 2015, Illumina submits a patent application with Jian Bing Fan listed as ”Inventor 1”, with an address of 5247 Pearlman Way, San Diego CA, Zip code 92130

Where is 5247 Pearlman Way, San Diego CA 92130?

- Imagine sending a letter to, say 5247 Pearlman Way, San Diego CA, Zip code 92130; the mailer may want to look up the location on Google Maps to make sure it gets to the right individual living in the house.

https://www.google.com/maps/place/5247+Pearlman+Way,+San+Diego,+CA+92130/@32.9537516,-117.205436,3a,75y,145.28h,90t/data=!3m6!1e1!3m4!1sVm1qQ7SYKEaG4_B30Y14vw!2e0!7i13312!8i6656!4m5!3m4!1s0x0:0xd3b437b3aeb2ef5e!8m2!3d32.9535502!4d-117.2052819!6m1!1e1

- What if it turned out the person living there was Jian-Bing Fan, an individual who works for Illumina, and filed a patent on June 17, 2015, literally weeks before deposition of the Cornell side, and at a time when Illumina claimed Jian-Bing Fan was not in the country and unavailable for deposition?

Full View: 5247 Pearlman Way, San Diego CA 92130

Detail 2: 5247 Pearlman Way, San Diego CA 92130

In May of 2015, Jian-Bing Fan submits a paper, where he is listed as an Illumina employee, but his email address is to “jianbing_fan@anchordx.com”, a company he founded in August, 2015.

www.impactjournals.com/oncotarget/ Oncotarget, Advance Publications 2015

Targeted or whole genome sequencing of formalin fixed tissue samples: Potential applications in cancer genomics

Sarah Munchel[1,*], Yen Hoang[2,5,*], Yue Zhao[1], Joseph Cottrell[1], Brandy Klotzle[1], Andrew K. Godwin[3], Devin Koestler[4], Peter Beyerlein[2], Jian-Bing Fan[1], Marina Bibikova[1], Jeremy Chien[5]

[1]Illumina, Inc., San Diego, CA

[2]Department of Bioinformatics and Biosystems Technology, University of Applied Sciences Wildau, Wildau, Germany

[3]Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS

[4]Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS

[5]Department of Cancer Biology, University of Kansas Medical Center, Kansas City, KS

[*]These authors have contributed equally to this work

Correspondence to:
Jeremy Chien, **e-mail:** jchien@kumc.edu
Marina Bibikova, **e-mail:** mbibikova@illumina.com
Jian-Bing Fan, **e-mail:** jianbing_fan@anchordx.com

Keywords: *cancer genomics, FFPE DNA, whole exome sequencing, whole genome sequencing, copy number alterations*

Received: *May 13, 2015* ***Accepted:*** *July 20, 2015* ***Published:*** *July 31, 2015*

Company Overview

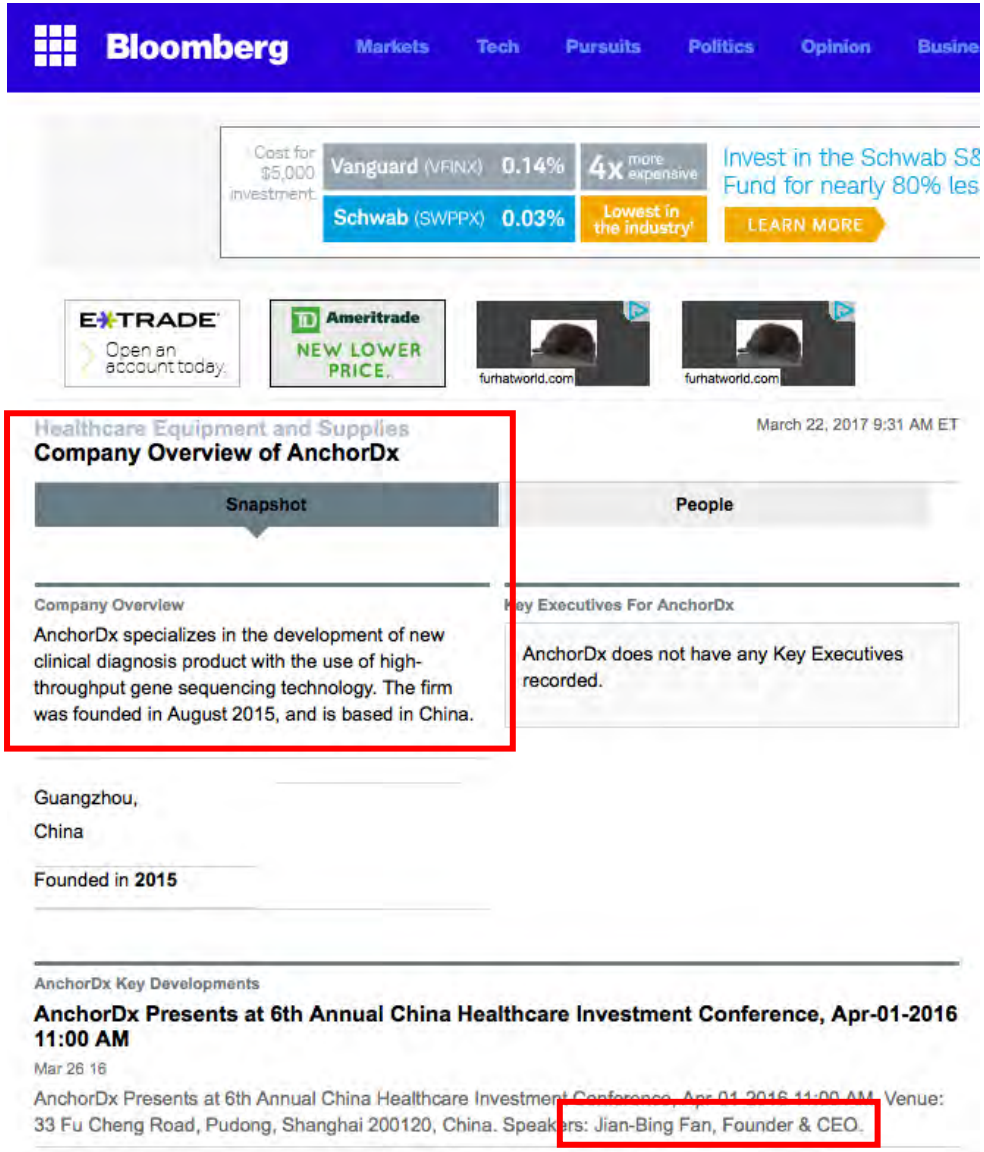
AnchorDx specializes in the development of new clinical diagnosis product with the use of high-throughput gene sequencing technology. The firm was founded in August 2015, and is based in China.

AnchorDx Presents at 6th Annual China Healthcare Investment Conference, Apr-01-2016 11:00 AM

Mar 26 16

AnchorDx Presents at 6th Annual China Healthcare Investment Conference, Apr-01-2016 11:00 AM. Venue: 33 Fu Cheng Road, Pudong, Shanghai 200120, China. Speakers: Jian-Bing Fan, Founder & CEO.

Bloomberg lists Jian-Bing Fan as founder and CEO of AnchorDx



Jian-Bing Fan's affiliation is listed at AnchorDx in San Diego, CA

ELSEVIER

Submit your paper

Browse journals > Genomics Data > Editorial Board > Fan

Jian-Bing Fan

- Editors-in-Chief

AnchorDx Corp., San Diego, California, USA

ELSEVIER

SEARCH

Home > Journals > Genomics Data > Editorial Board > Jian-Bing Fan

Submit Your Paper

View Articles

Guide for Authors

Journal Metrics

CiteScore: 0.42

More about CiteScore

Source Normalized Impact per Paper

Jian-Bing Fan

Editors-in-Chief, Genomics Data

AnchorDx Corp., San Diego, California, USA

Send an email to Jian-Bing Fan

To send an email to Jian-Bing Fan please complete the short form below. Please note that all enquiries should relate specifically to Genomics Data. All the fields are required so please make sure you complete them all otherwise we won't be able to send your message.

Illumina knew exactly where Jian-Bing Fan was in Sept. 2015

Dr. Jian-Bing Fan, a renowned NGS specialist, visits CNCL

时间: 2015-09-23 16:47 点击: 224次 字号: 小 | 中 | 大

Dr. Jian-Bing Fan, a renowned next generation sequencing (NGS) specialist and Chairman of AnchorDx Corporation (Guangzhou and San Diego), visited the Chinese National Compound Library (CNCL) in the noon of September 23, 2015. He was warmly received by his old friend, Dr. Ming-Wei Wang (Director of CNCL), as well as Dr. Caihong Zhou. They discussed collaborative research opportunities regarding novel therapeutic targets related to prostate cancer.

☆收藏 打印 关闭

Illumina knew exactly where Jian-Bing Fan was in June 2016

THE LIQUID BIOPSY SUMMIT Conference Agenda

WEDNESDAY, JUNE 22

3:45 pm Main Conference Registration

4:45 PANEL DISCUSSION: Current and Future Applications of Liquid Biopsies in Cancer

All agree CTCs and ctDNA are prognostic and predictive biomarkers for cancer. However, different approaches for CTCs/ctDNA detection and analysis to identify these tumor cell subpopulations need technical standardization before their clinical validity and biological specificity may be adequately investigated. Join these panelists as they discuss the current challenges and future opportunities for liquid biopsies.

Panelists:
Ellen M. Beasley, Ph.D., Genomic Health, Inc.
Geoff Otto, Ph.D., Foundation Medicine
Steven A. Soper, Ph.D., University of North Carolina, Chapel Hill
Rebecca (Becky) Suttmann, MS, Genentech, Inc.
Michael Kazinski, Senior Director, Molecular Pre-Analytical Technologies, QIAGEN
Judi Smith, MS, Vice President, In Vitro Diagnostics Regulatory and Quality, Precision for Medicine

9:30 Circulating RNAs as Noninvasive Biomarkers in Colorectal Cancer

Ajay Goel, Ph.D., Investigator/Professor & Director, Center for Gastrointestinal Research; Director, Center for Epigenetics, Cancer Prevention and Cancer Genomics, Baylor Research Institute and Charles A. Sammons Cancer Center, Baylor University Medical Center

Given their cancer-specific pattern of expression, remarkable stability and presence in blood and other body fluids, noncoding RNAs (ncRNAs) are considered to be highly promising "liquid biopsy" cancer biomarkers. Accumulating evidence firmly supports the existence of unique "ncRNA signatures" that can not only facilitate earlier detection of the tumor, but can also assist in predicting disease recurrence and therapeutic outcome to current treatment regimens.

10:00 Accessing Genetic Information with Liquid Biopsies

Jian-Bing Fan, Ph.D., CEO, AnchorDx Corp.

The molecular liquid biopsies approach provides non-invasive access to genetic information – somatic mutations, epigenetic changes, and differential expression – about the physiological conditions of our body and diseases. With the rapid development of highly sensitive and accurate technologies such as next-generation sequencing, it is now possible to reliably analyze CTCs and circulating nucleic acids in a clinic setting, which opens a valuable avenue for future genetic studies and human disease diagnosis.

Illumina knew exactly where Jian-Bing Fan was in Sept. 2016

DeviceChina 2016
中国医疗器械高峰论坛
Organized By: bioBAY
Co-Organized By:
September 9-11,2016
SUZHOU,CHINA
Register | Schedule Meetings | FAQ

HOME | ABOUT | AGENDA | SPEAKERS | SPONSORS | PARTNERS | COMMITTEE | VENUE | CONTACT

2016 SPEAKERS

KEYNOTE SPEAKERS

Yubo Fan President of the School of Biological Science and Medical Engineering	BUAA
Baoting Wang Deputy Secretary General	CHINA SOCIETY FOR DRUG REGULATION
Dongjin Wang Former Deputy Minister	MINISTRY OF HUMAN RESOURCES AND SOCIAL SECURITY

SPEAKERS

Kevin Chen partner at Sequoia Capital	Sequoia Capita
William Cao President &CEO	YIKON GENOMICS
David Cassak Managing Partner, Innovation In Medtech, LLC	Innovation In Medtech
Lifei Cheng Vice President	MEDTRONIC GREATER CHINA
Efi Cohen-Arazi Rainbow Medical Ltd	Cofounder & Chairman & CEO of Rainbow medical Ltd.
Jian-Bing Fan Founder and CEO	AnchorDx Corp
Yossi Gross Co-Founder of Rainbow Medical	Rainbow Medical Ltd

Jian-Bing Fan
Founder and CEO

AnchorDx Corp

Yossi Gross
Co-Founder of Rainbow Medical

Rainbow Medical Ltd

Binghua Su
Professor

College of Basic Medical S
Tong University

Jerry Wang
Head of Market Development, Illumina Greater China

Illumina

Xuyu Cai, lists her current address as Illumina, September 11th, 2014

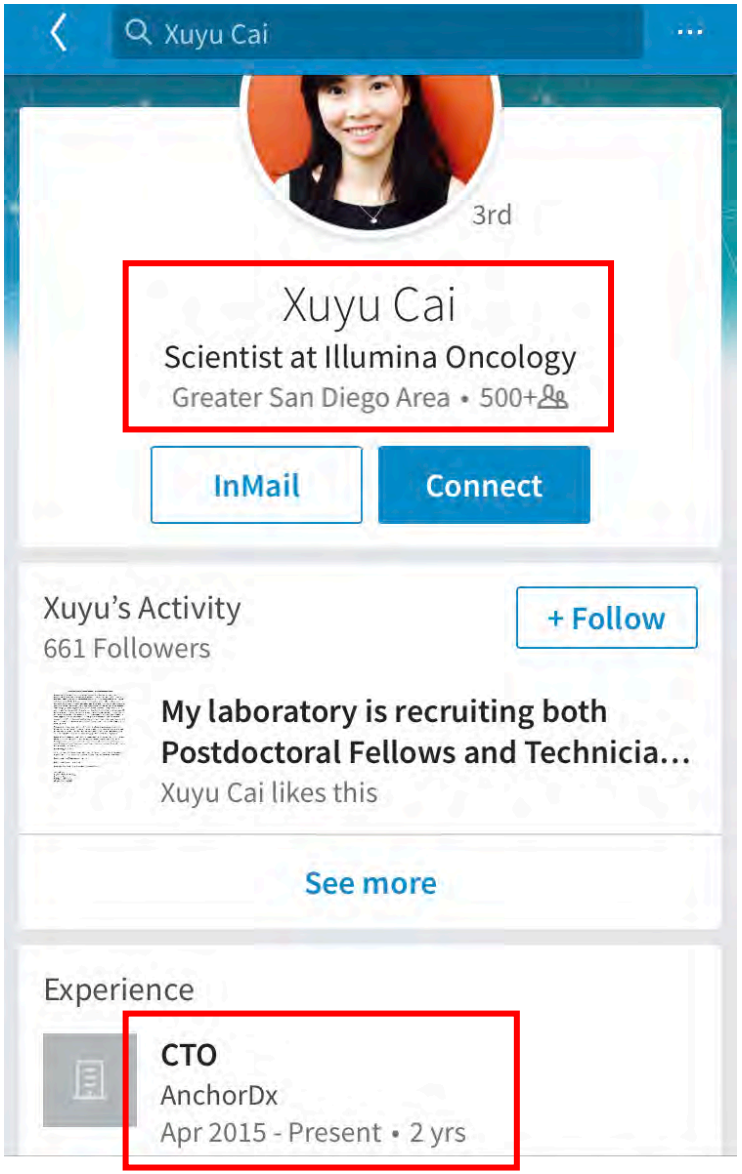
OPEN ACCESS CellPress

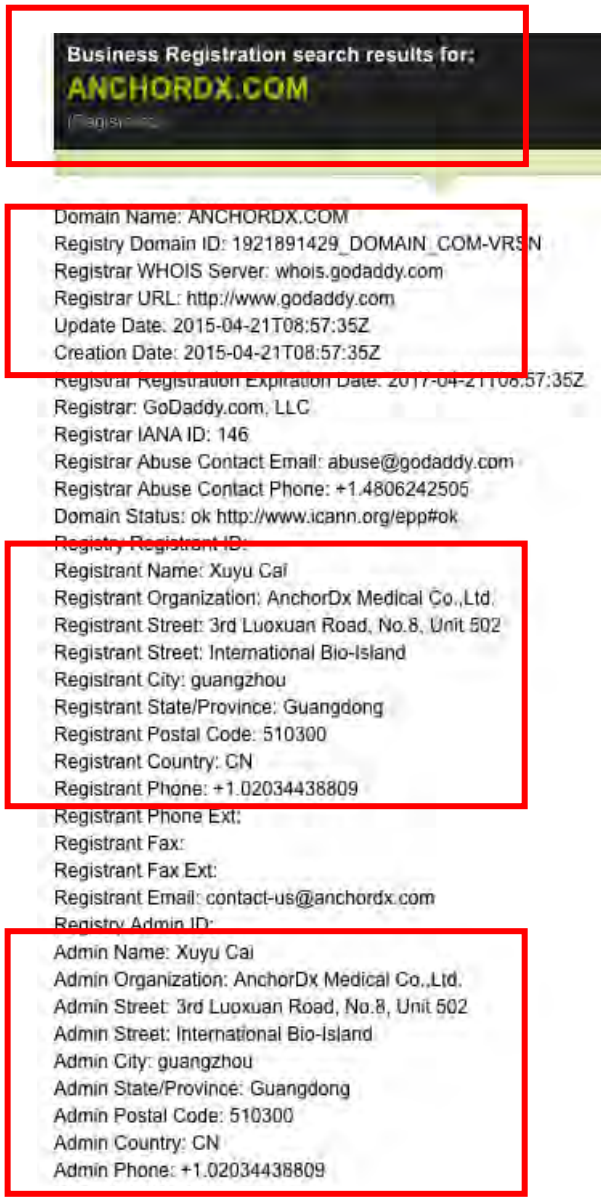
Cell Reports
Report

Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain

Xuyu Cai,[1,2,3,4,6] Gilad D. Evrony,[1,2,3,4] Hillel S. Lehmann,[1,2,3] Princess C. Elhosary,[1,2,3] Bhaven K. Mehta,[1,2,3] Annapurna Poduri,[1,2,3,5] and Christopher A. Walsh[1,2,3,4,*]
[1]Division of Genetics and Genomics, Manton Center for Orphan Disease Research and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA 02115, USA
[2]Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
[3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02138, USA
[4]Program of Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02115, USA
[5]Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA 02115, USA
[6]Present address: Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA
*Correspondence: christopher.walsh@childrens.harvard.edu
http://dx.doi.org/10.1016/j.celrep.2014.07.043

Xuyu Cai, a current Illumina employee, creates AnchorDx website, April 21, 2015





At the 2016 ASCO meeting Xuyu Cai and Jian-Bing Fan list their employment, stock ownership, and "Leadership" as: Illumina.

Ultra-deep next-generation sequencing of plasma cell-free (cf) DNA from patients with advanced cancers.

Subcategory:
Circulating Biomarkers

Category:
Tumor Biology

Meeting:
2016 ASCO Annual Meeting

Session Type and Session Title:
Poster Session, Tumor Biology

Abstract Number:
11537

Poster Board Number:
Board #234

Citation:
J Clin Oncol 34, 2016 (suppl; abstr 11537)

Author(s):
Filip Janku, Kiran Madwani, Shile Zhang, Helen J. Huang, Debra L. Andrews, David S. Hong, Daniel D. Karp, Siqing Fu, Xuyu Cai, Yue Zhao, Jian-Bing Fan, Jill Waters, Jonathan Toung, Li Liu, Min-Jui Richard Shen, Gordon B. Mills, Nelson Salathia; Department of Investigational Cancer Therapeutics (Phase 1 Program), The University of Texas MD Anderson Cancer Center, Houston, TX; The University of Texas MD Anderson Cancer Center, Houston, TX; Illumina, Inc., San Diego, CA; Department of Investigational Cancer Therapeutics (Phase I Program), University of Texas MD Anderson Cancer Center, Houston, TX; MD Anderson Cancer Center, Pearland, TX; Department of Investigational Cancer Therapeutics (Phase I Program), The University of Texas MD Anderson Cancer Center, Houston, TX; Illumina, San Diego, CA

Abstract Disclosures

Abstract:

Background: Plasma cfDNA offers an easily obtainable source of DNA for molecular analysis, which provides an attractive alternative to tumor tissue testing. Novel ultrasensitive multiplex technologies are needed for further implementation of cfDNA testing in personalized cancer therapy. **Methods:** We have developed an ultra-deep next-generation sequencing method for detection of somatic alterations in up to 62 common cancer related genes in plasma cfDNA. Each cfDNA fragment was uniquely barcoded and amplified prior to Illumina target enrichment workflow, followed by ultra-deep sequencing (> 10,000X). Proprietary data processing and analysis tools were developed to enable sensitive detection of rare mutant molecules over high wild-type background (detection of 1 in 1,000 - 10,000 molecules). Results were compared to molecular profiling of archival primary or metastatic tumor tissue obtained at different points of clinical care from a CLIA-certified laboratory. **Results:** Plasma cfDNA was extracted from 61 patients with advanced cancers (colorectal, n = 16; non-small cell lung [NSCLC], n = 12; breast, n = 6; melanoma, n = 4; cholangiocarcinoma, n = 3; thyroid, n = 3; ovarian, n = 3; pancreatic, n = 3; appendiceal, n = 2; cervical, n = 2; glioma, n = 2; other, n = 6) and for the presence of alterations in up to 62 common cancer related genes. Ultra-deep sequencing of plasma cfDNA detected key driver molecular alterations reported from archival tumor tissue samples in 66% (41/62) cases and absolute agreement between utra-deep sequencing of plasma cfDNA and archival tumor tissue was reported 53% (33/62) cases. There were no disagreements in samples from patients with cholangiocarcinoma, cervical, ovarian and pancreatic cancer, while none of the plasma cfDNA samples from patients with glioma revealed molecular alterations detected in the tumor tissue. **Conclusions:** Detecting common oncogenic alterations using ultra-deep sequencing of plasma cfDNA is feasible and should be further investigated for testing in patients with advanced cancer.

The following represents disclosure information provided by authors of this abstract. The program committee has reviewed all presenting author disclosure reports, identified potential conflicts of interest, and implemented strategies to manage those areas of conflict, where appropriate. All relationships are considered compensated. Relationships are self-held unless otherwise noted. I = Immediate Family Member, Inst = My Institution

Ultra-deep next-generation sequencing of plasma cell-free (cf) DNA from patients with advanced cancers.

Filip Janku
Consulting or Advisory Role - Deciphera; Foundation Medicine; Novartis; Sequenom; Trovagene
Research Funding - Agios; Astellas Pharma; Biocartis; BioMed Valley Discoveries; Deciphera; Foundation Medicine; Novartis; Roche; Trovagene
Other Relationship - Bio-Rad

Xuyu Cai
Employment - Illumina
Stock and Other Ownership Interests - Illumina

Yue Zhao
Employment - Illumina
Stock and Other Ownership Interests - Illumina

Jian-Bing Fan
Employment - Illumina
Leadership - Illumina
Stock and Other Ownership Interests - Illumina

Jian-Bing Fan, in a Science paper published on June 24th, 2016, lists his current address as Illumina, San Diego, CA, USA.

One day earlier, on June 23rd, 2016, Jian-Bing Fan spoke about "Accessing Genetic Information with Liquid Biopsies" at the Hotel Kabuki, in San Francisco, CA.

NEUROGENOMICS

Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain

Blue B. Lake,[1]* Rizi Ai,[2]* Gwendolyn E. Kaeser,[3,4]* Neeraj S. Salathia,[5]* Yun C. Yung,[3] Rui Liu,[1] Andre Wildberg,[2] Derek Gao,[1] Ho-Lim Fung,[1] Song Chen,[1] Raakhee Vijayaraghavan,[5] Julian Wong,[3] Allison Chen,[3] Xiaoyan Sheng,[3] Fiona Kaper,[5] Richard Shen,[5] Mostafa Ronaghi,[5] Jian-Bing Fan,[5]† Wei Wang,[2]† Jerold Chun,[3]† Kun Zhang[1]†

The human brain has enormously complex cellular diversity and connectivities fundamental to our neural functions, yet difficulties in interrogating individual neurons has impeded understanding of the underlying transcriptional landscape. We developed a scalable approach to sequence and quantify RNA molecules in isolated neuronal nuclei from a postmortem brain, generating 3227 sets of single-neuron data from six distinct regions of the cerebral cortex. Using an iterative clustering and classification approach, we identified 16 neuronal subtypes that were further annotated on the basis of known markers and cortical cytoarchitecture. These data demonstrate a robust and scalable method for identifying and categorizing single nuclear transcriptomes, revealing shared genes sufficient to distinguish previously unknown and orthologous neuronal subtypes as well as regional identity and transcriptomic heterogeneity within the human brain.

[1]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. [2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA. [3]Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA, USA. [4]Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA. [5]Illumina, San Diego, CA, USA.
*These authors contributed equally to this work. †Corresponding author. Email: kzhang@bioeng.ucsd.edu (K.Z.); jchun@scripps.edu (J.C.); wei-wang@ucsd.edu (W.W.); jianbing_fan@yahoo.com (J.-B.F.)

1586 24 JUNE 2016 • VOL 352 ISSUE 6293

Illumina files its 10/177,727 continuation application, June 20th, 2002, reaches back to provisional application 60/234,143 filed on Sept. 21st, 2000.

(12) **United States Patent**
Shen et al.

(10) **Patent No.: US 7,955,794 B2**
(45) **Date of Patent: *Jun. 7, 2011**

(54) **MULTIPLEX NUCLEIC ACID REACTIONS**

(75) Inventors: **Mun-Jui Richard Shen**, Poway, CA (US); **Arnold Oliphant**, Poway, CA (US); **Scott L. Butler**, San Diego, CA (US); **John E. Stuelpnagel**, Encinitas, CA (US); **Mark S. Chee**, Del Mar, CA (US); **Kenneth M. Kuhn**, San Diego, CA (US); **Jian-Bing Fan**, San Diego, CA (US)

(73) Assignee: **Illumina, Inc.**, San Diego, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 824 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **10/177,727**

(22) Filed: **Jun. 20, 2002**

(65) **Prior Publication Data**

US 2003/0211489 A1 Nov. 13, 2003

Related U.S. Application Data

(60) Provisional application No. 60/234,143, filed on Sep. 21, 2000, provisional application No. 60/234,732, filed on Sep. 22, 2000, provisional application No. 60/297,609, filed on Jun. 11, 2001, provisional application No. 60/311,271, filed on Aug. 9, 2001, provisional application No. 60/336,958, filed on Dec. 3, 2001, provisional application No. 60/305,118, filed on Jul. 12, 2001, provisional application No. 60/341,827, filed on Dec. 17, 2001.

4,882,269 A 11/1989 Schneider et al.
4,883,750 A 11/1989 Whiteley et al.
4,921,805 A 5/1990 Gebeyehu et al.
4,988,617 A * 1/1991 Landegren et al. 435/6
4,999,306 A * 3/1991 Yafuso et al. 436/68
5,002,867 A * 3/1991 Macevicz 435/6
5,011,769 A * 4/1991 Duck et al. 435/6
5,028,545 A * 7/1991 Soini 436/501
5,104,791 A 4/1992 Abbott et al.
5,105,305 A * 4/1992 Betzig et al. 359/368
5,114,864 A * 5/1992 Walt 436/528
5,132,242 A * 7/1992 Cheung 436/501
5,143,853 A * 9/1992 Walt 436/501
5,175,082 A 12/1992 Jeffreys
5,185,243 A * 2/1993 Ullman et al. 435/6
5,194,300 A * 3/1993 Cheung 427/213.31
5,232,829 A 8/1993 Longiaru et al.
5,244,636 A * 9/1993 Walt et al. 422/82.07
5,244,813 A * 9/1993 Walt et al. 436/172
5,250,264 A * 10/1993 Walt et al. 422/82.07
5,252,494 A * 10/1993 Walt 436/528
5,254,477 A * 10/1993 Walt 436/172
5,298,741 A * 3/1994 Walt et al. 250/227.23
5,302,509 A * 4/1994 Cheeseman 435/6
5,314,809 A 5/1994 Erlich et al.
5,320,814 A * 6/1994 Walt et al. 422/82.07
5,357,590 A * 10/1994 Auracher 385/33
5,380,489 A * 1/1995 Sutton et al. 422/68.1
5,387,505 A 2/1995 Wu et al.
5,399,491 A * 3/1995 Kacian et al. 435/91.21
5,403,711 A 4/1995 Walder et al.
5,409,818 A * 4/1995 Davey et al. 435/91.21
5,427,930 A * 6/1995 Birkenmeyer et al. 435/91.52
5,435,724 A * 7/1995 Goodman et al. 433/215
5,445,934 A * 8/1995 Fodor et al. 435/6
(Continued)

FOREIGN PATENT DOCUMENTS

EP 0139489 5/1985
(Continued)

OTHER PUBLICATIONS

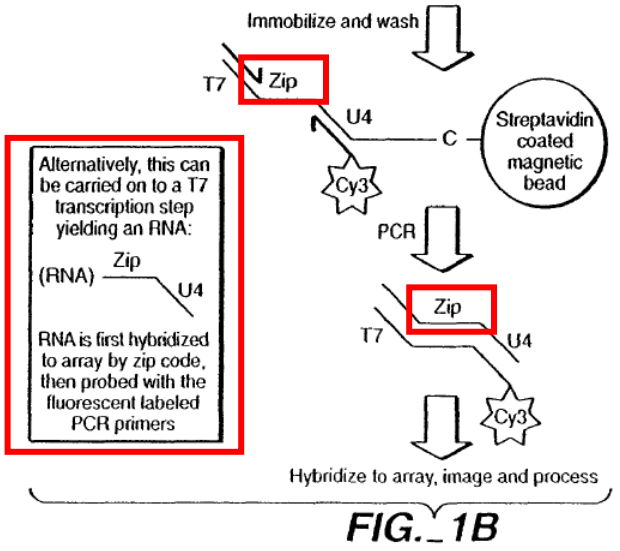


FIG._1B

FIG. 1 depicts a schematic of a preferred embodiment of the invention. The primary steps of the method include annealing oligonucleotides to immobilized target (e.g. genomic) DNA, a chain extension reaction that is terminated by tagged (e.g. biotinylated) ddNTPs (FIG. 1A), isolation and amplification of the tagged extension products (FIG. 1B).

Illumina uses word "Zip" for "Zipcode" in figures, but does not use or define in figure legends.

hybridized primers. An extension reaction is carried out. The extended products are eluted and used in a PCR amplification reaction (using the universal PCR primers specific for these oligos U1, U2 and U3). U2 and U3 are labeled with different fluorescent tags. The ratio in the amount of one allele relative to another is determined by the ratio of the fluorescent tags.

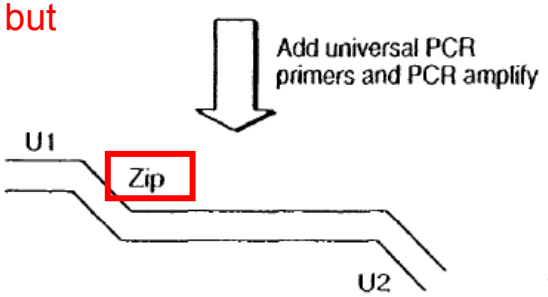


FIG._3

Stanford files its Padlock Probe provisional application, Oct. 24th, 2000, one month after Illumina's provisional application 60/234,143 (Sept. 21st, 2000).

(12) **United States Patent**
Willis et al.

(10) **Patent No.: US 6,858,412 B2**
(45) **Date of Patent: Feb. 22, 2005**

(54) **DIRECT MULTIPLEX CHARACTERIZATION OF GENOMIC DNA**

(75) Inventors: **Thomas D. Willis**, San Francisco, CA (US); **Paul Hardenbol**, Los Altos, CA (US); **Maneesh Jain**, Menlo Park, CA (US); **Viktor Stolc**, Cupertino, CA (US); **Mostafa Ronaghi**, Palo Alto, CA (US); **Ronald W. Davis**, Palo Alto, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Palo Alto, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/999,362**
(22) Filed: **Oct. 24, 2001**
(65) **Prior Publication Data**

US 2004/0101835 A1 May 27, 2004

Related U.S. Application Data
(60) Provisional application No. 60/242,901, filed on Oct. 24, 2000.

FOREIGN PATENT DOCUME

EP	0 320 308 B1	6/1989
EP	0 336 731 B1	10/1989
EP	0 439 182 B1	7/1991
EP	0 799 897 A1	10/1997
WO	WO 89/09835	10/1989
WO	WO 89/12696	12/1989
WO	WO 90/01069	2/1990
WO	WO 93/25563	12/1993
WO	WO 95/00667	1/1995
WO	WO 95/05480	2/1995
WO	WO 95/14106	5/1995
WO	WO 96/15271 A	5/1996
WO	WO 97/31256	8/1997
WO	WO 98/04746 A1	2/1998
WO	WO 96/17958	6/1999
WO	WO 00/55369	9/2002
WO	WO 03/006677 A2	1/2003

OTHER PUBLICATIONS

Barany F, "Genetic disease detection and DI using cloned thermostable ligase," *Proc Na* Jan. 1;88(1):189–93 (1991).
Beaucage and Caruthers, "Deoxynucleos mides," *Tetrahedron Letts.*, 22(20:1859–18
Chee M. "Enzymatic multiplex DNA sequ *Acids Res.* Jun. 25;19(12):3301–5 (1991).
Fodor et al., "Light–directed, spatially add chemical synthesis," *Science*. Feb. 15;25

What is claimed is:

1. A method for detecting a target sequence comprising a first and second target domain in a sample, said method comprising:

a) hybridizing said target sequence to a precircle probe to form a first hybridization complex, said precircle probe comprising:
i) a first targeting domain;
ii) a second targeting domain;
iii) at least a first universal priming site; and
iv) a cleavage site;

wherein said first and second targeting domains hybridize to said first and second target domains;

b) contacting said first hybridization complex with a ligase to form a closed circular probe;

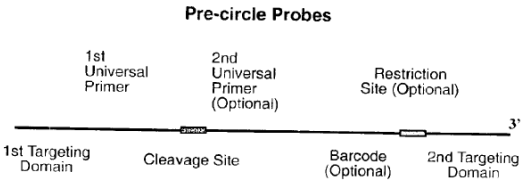


FIG._1

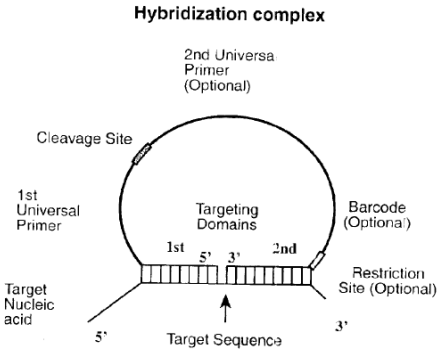


FIG._2A

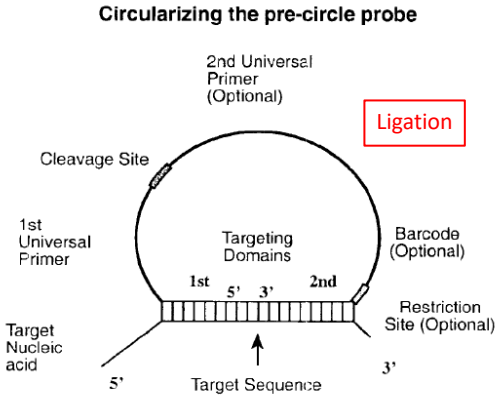


FIG._2B

Exonuclease

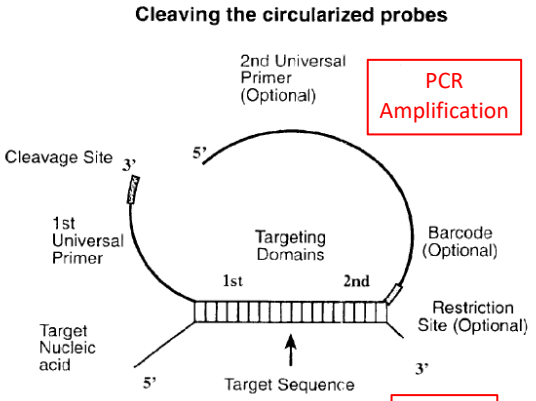


FIG._2C

Zip code Capture

Cornell files its Padlock Probe provisional application, May 29th, 1996, four years ahead of both Illumina and Stanford; it teaches ligation, exonuclease, universal PCR amplification, and zip code capture.

(12) **United States Patent**
Barany et al.

(10) **Patent No.: US 7,556,924 B2**
(45) **Date of Patent: Jul. 7, 2009**

(54) **DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING COUPLED LIGASE DETECTION WITH PADLOCK PROBES AND POLYMERASE CHAIN REACTIONS**

(75) Inventors: **Francis Barany**, New York, NY (US); **Matthew Lubin**, Rye Brook, NY (US); **George Barany**, Falcon Heights, MN (US); **Robert P. Hammer**, Baton Rouge, LA (US)

(60) Provisional application No. 60/018,532, filed on May 29, 1996.

(51) **Int. Cl.**
C12Q 1/68 (2006.01)
C12P 19/34 (2006.01)
(52) **U.S. Cl.** **435/6**; 435/91.2
(58) **Field of Classification Search** 435/6, 435/91.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

What is claimed:

1. A method for identifying one or more different target nucleotide sequences comprising:

providing a sample potentially containing one or more target nucleotide sequences comprising sequence differences;

providing one or more padlock oligonucleotide probes, each probe comprising (a) a first target-specific portion and a 5' upstream primer-specific portion and (b) a second target-specific portion and a 3' downstream primer-specific portion, wherein the first and second target-specific portions of a particular probe are suitable for ligation together when hybridized on a corresponding target nucleotide sequence, but have a mismatch which interferes with such ligation when first and second target-specific portions of a particular probe are hybridized to any other nucleotide sequence present in the sample;

providing a ligase;

blending the sample, the one or more padlock oligonucleotide probes, and the ligase to form a ligase detection reaction mixture;

subjecting the ligase detection reaction mixture to one or more ligase detection reaction cycles to form a ligation product sequence comprising (a) the 5' upstream primer specific portion, (b) the target-specific portions, and (c) the 3' downstream primer-specific portion, when the respective target nucleotide sequence of the corresponding padlock oligonucleotide probe is present in the sample;

providing one or a plurality of oligonucleotide primer sets, each set comprising (a) an upstream primer and (b) a downstream primer;

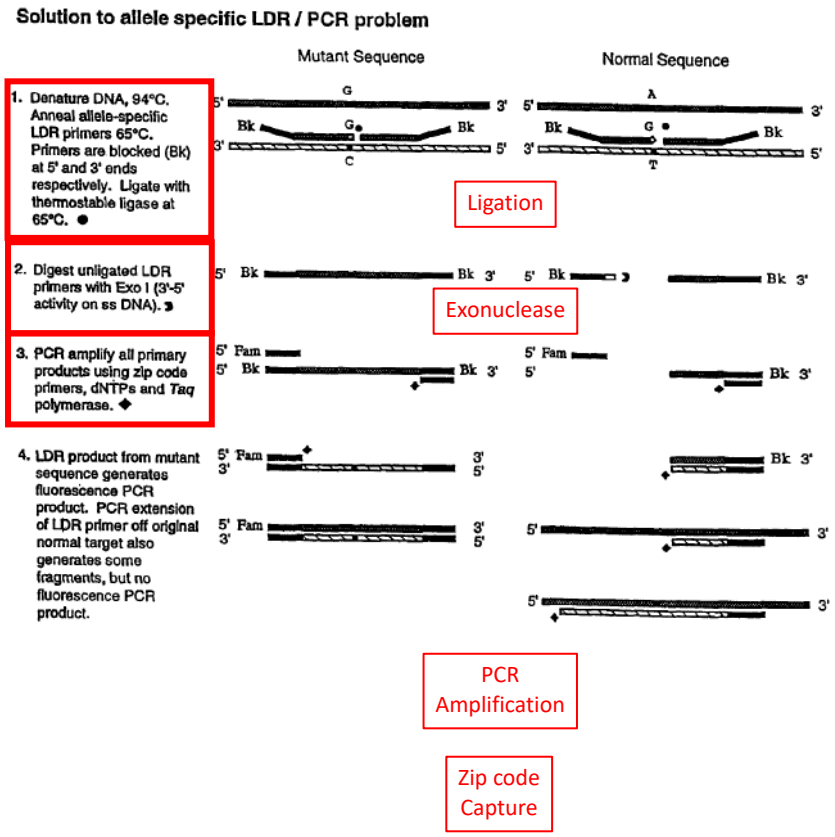
providing a polymerase;

blending the ligase detection reaction mixture with the one or a plurality of oligonucleotide primer sets, and the polymerase to form a polymerase chain reaction mixture;

subjecting the polymerase chain reaction mixture to one or more polymerase chain reaction cycles to form extension products comprising the ligation product sequence and/or complements thereof; and

detecting the extension products to identity one or more target nucleotide sequences in the sample.

2. The method according to claim **1**, wherein the upstream primer is complementary to the 5' upstream primer-specific portion of the ligation product sequence and the downstream primer is complementary to the 3' downstream primer-specific portion of the ligation product sequence.



Cornell '470 family patents are 5 years ahead of Illumina.

Cornell, Filed May 29, 1996

(12) **United States Patent**
Barany et al.

(10) **Patent No.: US 6,797,470 B2**
(45) **Date of Patent: Sep. 28, 2004**

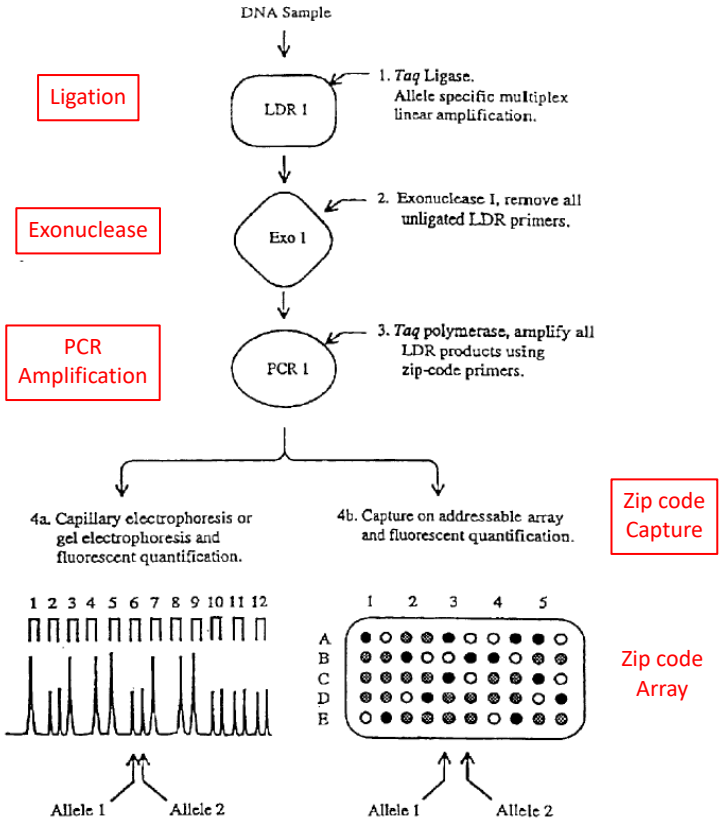
(54) **DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING COUPLED LIGASE DETECTION AND POLYMERASE CHAIN REACTIONS**

(75) Inventors: **Francis Barany**, New York, NY (US); **Matthew Lubin**, Rye Brook, NY (US); **George Barany**, Falcon Heights, MN (US); **Robert P. Hammer**, Baton Rouge, LA (US)

5,035,996 A 7/1991 Hartley
5,143,854 A 9/1992 Pirrung et al.
(List continued on next page.)

FOREIGN PATENT DOCUMENTS

EP WO 92/10588 6/1992
EP WO 92/16655 10/1992
EP WO 93/17126 9/1993
EP WO 93/20236 10/1993
EP 0 601 714 A1 6/1994
EP WO 96/06190 2/1996



Illumina Provisional, Filed Feb. 7, 2001

(19) **United States**
(12) **Patent Application Publication**
Fan et al.

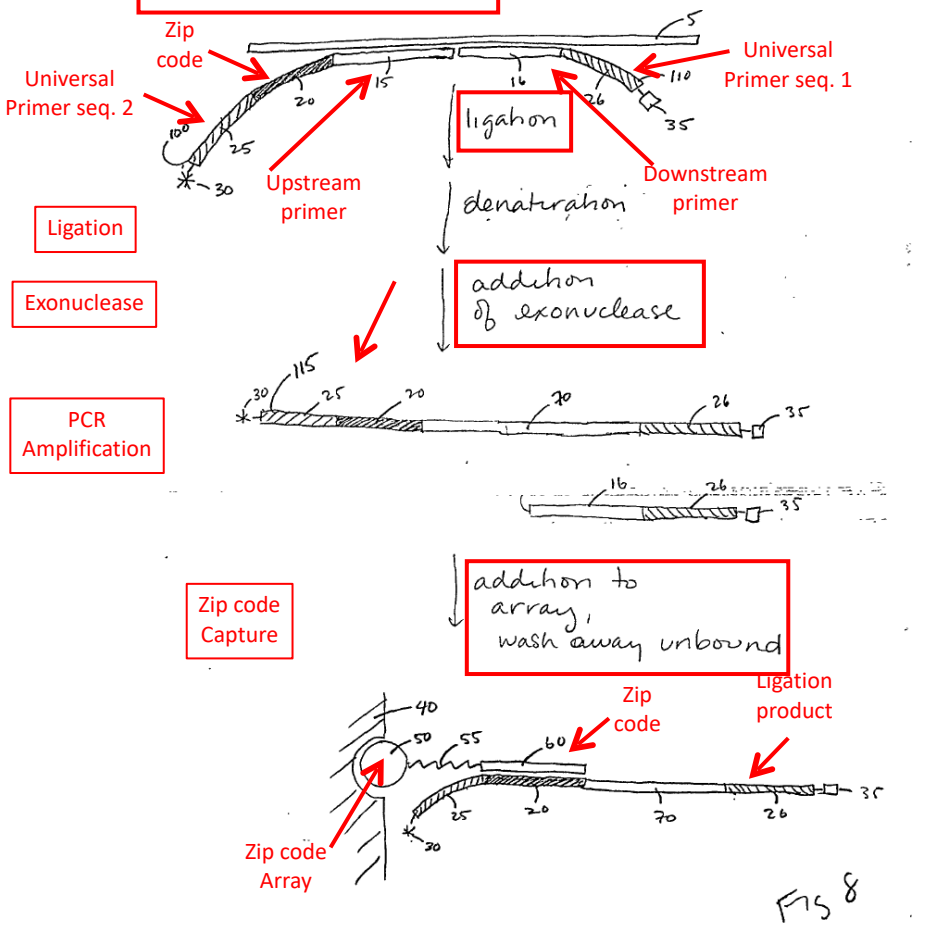
(10) **Pub. No.: US 2002/0172946 A1**
(43) **Pub. Date: Nov. 21, 2002**

(54) **NUCLEIC ACID DETECTION METHODS USING UNIVERSAL PRIMING**

(76) Inventors: **Jian-Bing Fan**, San Diego, CA (US); **Xiang-Dong Fu**, San Diego, CA (US)

Related U.S. Application Data

(60) Provisional application No. 60/180,810, filed on Feb. 7, 2000. Provisional application No. 60/234,731, filed on Sep. 22, 2000.



Jian-Bing Fan declaration on Oct. 30, 2007 (re: Illumina continuation application 10/177,727, filed June 20th, 2002), claims ability to accurately and reproducibly detect a "100plex" in a single reaction "was very much contrary to expectations."

Docket No.: 067234-0021 **PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Shen, Min-Jui Richard, et al.
Appl. No. : 10/177,727
Filed : June 20, 2002
Title : MULTIPLEX NUCLEIC ACID REACTIONS

Customer No.: 41552
Confirmation No.: 9974

Grp./A.U. : 1634
Examiner: : Lu, Frank Wei Min

DECLARATION UNDER 37 C.F.R. § 1.132

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

I, Jian-Bing Fan, Ph.D., declare as follows:

1) I am a Director of Scientific Research at Illumina, Inc. (Illumina), where I have held this position for 7 years. Prior to my current position, I worked as a Senior Scientist from June 1999 to March 2000 and an Associate Director of Genetic Analysis from April 2000 to October 2000 at Illumina.

5) I am a co-inventor on the application entitled "Multiplex Nucleic Acid Reactions" having U.S. Serial No. 10/177,727, filed June 20, 2002. I have read the Office Action mailed July 13, 2007, and understand that claims 1-3 and 10 are rejected under 35 U.S.C. § 103(a) as obvious over Barany et al., U.S. Patent No. 6,027,889 ("Barany et al."), in view of Kohne, U.S. Patent No. 5,567,587 ("Kohne"). I also understand that to the extent the current amendments to claim 1 and new claims 12-36 might be considered to fall under this ground of rejection then such claims also would be rejected similar reasons. I have read and am familiar with Barany et al. and Kohne.

6) I understand that the Examiner cites Barany et al. for describing a detection assay using at least first and a second target sequences corresponding to the first and second parts of a target nucleic acid sequence to form first and second hybridization complexes, amplification and detection. Kohne is cited for describing methods for fixing sample nucleic acid onto an inert support . The Examiner concludes that one of ordinary skill in the art would have been motivated to apply the method of Kohne to attach at least first and second different target sequences to a solid support to the assay of Barany et al. because it would enhance separation of hybridized from non-hybridized complexes.

7) For the reasons summarized in this paragraph and detailed in the paragraphs that follow, it is my opinion that, based on the cited references, the ordinary skilled person in the art would not have expected to combine a solid phase immobilization step of a plurality of at least 100 different target sequences as described and claimed in the application to achieve the claimed results. The invention claims a method of detecting target sequences in a multiplex format where at least 100 different target sequence determinations are assayed in a single reaction mixture. At the time the invention was made, multiplexing samples for nucleotide detection at this level was not viewed as possible. The ability to accurately and reproducibly detect at least 100 different target sequence determinations ("100plex") in a single reaction mixture was very much contrary to expectations. At the time, ordinary solution phase assays such as the one described by Barany et al. and Kohne were used in a multiplex format, but only at very low levels of about 12-24 targeted nucleotide positions in a single assay. Therefore, achieving multiplex detection of at least 100 different nucleotide positions was very much considered a great advancement.

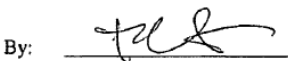
Jian-Bing Fan declaration on Oct. 30, 2007, claims initial solid phase immobilization step enabled an accurate assay for 1,536 different SNPs in a single reaction, with call rates of 0.989 and concordance of 0.982, while performing the same reactions in liquid gave inferior call rates of 0.548 and concordance of 0.783.

8) To demonstrate the extreme disparity between multiplex detection without the claimed immobilization step and the multiplex method of the invention using attachment of a plurality of target nucleic acid sequences to a solid phase I have performed a side-by-side comparison. Multiplex assays were performed following the methods of the invention. A copy of the procedure employing immobilization of a plurality of target nucleic acid sequences is attached as Exhibit 2. A copy of this procedure employed without this immobilization step is attached as Exhibit 3. The procedures are identical except for the first several steps. In Exhibit 2, genomic DNA was first annealed with oligonucleotide probes and immobilized onto a solid-support; several wash steps were performed to remove excess and mis-hybridized oligonucleotides; subsequent oligonucleotide probe extension and ligation reactions were carried out on solid-phase. While in Exhibit 3, the oligonucleotide probe annealing, extension and ligation reactions occur simultaneously in a reaction solution. These modifications to the protocol of Exhibit 3 were mandated by, and incurred because of, the lack of the immobilization step in method of the invention (Exhibit 2).

9) The results of this side-by-side comparison are attached as Exhibits 4-9 and demonstrate that the claimed methods yield exceedingly high levels of call rate and reproducibility when at least 100 different detection positions are determined in a single reaction. In contrast, employing the same method without immobilization of the target nucleic acid sequences produced essentially random scattered data that cannot be used to make reliable genotype call.

10) Exhibit 4 is a summary of the results comparing the claimed method of the invention with or without immobilization of target nucleic acid sequences. Each of sixteen different genomic DNA ("gDNA") samples was assayed for 1,536 different SNPs in a single reaction mixture. Therefore, for each gDNA sample, the multiplex level is 1,536. For each multiplex assay, Exhibit 4 provides the call rate, which is the fraction of correctly determined nucleotide identity at particular positions ("genotype calls") expressed as a fraction of all possible calls, and the concordance, which is a measurement for the degree of reproducibility (i.e., same genotypes reported by duplicated experiments). Experiments preformed in our assay without any genomic DNA input yield a call rate from randomly scattered data points of about 0.6 (data not shown). Therefore, call rates below 0.6 are considered to be akin to random background noise. The results shown in Exhibit 4 indicate that immobilization of the genomic target sequences dramatically increases both the call rate and the reproducibility of each multiplex sample compared to the procedure without an immobilization step. The mean call rate and concordance was 0.548 and 0.783, respectively, without the immobilization, indicating that these multiplex determinations were essentially random. In contrast, immobilization of genomic target sequences increased the mean call rate to 0.989 and the concordance to 0.982, respectively.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that any such willful false statement may jeopardize the validity of the application or any patent issued thereon.

Date: 10/30/2007 By: 

Jian-Bing Fan, Ph.D.

Despite Jian-Bing Fan declaration on Oct. 30, 2007, Ron Davis's group filed for a patent in Oct. 2000, and published in June 2003 the ability to genotype 1,517 loci in a single liquid reaction with an accuracy of 99.95%.

Multiplexed genotyping with sequence-tagged molecular inversion probes

Paul Hardenbol[1,3], Johan Banér[2], Maneesh Jain[1,3], Mats Nilsson[2], Eugeni A Namsaraev[1,3], George A Karlin-Neumann[1,3], Hossein Fakhrai-Rad[1,3], Mostafa Ronaghi[1], Thomas D Willis[1,3], Ulf Landegren[2] & Ronald W Davis[1]

We report on the development of molecular inversion probe (MIP) genotyping, an efficient technology for large-scale single nucleotide polymorphism (SNP) analysis. This technique uses MIPs to produce inverted sequences, which undergo a unimolecular rearrangement and are then amplified by PCR using common primers and analyzed using universal sequence tag DNA microarrays, resulting in highly specific genotyping. With this technology, multiplex analysis of more than 1,000 probes in a single tube can be done using standard laboratory equipment. Genotypes are generated with a high call rate (95%) and high accuracy (>99%) as determined by independent sequencing.

Here we present a strategy that combines DNA detection specificity and sensitivity with the potential to analyze large numbers of target sequences in parallel. Sets of padlock probes with universal tag sequences were reacted with target DNA, molecularly inverted, amplified together and identified in a multiplex analysis yielding more than 1,000 genotypes simultaneously. Using molecular inversion probes, the information content of the SNPs was reformatted into tag sequences that could be detected using a universal oligonucleotide detection array[13]. We report the application of this technique at unprecedented levels of multiplexing, resulting in a lowering of the scale, cost and sample requirements of high-throughput genotyping. The approach retained high accuracy through multiple hybridization and enzymatic processing events, and provided inherent quality control checking.

NATURE BIOTECHNOLOGY VOLUME 21 NUMBER 6 JUNE 2003

Accuracy was determined through independent sequencing. 1,517 loci were genotyped in a 1,517-probe multiplex analysis with ten individuals. Forward and reverse Sanger sequencing was performed on a subset (129) of PCR amplicons of 1,517 loci amplified from the same 10 individuals. Conservative reads were made manually with the identity of the forward and reverse loci blinded at the time of sequence interpretation. Accuracy of Sanger sequencing was measured by comparing reads for which the sequence of both strands existed. 359 of 367 sequence pairs were identical, for an accuracy of 97.8%. The accuracy of the 359 agreeing sequence pairs was therefore 99.95%. In the genotyping data set, 312 full genotypes and 23 half genotypes were identical with the Sanger sequence pairs (643/647 chromosome comparisons), for an accuracy of 99.4%. Similar accuracy was achieved when the data were compared with pyrosequencing[18,19] data on different sets of SNPs. The four discordant genotypes were found in a single locus (probe 2101) in four individuals. Because the sequencing data were nonpolymorphic and subsequent sequencing of these loci with newly generated amplicons was concordant with our genotyping results, it is likely that the original amplification for Sanger sequencing was contaminated with other template DNA or PCR product from another individual.

DISCUSSION

The MIP genotyping method described here has several advantages over alternative techniques. No singleplex PCR amplification is required before mutation detection, thereby reducing labor and expense. PCR is applied only after mutation detection, at which time all molecular inversion probes are converted to standard-length oligonucleotides of similar sequence composition and common primers. This results in a high degree of multiplexing capacity. We have not observed any change in performance in multiplexing from a single probe up to 1,500 probes and speculate that a further increase to 10,000 probes might be possible because sufficient signal is generated in the assay to support that many probes. The data

Ron Davis's group was successful in achieving this high accuracy by following the teachings of the Cornell '470 patent series by using ligation, exonuclease, universal PCR amplification, and zip code capture.

Multiplexed genotyping with sequence-tagged molecular inversion probes

Paul Hardenbol[1,3], Johan Banér[2], Maneesh Jain[1,3], Mats Nilsson[2], Eugeni A Namsaraev[1,3], George A Karlin-Neumann[1,3], Hossein Fakhrai-Rad[1,3], Mostafa Ronaghi[1], Thomas D Willis[1,3], Ulf Landegren[2] & Ronald W Davis[1]

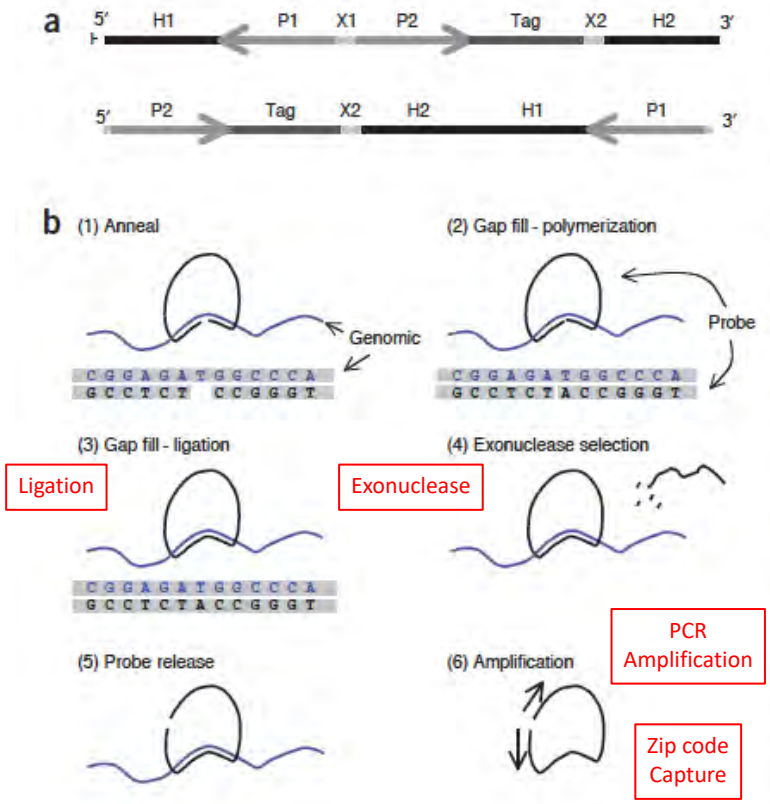


Figure 2 Molecular inversion probes. (**a**) Unreacted probe (top) and inverted probe (bottom). A single probe is used to detect both alleles of each SNP and consists of seven segments: two regions of homology to target genomic DNA, H1 and H2 (unique to each probe) at the termini of the probe, two PCR primer regions common to all probes, one bar code specific for each locus and two common cleavage sites, X1 and X2. Successfully reacted probes are amplified using primers P1 and P2. A universal detection tag sequence, one of 16,000, is for array detection of amplified probe. Cleavage sites X1 and X2 are used to release the circularized probe from genomic DNA and for post-amplification processing, respectively. (**b**) Enzymatic probe inversion. (1) A mixture of Genomic DNA, 1,000 or more probes, and thermostable ligase and polymerase is heat-denatured and brought to annealing temperature. Two sequences targeting each terminus of the probe hybridize to complementary sites in the genome, creating a circular conformation with a single-nucleotide gap between the termini of the probe. (2) Unlabeled dATP, dCTP, dGTP or dTTP, respectively, is added to each of the four reactions. In reactions where the added nucleotide is complementary to the single-base gap, DNA polymerase adds the nucleotide and (3) DNA ligase closes the gap to form a covalently closed circular molecule that encircles the genomic strand to which it is hybridized. (4) Exonucleases are added to digest linear probes in reactions where the added nucleotide was not complementary to the gap and excess linear probe in reactions where circular molecules were formed. The reactions are then heated to inactivate the exonucleases. (5) To release probes from genomic DNA, uracil-*N*-glycosylase is added to depurinate the uracil residues in the probes. The mixture is then heated to cleave the molecule at the abasic site and release it from genomic DNA. (6) PCR reagents are added, including a primer pair common to all probes. The reactions are then subjected to thermal cycling, with the result that only probes circularized in the allele-specific gap-fill reaction are amplified.

Ron Davis's group successfully genotypes 10,000 and then 20,000 loci in a single tube, publishing in 2005, two years before Jian-Bing Fan's declaration.

Methods

Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a single tube assay

Paul Hardenbol,[1] Fuli Yu,[2] John Belmont,[2] Jennifer MacKenzie,[1] Carsten Bruckner,[1] Tiffany Brundage,[1] Andrew Boudreau,[1] Steve Chow,[1] Jim Eberle,[1] Ayca Erbilgin,[1] Mat Falkowski,[1] Ron Fitzgerald,[1] Sy Ghose,[2] Oleg Iartchouk,[1] Maneesh Jain,[1] George Karlin-Neumann,[1] Xiuhua Lu,[2] Xin Miao,[1] Bridget Moore,[1] Martin Moorhead,[1] Eugeni Namsaraev,[1] Shiran Pasternak,[2] Eunice Prakash,[1] Karen Tran,[1] Zhiyong Wang,[1] Hywel B. Jones,[1] Ronald W. Davis,[3] Thomas D. Willis,[1,4] and Richard A. Gibbs[2]

[1]ParAllele BioScience, Inc., South San Francisco, California 94080, USA; [2]Baylor College of Medicine, Human Genome Sequencing Center, Houston, Texas 77030, USA; [3]Stanford Genome Technology Center, Stanford University, California 94305, USA

Large-scale genetic studies are highly dependent on efficient and scalable multiplex SNP assays. In this study, we report the development of Molecular Inversion Probe technology with four-color, single array detection, applied to large-scale genotyping of up to 12,000 SNPs per reaction. While generating 38,429 SNP assays using this technology in a population of 30 trios from the Centre d'Etude Polymorphisme Humain family panel as part of the International HapMap project, we established SNP conversion rates of ~90% with concordance rates >99.6% and completeness levels >98% for assays multiplexed up to 12,000plex levels. Furthermore, these individual metrics can be "traded off" and, by sacrificing a small fraction of the conversion rate, the accuracy can be increased to very high levels. No loss of performance is seen when scaling from 6,000plex to 12,000plex assays, strongly validating the ability of the technology to suppress cross-reactivity at high multiplex levels. The results of this study demonstrate the suitability of this technology for comprehensive association studies that use targeted SNPs in indirect linkage disequilibrium studies or that directly screen for causative mutations.

15:269–275 ©2005 by Cold Spring Harbor Laboratory Press;

Genome Research 269
www.genome.org

Published online November 28, 2005

Nucleic Acids Research, 2005, Vol. 33, No. 21 e183
doi:10.1093/nar/gni177

Allele quantification using molecular inversion probes (MIP)

Yuker Wang, Martin Moorhead, George Karlin-Neumann, Matthew Falkowski, Chunnuan Chen, Farooq Siddiqui, Ronald W. Davis[1], Thomas D. Willis and Malek Faham*

ParAllele BioScience, 7300 Shoreline Boulevard, South San Francisco, CA 94080, USA and [1]Stanford Genome Technology Center, 855 California Avenue, Palo Alto, CA 94304, USA

Received July 22, 2005; Revised and Accepted October 25, 2005

ABSTRACT

Detection of genomic copy number changes has been an important research area, especially in cancer. Several high-throughput technologies have been developed to detect these changes. Features that are important for the utility of technologies assessing copy number changes include the ability to interrogate regions of interest at the desired density as well as the ability to differentiate the two homologs. In addition, assessing formaldehyde fixed and paraffin embedded (FFPE) samples allows the utilization of the vast majority of cancer samples. To address these points we demonstrate the use of molecular inversion probe (MIP) technology to the study of copy number. MIP is a high-throughput genotyping technology capable of interrogating >20000 single nucleotide polymorphisms in the same tube. We have shown the ability of MIP at this multiplex level to provide copy number measurements while obtaining the allele information. In addition we have demonstrated a proof of principle for copy number analysis in FFPE samples.

Ulf Landegren's group also publishes on using padlock probes with zip code arrays in 2003, comparing to the Ron Davis publication.

Nucleic Acids Research, 2003, Vol. 31, No. 17 ***e103***
DOI: 10.1093/nar/gng104

Parallel gene analysis with allele-specific padlock probes and tag microarrays

Johan Banér, Anders Isaksson, Erik Waldenström[1], Jonas Jarvius, Ulf Landegren* and Mats Nilsson

The Beijer Laboratory, Department of Genetics and Pathology, Rudbeck Laboratory, SE-751 85 Uppsala, Sweden and [1]Department of Medical Sciences, Uppsala University Hospital, SE-751 85 Uppsala, Sweden

We have recently presented an approach for parallel gene analysis using large pools of sequence-tagged padlock probes. These were added to individual DNA samples and subjected to gap-fill ligation followed by exonuclease treatment to remove both unreacted and any cross-reactive products that might arise through intermolecular ligation. Next, the circularized probes were molecularly inverted by linearization at a position remote from the site of ligation and amplified by PCR. Finally, the probes were analyzed on commercially available tag sequence arrays (6). Using this strategy individual DNA samples were analyzed for more than 1500 SNPs in parallel, and more recently the technique has been used to analyze sets of 13 000 markers (T.Willis, ParAllele BioScience, personal communication). The approach exploits some useful properties of padlock probes, oligonucleotides that circularize in target sequence-dependent ligation reactions (7). The requirement for recognition of adjacent target sequences by the two ends of these linear probes provide sufficient specificity to analyze SNPs in total genomic DNA without prior target amplification (6,8–10). Moreover, as shown by Hardenbol *et al.* (6), large numbers of padlock probes can be combined in single reactions since amplification of unreacted and cross-reacted probes can be counteracted by exonuclease treatment.

ABSTRACT

Parallel, highly specific analysis methods are required to take advantage of the extensive information about DNA sequence variation and of expressed sequences. We present a scalable laboratory technique suitable to analyze numerous target sequences in multiplexed assays. Sets of padlock probes were applied to analyze single nucleotide variation directly in total genomic DNA or cDNA for parallel genotyping or gene expression analysis. All reacted probes were then co-amplified and identified by hybridization to a standard tag oligonucleotide array. The technique was illustrated by analyzing normal and pathogenic variation within the Wilson disease-related ATP7B gene, both at the level of DNA and RNA, using allele-specific padlock probes.

In this paper we report an alternative protocol for padlock-based analyses of sequence variants in genomic DNA or transcribed sequences. Sets of allele-specific padlock probes were combined in a single ligation reaction, followed by exonucleolytic removal of unreacted and cross-reacted probes. The remaining circularized probes were amplified and the products identified in an array-of-arrays, prepared using a standard printing robot. Twenty-six allele-specific padlock probes were used to genotype 75 individuals at 100% call rate and accuracy, and with sufficient signal strength for further scale-up. The method was also used to evaluate allele-specific gene expression by comparing genomic DNA samples from liver necropsies with cDNA prepared from the same samples. We discuss differences between the present approach and that used in Hardenbol *et al.* (6).

Ulf Landegren's group also followed the teachings of the Cornell '470 patent series by using ligation, exonuclease, universal PCR amplification, and zip code capture.

Parallel gene analysis with allele-specific padlock probes and tag microarrays

Johan Banér, Anders Isaksson, Erik Waldenström[1], Jonas Jarvius, Ulf Landegren* and Mats Nilsson

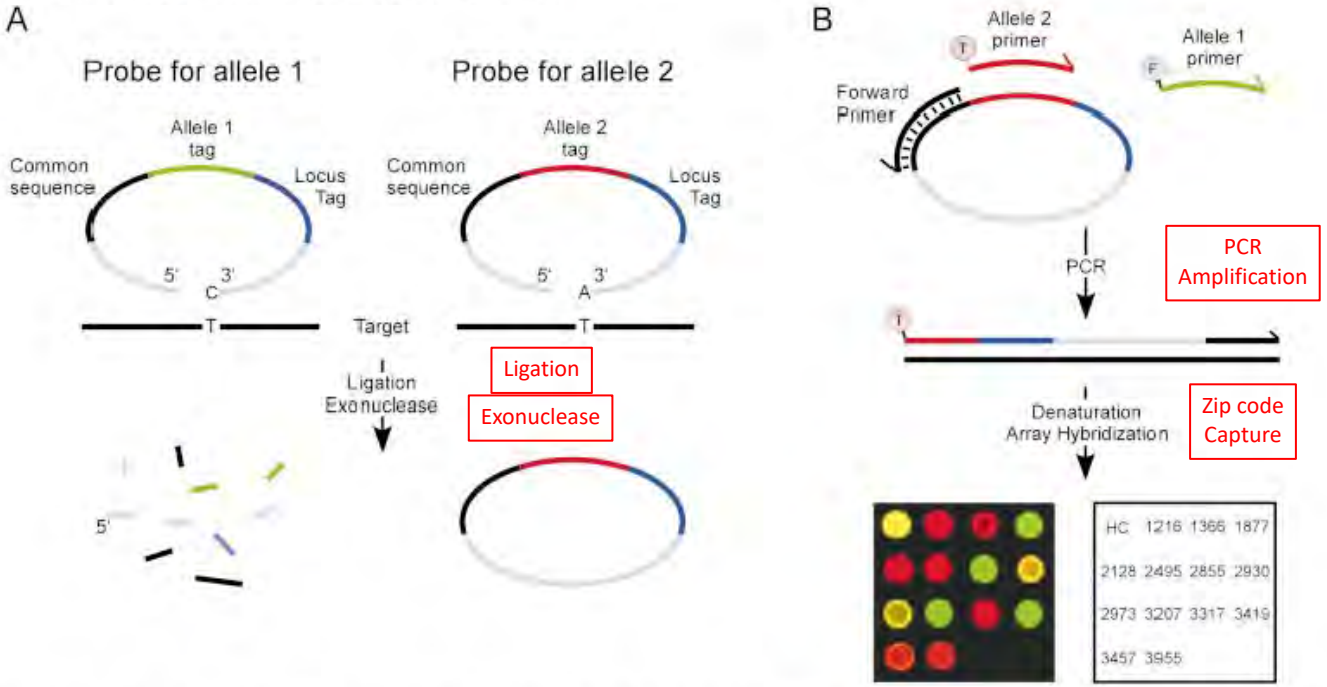


Figure 1. Parallel padlock probe analysis of single nucleotide variation. (**A**) Padlock probes included two target-complementary sequences at both the 5′ and 3′ ends (grey) and the allele-specific nucleotide was positioned at the 3′-end. The target-specific ends were connected by the following sequence elements, 20 nt each, in 5′→3′ order: one sequence common for all probes used for amplification (black), either of two allele-specific tags (green or red) and, finally, a tag sequence unique for each locus (blue). Only probe arms correctly hybridized to a target sequence (black line) are efficiently joined by a ligase. After ligation, unreacted or dimerized probes are degraded by exonuclease treatment, preserving circularized ones. (**B**) Circularized probes at any locus serve as templates in a PCR with one primer specific for the common sequence (black) and two allele-specific primers, labeled with either FITC for allele 1 (green primer, marked F) or TAMRA for allele 2 (red primer, marked T). PCR products are then denatured and applied to a DNA microarray with complementary tag sequences, guiding the labeled strands to their corresponding positions. An overlaid scanning image is shown from an analysis of one individual (red color, TAMRA; green color, FITC; a combination of TAMRA and FITC appears as yellow). The spot position key identifies the investigated loci. Numbers represent locus nomenclature according to Petrukhin *et al.* (19).

Ulf Landegren's group confirms the teachings of the Cornell '470 patent series by using direct ligation, exonuclease, universal PCR amplification, and zip code capture. He shows the results are as accurate using direct ligation as using the gap ligation approach of Ron Davis's group, i.e. there is insubstantial difference between the two approaches.

Parallel gene analysis with allele-specific padlock probes and tag microarrays

Johan Banér, Anders Isaksson, Erik Waldenström[1], Jonas Jarvius, Ulf Landegren* and Mats Nilsson

DISCUSSION

We present a strategy for parallel analyses of large sets of DNA or RNA sequence variants using padlock probes to detect and distinguish gene sequences in total genomic DNA or cDNA. The approach represents a scalable alternative to the highly multiplexed method we recently described (6), differing in a few aspects as described below.

In the method used by Hardenbol *et al.* (6), circularization of probes was affected in combined polymerization and ligation reactions, i.e. separate gap fill reactions were performed for each nucleotide triphosphate (14). These procedures require only one probe per polymorphism, eliminating probe pair imbalances and reducing probe cost, but four reactions must be performed per individual typed. In the present strategy a single ligation reaction was used to determine all genotypes in a sample which ensured that identical reaction conditions were used, but twice as many probes had to be prepared.

The combination of allele-specific incorporation of nucleotides followed by ligation in the previous study provides two levels of allele distinction. First, the correct nucleotide must be incorporated, and any misincorporated nucleotide would still hamper the ligation reaction. However, ligation reactions by themselves are known to distinguish single nucleotide variants by factors of thousands (15,16). Accordingly, the allele-specific padlock probe ligation reactions used in the present study also yield robust discrimination of alleles.

15. Luo,J., Bergstrom,D.E. and Barany,F. (1996) Improving the fidelity of *Thermus thermophilus* DNA ligase. *Nucleic Acids Res.*, **24**, 3071–3078.

16. Nilsson,M., Banér,J., Mendel-Hartvig,M., Dahl,F., Antson,D.O., Gullberg,M. and Landegren,U. (2002) Making ends meet in genetic analysis using padlock probes. *Hum. Mutat.*, **19**, 410–415.

Concerns with Jian-Bing Fan's Declaration of Oct. 30th 2007:

- Jian-Bing Fan declares: " The ability to accurately and reproducibly detect at least a 100 different target sequence determinations ("100plex") in a single reaction mixture was very much contrary to expectations."

- Jian-Bing Fan then proceeds to give results using his own experiments comparing an initial solid phase capture vs. a liquid version, which were done at some unknown time. However, in the liquid version, Jian-Bing Fan knowingly fails to get rid of extra probe using exonuclease, as taught in the Cornell '470 series patents filed and published 5 years prior to Illumina's 10/177,727 filing, which is the subject of the declaration.

- However, contrary to Jian-Bing Fan's assertions of the declaration, it was well known at the time that highly multiplexed reaction of 1,517 probes (Hardenbol, et. al., Nature Biotech 2003), 10,000 probes (Hardenbol, et. al., Genome Research 2005), and even 20,000 probes (Wang et. al., NAR, 2005) in a single reaction mixture were achieved with high accuracy.

- Landegren's group also confirms highly multiplexed reaction in liquid, and demonstrates equal accuracy when using direct ligation, and gap ligation.

- Thus, an impartial outside observer may conclude that Fan misled and/or provided false testimony to the USPTO in order to get his claims allowed so that Illumina would get a foundational patent.

Jian-Bing Fan, from Affymetrix, but while employed at Illumina, tries to re-patent Universal Array and composite oligonucleotides described by Cornell 3 years earlier:

(19) **United States**

(12) **Patent Application Publication**

Fan et al.

(10) **Pub. No.: US 2005/0074787 A1**

(43) **Pub. Date: Apr. 7, 2005**

(54) **UNIVERSAL ARRAYS**

(75) Inventors: **Jian-Bing Fan**, San Diego, CA (US); **Joel N. Hirschhorn**, Newton, MA (US); **Xiaohua Huang**, Mountain View, CA (US); **Paul Kaplan**, Campbell, CA (US); **Eric S. Lander**, Cambridge, MA (US); **David J. Lockhart**, Del Mar, CA (US); **Thomas Ryder**, Los Gatos, CA (US); **Pamela Sklar**, Brookline, MA (US)

Correspondence Address:
Lisa M. Treannie, Esq.
Ropes & Gray LLP
One International Place
Boston, MA 02110-2624 (US)

(73) Assignees: **Whitehead Institute for Biomedical Research**, Cambridge, MA; **The General Hospital Corporation**, Boston, MA; **Affymetrix, Inc**, Santa Clara, CA

(21) Appl. No.: **10/730,771**

(22) Filed: **Dec. 8, 2003**

Related U.S. Application Data

(62) Division of application No. 09/536,841, filed on Mar. 27, 2000.

(60) Provisional application No. 60/126,473, filed on Mar. 26, 1999. Provisional application No. 60/140,359, filed on Jun. 23, 1999.

Publication Classification

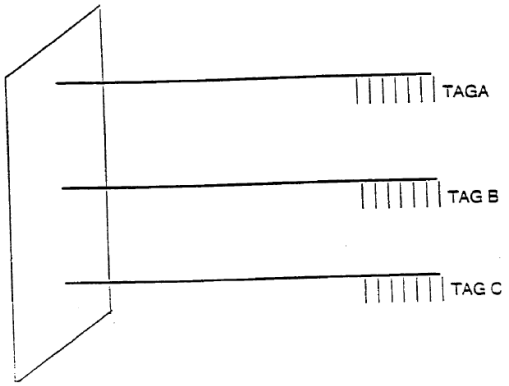
(51) **Int. Cl.**[7] **C12Q 1/68**

(52) **U.S. Cl.** **435/6**

(57) **ABSTRACT**

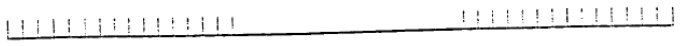
An array of oligonucleotides on a solid substrate is disclosed, which can be used for multiple purposes. Methods and reagents are provided for performing genotyping to determine the identity or ratio of allelic forms of a gene in a sample. A single base extension primer is coupled to a sequence identity code. During the primer extension reaction a distinctive label is incorporated which identifies the allelic form present in the sample. This permits multiple simultaneous analyses to be performed easily and efficiently.

GENERIC HYBRIDIZATION ARRAYS
SOLID SUBSTRATE WITH GENERIC
MOLECULAR ADDRESSES



[0011] FIG. 1 is a diagram of the universal array: The solid substrate (e.g., a glass slide) is depicted on the left, and different oligonucleotide tags ("A", "B", "C", etc.) are shown attached to the solid substrate. The nucleotide sequence on the right-hand end of each oligonucleotide tag ("Tag A", "Tag B", "Tag C") is arbitrary unique sequence; that is, it is designed and synthesized to be unique to each oligonucleotide tag.

Locus-specific tagged oligonucleotide:



synthetic sequence specific for a particular oligonucleotide tag on the array (e.g., "Tag A," "Tag B," "Tag C," etc.)

sequence specific for the amplification product of a particular SNP (e.g., SNP "A," SNP "B," SNP "C," etc.)

[0012] FIG. 2 is a diagram depicting a locus-specific tagged oligonucleotide. The nucleotide sequence at the left-hand end is complementary to the arbitrary sequence of one of the oligonucleotide tags depicted in **FIG. 1**. The nucleotide sequence at the right-hand end is complementary to the amplification product of a known polymorphic locus (e.g., a single nucleotide polymorphism (SNP)). Therefore, locus-specific tagged oligonucleotide "A" comprises a nucleotide sequence complementary to the arbitrary sequence of the "Tag A" oligonucleotide tag depicted in **FIG. 1**, and also comprises sequence complementary to SNP "A".

Why did key Illumina employees Mark Chee and Jian Bing Fan BOTH disappear (to China) during the time of deposition in the Cornell Case?

2nd Principal Investigators Meeting Abstracts

2nd Principal Investigators Meeting
Innovative Molecular Analysis Technologies Programs
June 27-29, 2001
Ritz-Carlton
Washington, DC

Mark Chee
Illumina, Inc.
Randomly Self-Assembled Bead-Based Arrays for Gene Expression Analysis

Mark Chee, Tim McDaniel, Shawn Baker, Semyon Kruglyak, Francisco Garcia, Kenneth Kuhn, Csilla Fenczik, Kevin Gunderson, and Jian-Bing Fan

Illumina, Inc., San Diego, CA

Self-assembled arrays of bead-based sensors have been developed. Each bead contains oligonucleotide probes that can hybridize with high specificity to complementary sequences in a complex nucleic acid mixture. The identity of each bead in the random array is determined by a hybridization-based decoding procedure. By formatting the miniaturized arrays into a matrix that matches a 96-well microtiter plate, many samples can be processed efficiently in parallel. We are using this platform to develop assays in the areas of SNP genotyping, gene expression profiling, and protein analysis. Progress in the area of gene expression profiling will be discussed.

Xiang-Dong Fu
University of California, San Diego
Cancer Classification Based on MRNA Isoforms

Xiang-Dong Fu[1], Joanne M. Yeakley[1], Jian-Bing Fan[2], Mark Chee[2], David Tarin[3], Michael Gribskov[4], and Michael Zhang[5]

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, [2]Illumina, Inc., San Diego, [3]UCSD Cancer Center, [4]San Diego Supercomputer Center, and [5]Cold Spring Harbor Laboratory

Our project aims at developing a genomic approach to alternative splicing because current techniques for assaying alternative splicing rely on low throughput methods that are not suitable for genome-wide analysis. The technology utilizes the addressable zip-code strategy on fiber optic-based microarrays developed at Illumina. Briefly, DNA oligos are used to cover specific splice junction sequences and each isoform-sepcific oligo is linked to a unique zip-code sequence. The oligos are first hybridized to total RNA and those detecting specific alternative splicing events are selectively amplified by PCR. Individual splicing events are then sorted and quantified on a universal Zipcode array. Using this novel microarray technology, we can detect specific alternative splicing events with unprecedented specificity and sensitivity, reaching to the single cell level in certain cases. Furthermore, semi-quantitative data are obtained from a panel of experimental cell lines and 90% of data points match the results obtained by conventional RT-PCR. Although we have been focusing on technology development in the initial R21 phase, our research has already led to the discovery of a number of striking cell-specific alternative splicing events in our model systems.

Was Illumina aware that Mark Chee and Jian-Bing Fan were the only two Illumina employees tied to abstracts provided at a June 27-29, 2001 meeting at the Ritz-Carlton in Washington DC, where the “hybridization-based decoding procedure” was incontrovertibly linked to “a universal Zipcode array”, thus exposing Illumina’s fraudulent arguments?

Summary of Concerns with Illumina's missing witnesses: Part 1

- Two key witnesses, Mark Chee and Jian-Bing Fan of Illumina, in the Cornell vs. Illumina case were said to be "missing" or "could not be found" according to Illumina, and therefore were not made available for deposition, although they were originally offered as witnesses in their defense by Illumina.

- Their depositions would have been valuable to expose Illumina's infringement of Cornell's Universal DNA Arrays and Zipcodes. Chee and Fan both claimed to have invented these on declarations and oaths to the USPTO. Since there is an abundance of proof that they were fully aware of Cornell's WO 97/31256 patent, this calls all of their oaths and declarations into question.

- The "disappearance" of Jian-Bing Fan appears to have been deliberately orchestrated with the apparent help of Illumina via at least one of its current employees at the time of the subpoena.

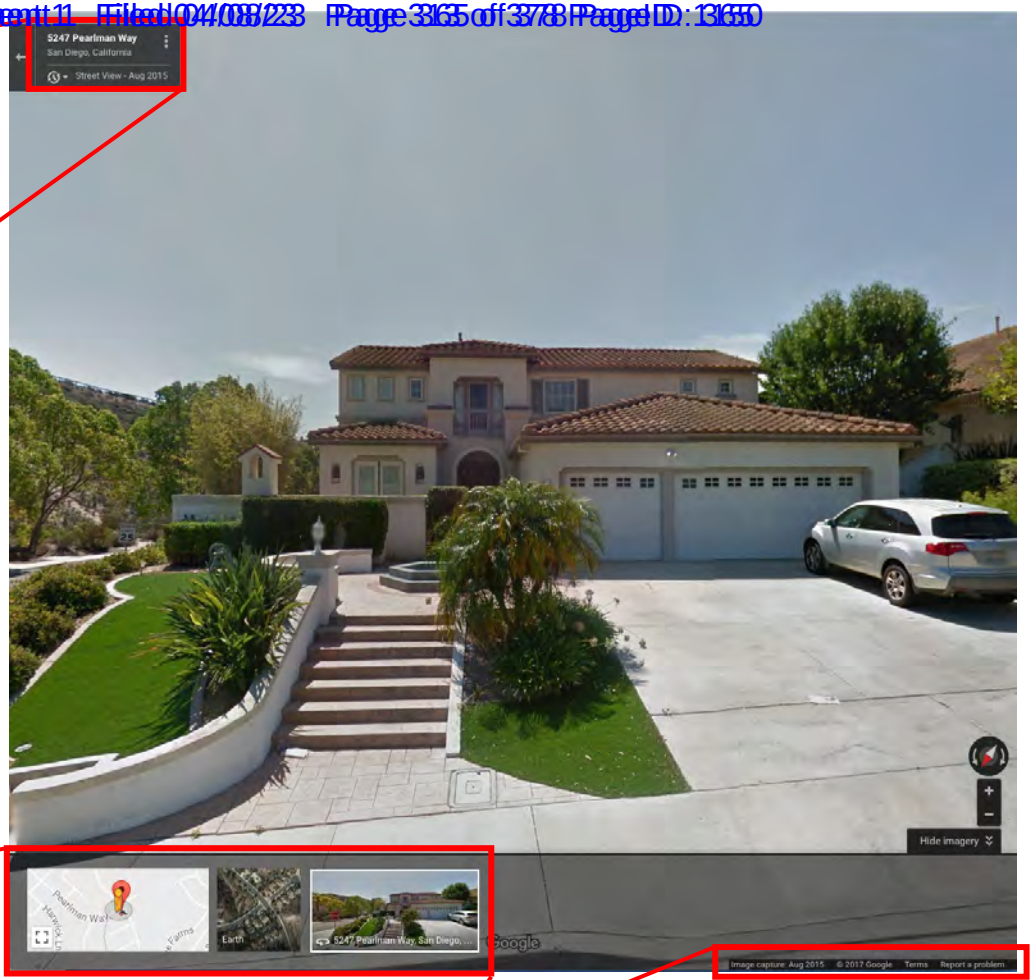
Summary of Concerns with Illumina's missing witnesses: Part 2

- Jian-Bing Fan made a misleading and/or provided false testimony to the USPTO in order to get a foundational Illumina patent issued.

- Jian-Bing Fan tried to re-patent Cornell's invention of a Universal Array, and use of composite oligonucleotides for Affymetrix in June 1999, while already an employee at Illumina, during the same time that Kevin Gunderson and Mark Chee were trying to re-patent Cornell's invention for Illumina.

- Mark Chee and Jian-Bing Fan were the only two Illumina employees tied to abstracts provided at a June 27-29, 2001 meeting at the Ritz-Carlton in Washington DC, where the "hybridization-based decoding procedure" was incontrovertibly linked to "a universal Zipcode array".

Detail1: 5247 Pearlman Way, San Diego CA 92130



Xuyu Cai, lists her current address as Illumina, January 7th, 2015

Neuron
NeuroResource

CellPress

Cell Lineage Analysis in Human Brain Using Endogenous Retroelements

Gilad D. Evrony,[1,2,3,9] Eunjung Lee,[4,5,9] Bhaven K. Mehta,[1,2,3] Yuval Benjamini,[6] Robert M. Johnson,[7] Xuyu Cai,[1,2,3,8] Lixing Yang,[4,5] Psalm Haseley,[4,5] Hillel S. Lehmann,[1,2,3] Peter J. Park,[4,5,10,*] and Christopher A. Walsh[1,2,3,10,*]

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease Research, and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA 02115, USA
[2]Departments of Neurology and Pediatrics, Harvard Medical School, Boston, MA 02115, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA
[5]Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA
[6]Department of Statistics, Stanford University, Stanford, CA 94305, USA
[7]NIH NeuroBioBank, University of Maryland, Baltimore, MD 21201, USA
[8]Present address: Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA
[9]Co-first author
[10]Co-senior author
*Correspondence: peter_park@hms.harvard.edu (P.J.P.), christopher.walsh@childrens.harvard.edu (C.A.W.)
http://dx.doi.org/10.1016/j.neuron.2014.12.028

Neuron *85*, 49–59, January 7, 2015

Illumina's initial filing of a patent on April 28th, 2015, is followed by two more on July 16th, 2015, and December 18th, 2015, where Illumina adds Xuyu Cai as a coinventor.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
62/153,699	04/28/2015		260	IL.MNP008P/IP-1360-PRV		

CONFIRMATION NO. 5458

FILING RECEIPT

22434
Weaver Austin Villeneuve & Sampson LLP
P.O. BOX 70250
OAKLAND, CA 94612-0250

Inventor(s)

Sante Gnerre, Mountain View, CA;
Byoungsok Jung, Atherton, CA;
Emrah Kostem, Redwood City, CA;
Alex Aravanis, San Mateo, CA;

Applicant(s)

Illumina, Inc., San Diego, CA

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
62/193,469	07/16/2015		260	IL.MNP008P2/IP-1360-PRV		

CONFIRMATION NO. 6675

CORRECTED FILING RECEIPT

22434
Weaver Austin Villeneuve & Sampson LLP
P.O. BOX 70250
OAKLAND, CA 94612-0250

Inventor(s)

Sante Gnerre, Mountain View, CA;
Byoungsok Jung, Atherton, CA;
Emrah Kostem, Redwood City, CA;
Alex Aravanis, San Mateo, CA;
Alex So, San Diego, CA;
Xuyu Cai, Natick, MA;
Zhihong Zhang, San Diego, CA;

Applicant(s)

Illumina, Inc., San Diego, CA

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
62/269,485	12/18/2015		260	IL.MNP008P3/IP-1360-PRV3		

CONFIRMATION NO. 8199

FILING RECEIPT

22434
Weaver Austin Villeneuve & Sampson LLP
P.O. BOX 70250
OAKLAND, CA 94612-0250

Inventor(s)

Sante Gnerre, Mountain View, CA;
Byoungsok Jung, Atherton, CA;
Emrah Kostem, Redwood City, CA;
Alex Aravanis, San Mateo, CA;
Alex So, San Diego, CA;
Xuyu Cai, Natick, MA;
Zhihong Zhang, San Diego, CA;

Applicant(s)

Illumina, Inc., San Diego, CA

Title

ERROR SUPPRESSION IN SEQUENCED DNA FRAGMENTS USING REDUNDANT READS WITH UNIQUE MOLECULAR INDICES (UMIS)

Illumina submits provisional application 09/517945 on March 03, 2000, where they claim “adapter sequences” which are zipcodes.

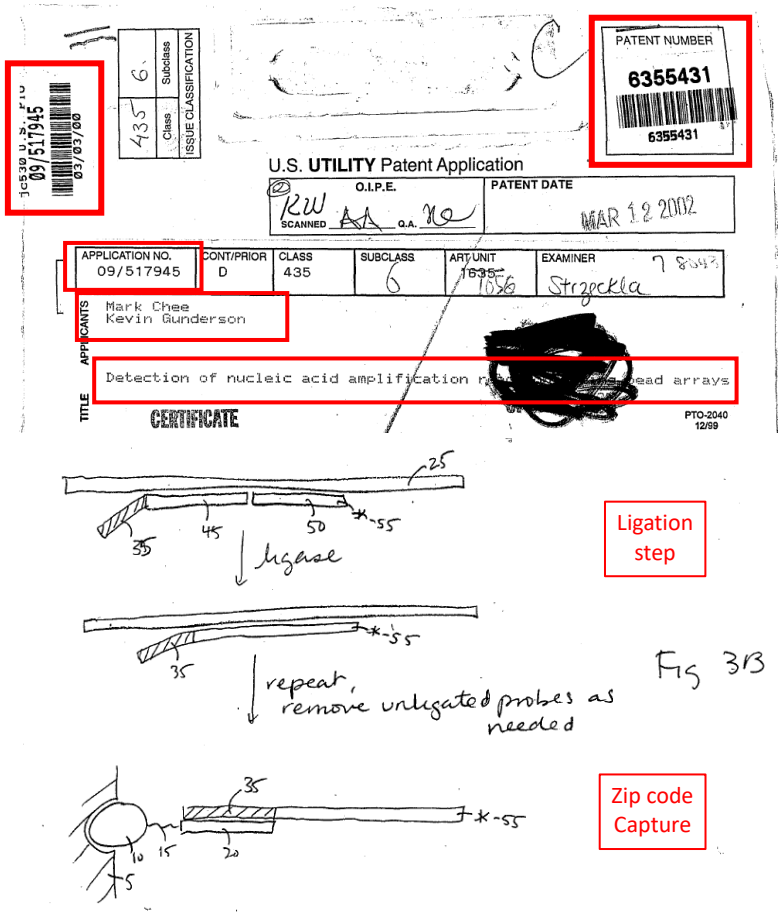


Fig 3B

Figures 3A and 3B depict two preferred embodiments of OLA amplification. Figure 3A depicts a first ligation probe **45** and a second ligation probe **50** with a label **55**. Upon addition of the ligase, the probes are ligated. The reaction can be repeated and then the ligated primer is added to the array as above. Figure 3B depicts the same reaction but using adapter sequences.

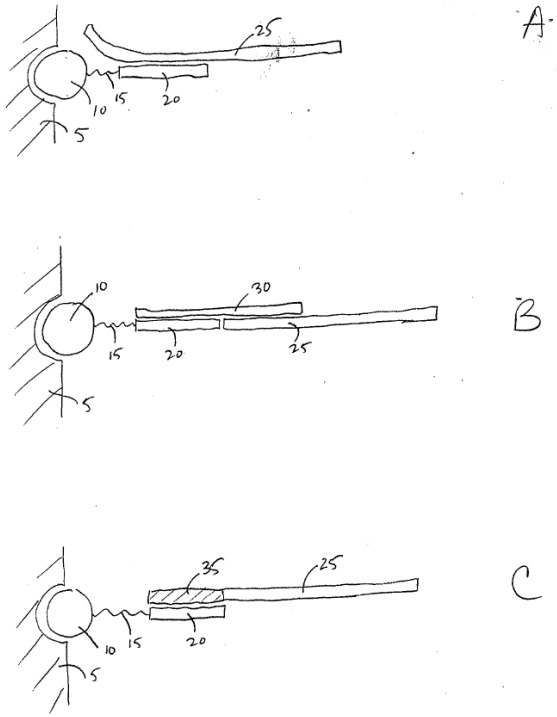


Fig 1

Figures 1A, 1B and 1C depict three different embodiments for attaching a target sequence to an array. The solid support **5** has microsphere **10** with capture probe **20** linked via a linker **15**. Figure 1A depicts direct attachment; the capture probe **20** hybridizes to a first portion of the target sequence **25**. Figure 1B depicts the use of a capture extender probe **30** that has a first portion that hybridizes to the capture probe **20** and a second portion that hybridizes to a first domain of the target sequence **25**. Figure 1C shows the use of an adapter sequence **35**, that has been added to the target sequence, for example during an amplification reaction as outlined herein.

Illumina again tries to re-patent Cornell's ideas from 7 years earlier:

Cornell, Filed Feb 9, 1996

(12) **United States Patent**
Barany et al.

(10) **Patent No.: US 7,083,917 B2**
(45) **Date of Patent: *Aug. 1, 2006**

(54) **DETECTION OF NUCLEIC ACID SEQUENCE DIFFERENCES USING THE LIGASE DETECTION REACTION WITH ADDRESSABLE ARRAYS**

(75) Inventors: **Francis Barany**, New York, NY (US); **George Barany**, Falcon Heights, MN (US); **Robert P. Hammer**, Baton Rouge, LA (US); **Maria Kempe**, Minneapolis, MN (US); **Herman Blok**, Wemeldinge (NL); **Monib Zirvi**, New York, NY (US)

5,143,854	A	9/1992	Pirrung et al.
5,202,231	A	4/1993	Drmanac et al.
5,258,506	A	11/1993	Urdea et al.
5,278,298	A	1/1994	Chakraborty et al.
5,288,468	A	2/1994	Church et al.
5,290,925	A	3/1994	Fino
5,314,809	A	5/1994	Erlich et al.
5,324,633	A	6/1994	Fodor et al.
5,352,582	A	10/1994	Lichtenwalter et al.
5,371,241	A	12/1994	Brush et al.
5,391,480	A	2/1995	Davis et al.
5,405,783	A	4/1995	Pirrung et al.
5,407,798	A	4/1995	Martinelli et al.
5,412,087	A	5/1995	McGall et al.

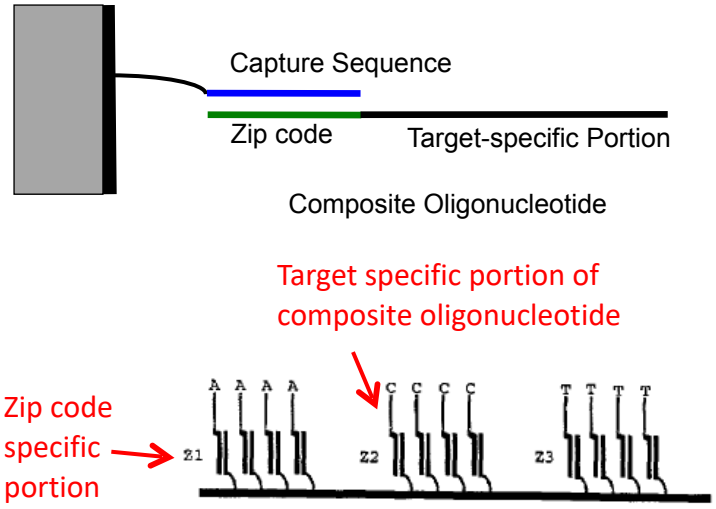


Fig. 7 Top

Illumina Provisional, Filed Dec. 23, 2003

(19) **United States**
(12) **Patent Application Publication**
Gunderson et al.

(10) **Pub. No.: US 2005/0136414 A1**
(43) **Pub. Date: Jun. 23, 2005**

(54) **METHODS AND COMPOSITIONS FOR MAKING LOCUS-SPECIFIC ARRAYS**

(76) Inventors: **Kevin Gunderson**, Encinitas, CA (US); **David Barker**, Del Mar, CA (US); **Mark Chee**, Del Mar, CA (US); **Tim McDaniel**, San Diego, CA (US); **Robert Yang**, San Diego, CA (US)

Correspondence Address:
PILLSBURY WINTHROP SHAW PITTMAN LLP
ATTENTION: DOCKETING DEPARTMENT
11682 EL CAMINO REAL, SUITE 200
SAN DIEGO, CA 92130 (US)

(21) Appl. No.: **10/746,249**

(22) Filed: **Dec. 23, 2003**

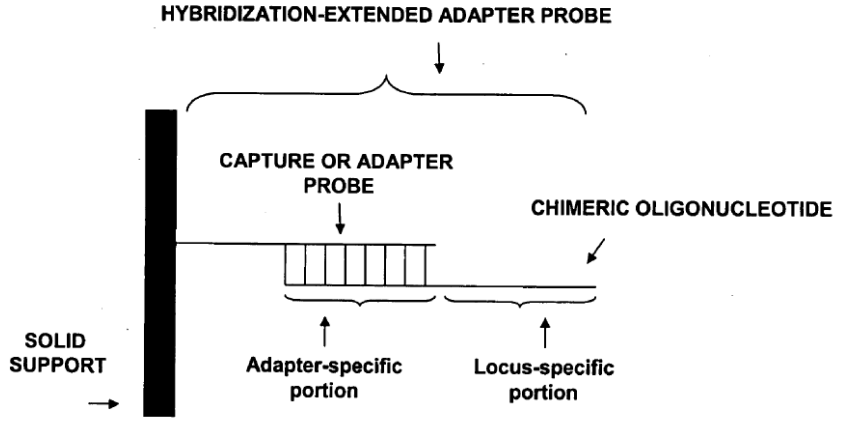
Publication Classification

(51) **Int. Cl.**[7] **C12Q 1/68**; C12P 19/34; C12M 1/34

(52) **U.S. Cl.** **435/6**; 435/91.2; 435/287.2

(57) **ABSTRACT**

The present invention includes methods and compositions relating to locus-specific arrays. More specifically, this invention includes methods for making locus-specific arrays from universal arrays in situ, the custom arrays made using those methods, and methods of using the custom arrays to detect target nucleotides.



Illumina's own patent application – filed 7 years after Cornell's zip code patent was filed – claims use of a zip code array, and "chimeric oligonucleotide" comprised of an "Adapter-specific portion" and a "Locus-specific portion" for making locus-specific arrays – which is identical to the Cornell teachings, and equivalent to an infinium array .

Cornell '293 family patents are 2 years ahead of Illumina.

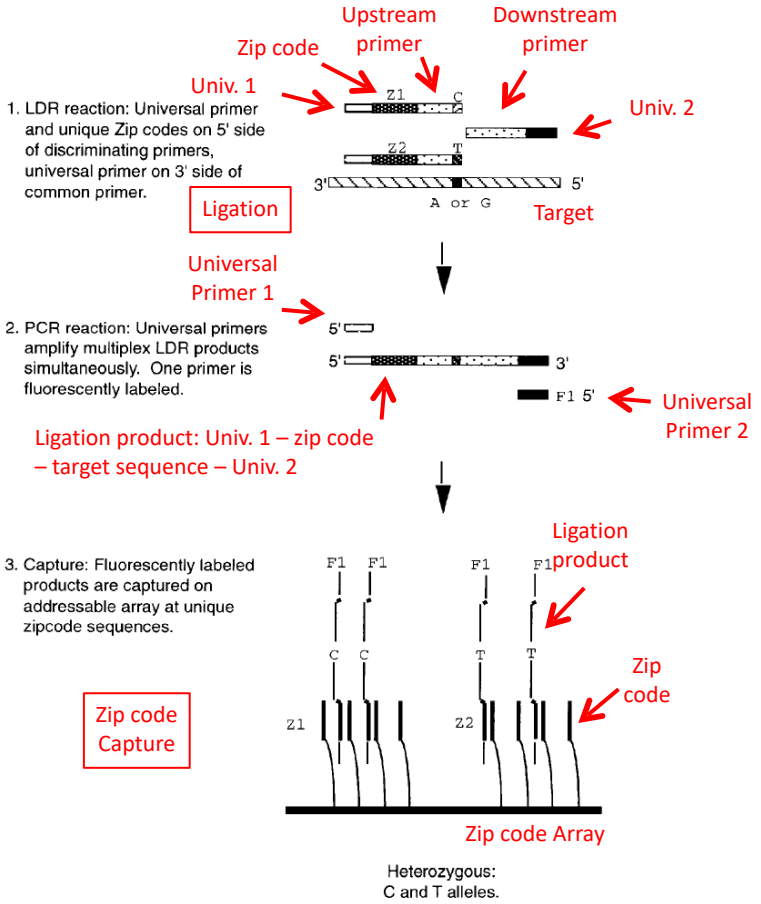
Cornell, Filed Jan 6, 1999

(12) **United States Patent**
Barany et al.

(10) **Patent No.:** **US 6,534,293 B1**
(45) **Date of Patent:** ***Mar. 18, 2003**

(54) **ACCELERATING IDENTIFICATION OF SINGLE NUCLEOTIDE POLYMORPHISMS AND ALIGNMENT OF CLONES IN GENOMIC SEQUENCING**

LDR/PCR with Addressable Array Capture



Illumina Provisional, Filed Feb. 7, 2001

(19) **United States**
(12) **Patent Application Publication**
Fan et al.

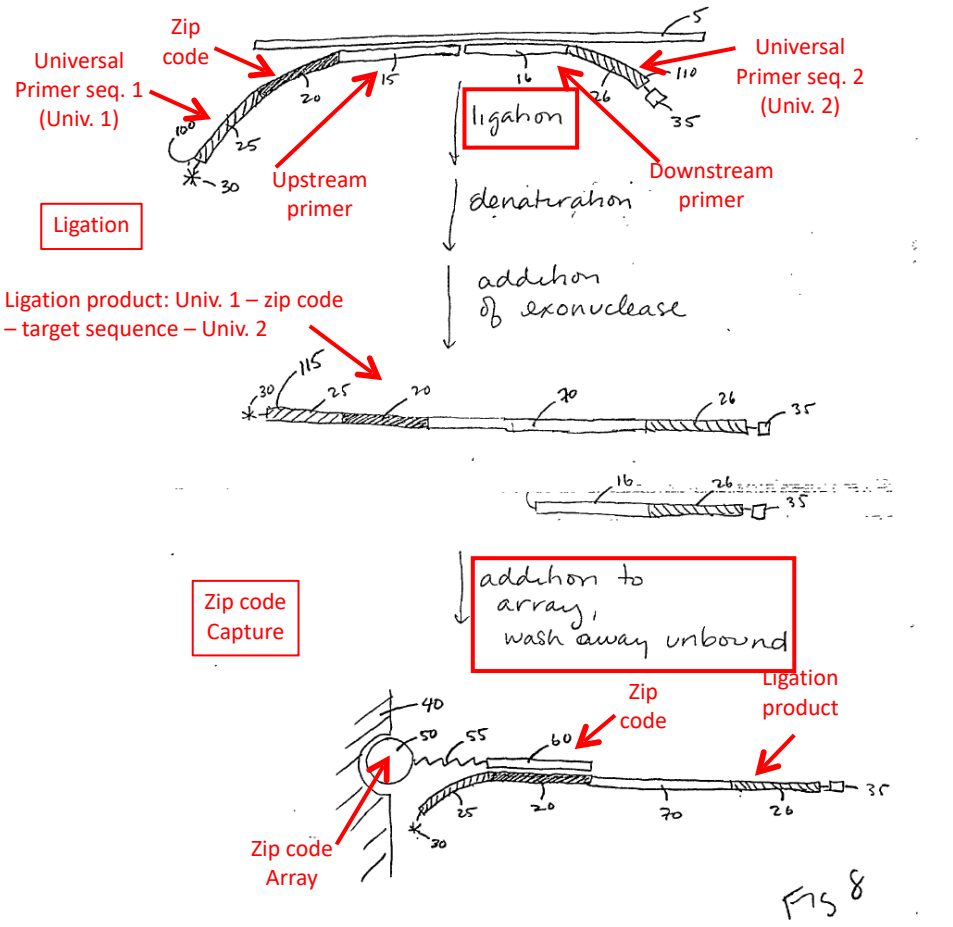
(10) **Pub. No.: US 2002/0172946 A1**
(43) **Pub. Date: Nov. 21, 2002**

(54) **NUCLEIC ACID DETECTION METHODS USING UNIVERSAL PRIMING**

(76) Inventors: **Jian-Bing Fan**, San Diego, CA (US); **Xiang-Dong Fu**, San Diego, CA (US)

Related U.S. Application Data

(60) Provisional application No. 60/180,810, filed on Feb. 7, 2000. Provisional application No. 60/234,731, filed on Sep. 22, 2000.



Subpoena sent by Affymetrix to Illumina to depose Jian-Bing Fan on November 16, 2005 at the Hilton San Diego Airport Hotel.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

AFFYMETRIX, INC.,	)	
	)	
Plaintiff/Counter-Defendant,	)	C. A. No. 04-901 (JJF)
	)	
	)	
	)	
ILLUMINA, INC.,	)	
	)	
Defendant/Counter-Plaintiff.	)	

NOTICE OF DEPOSITION OF JIAN-BING FAN

TO ALL PARTIES AND THEIR COUNSEL OF RECORD:

PLEASE TAKE NOTICE that, pursuant to Fed. R. Civ. P. 30, plaintiff Affymetrix, Inc., through their attorneys, will take the deposition upon oral examination of Jian-Bing Fan beginning at 9:30 a.m. on November 16, 2005 at the Hilton San Diego Airport/Harbor Island Hotel, 1960 Harbor Island Drive, San Diego, California, 92101.

The deposition will be recorded stenographically, with the capacity to provide instant visual display or playback of the testimony, and by video by an officer authorized to administer oaths. All counsel are invited to attend and examine the witness.

Illumina knew exactly where Mark Chee was in January 2017.

Garvan Institute of Medical Research

Leaders in Science & Society

GARVAN INSTITUTE

Dr Mark Chee

Chief Executive Officer
Prognosys Biosciences

"A perspective on Genomics Technologies"

Monday 23 January 2017 12PM, AUDITORIUM

Host: Prof Chris Goodnow

Mark Chee received his B.Sc. in Biochemistry from the University of New South Wales and his Ph.D. in Molecular Biology from the University of Cambridge. Dr. Chee is currently CEO of Prognosys Biosciences and is involved in developing next-generation protein assay technologies. He co-founded Illumina, where he served as VP of Genomics, and prior to that was Director of Genetics Research at Affymetrix. He also serves as a scientific advisor to Complete Genomics, Inc.

EXHIBIT 14

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF NEW JERSEY

MONIB ZIRVI, M.D., Ph.D. Plaintiff, v. ILLUMINA INC., THERMO FISHER SCIENTIFIC, GUMP STRAUSS HAUER & FELD LLP, LATHAM & WATKINS, RIP FINST, SEAN BOYLE, MATTHEW A. PEARSON, ANGELA VERRECCHIO, ROGER CHIN, and DOUGLAS LUMISH Defendants.	)	**AFFIDAVIT OF MERIT**

I, Jeffrey B. Oster, submit this Affidavit of Merit pursuant to N.J.S.A. 2A:53A-27, which provides that "the person executing the Affidavit shall be licensed in this or any other State; have particular expertise in the general area of specialty involved in the action…for a period of at least five years".

1. I am an attorney currently admitted in Washington (State) to which I was admitted on May 31, 1988. I was admitted as an attorney in New York in 1986 and I believe I'm currently listed as an out-of-state attorney in New York. And I was admitted in New Jersey (in 1986) where I resided at the time, but did not renew my admission in New Jersey after relocating to Washington in late 1987.

2. I am admitted at the US Patent and Trademark Office (Registration No. 32,585).

3. I have 37 years of experience in both law firm and in-house positions practicing law, including: (1) IP litigation in Federal Courts; (2) post-grant procedures at the PTAB (Patent Trial and Appeal Board) and predecessor administrative panels including successfully defending patent validity challenges filed by defendant Illumina (both Zebala/Syntrix patent and The Scripps

Research Institute DNA bar coding patent) and defendant Thermo Fisher Scientific (Life Technologies) for the Black Hole Quencher patents and various second generation PCR patents; (3) the Opposition Division of the European Patent Office (dual citizen of the United States and Germany) including multiple European oppositions filed against Thermo Fisher Scientific (Affymetrix); (4) in patent preparation and prosecution matters: (5) and general legal corporate and securities matters, including a dissident directors slate proxy campaign opposed to Latham & Watkins (defendants Chin and Lumish) representing existing Board of Sorrento.

4. My more recent legal experience in federal courts includes, but is not limited to, providing intellectual property expertise in: (i) The *Syntrix Biosystems, Inc. v. Illumina Inc*, W.D. Wash No. 3_10-cv-05870, at the time, the largest patent litigation damage award in Washington, (ii) *The Scripps Research Institute v. Illumina* S.D. Calif No. 3:16-cv-00661-JLS, where one of the inventors won a Nobel Prize, including being the sole Attorney for Scripps on IPR PR2016-01619; and (iii) *Troll Busters LLV v. Roche Applied Sciences et al.* S.D. Calif No. 3:-cv-00056 wherein both Thermo Fisher and Life Technologies were defendants (case dropped when lobbyist (upon information and belief, it was a Life Technology lobbyist) inserted a retroactive clause in a provision of the America Invents Act of 2012). Thus, I have the experience to evaluate this professional malpractice claim for reasonability.

5. I received a B.A. degree from Johns Hopkins University, Ph.D. degree in pharmacology from the University of Pennsylvania, and a JD degree from Rutgers Law School Newark (evening program).

6. I have reviewed the complaint for legal malpractice in this case as it relates to the Plaintiff, Dr. Monib Zirvi's claims that, as an inventor of intellectually property at the center of the case of *Cornell et al v Illumina*, Case 1:10-cv-00433-LPS-MPT, the actions and legal advice by the lawyers who represented Dr. Zirvi fell outside the acceptable professional standards for Attorneys engaged in the practice of law for a patent infringement case in a Federal Court.

7. Upon information and belief Dr. Zirvi has made allegations that state a claim, which evidences a reasonable probability that the licensed attorney Defendants' actions deviated from accepted standards of care for an attorney representing Dr. Zirvi in *Cornell et al v Illumina* ("The Case"). Dr. Zirvi's attorneys breach of professional conduct included:

a. Upon information and belief, all the inventors, including Dr. Zirvi, were told that they did not need separate representation because the inventors' interest were aligned with

the Thermo Fisher (then Life Technologies as licensee of Cornell) for The Case. The defendant lawyers should have advised the inventors of a clear conflict of interest either before indicating the need for separate representation or during the representation when the defendant lawyers were clearly aware of the negotiation and collaboration going on between Life Technologies (The Case plaintiff) and Illumina (The Case defendant) and that there may be reasons for retaining separate counsel to advise them during the litigation.

b. During the litigation of The Case and upon information and upon belief, the defendant lawyers made representations and took actions that established an attorney client relationship including providing documents marked "Privileged" for The Case plaintiff, Life Technologies (now Thermo Fisher), for whom all of the lawyer defendants worked for as either outside counsel or in-house counsel. Moreover, upon information and belief, the Thermo Fisher lawyers would not allow separate representation for Cornell (licensor) or the inventors (including the current Plaintiff) as no other lawyer would be allowed to participate, or view discovery, under a protective order.

c. Upon information and belief, throughout the representation the defendant lawyers were provided access to discovery or The Case and/or in "The Other Case" (*Illumina, Inc. v. Life Technologies Corporation et al*, 3_11-cv-03022). This "Zip Code Discovery" was not shared or even brought to the attention of Dr. Zirvi. Upon information and belief the Zip Code Discovery indicated Illumina's use of Zip Codes, including electronic files and other evidence that Illumina was using the identical Zip Codes invented by Dr. Zirvi internally for years. The defendant lawyers, who were also representing Dr. Zirvi, never brought the issue to Dr. Zirvi, or even acted in a professional manner if this created a clear conflict of interest.

d. During the course of defendant lawyers alleged representation of Dr. Zirvi, there was an ongoing secret joint "Ampliseq for Illumina" project, which involved a collaboration between Illumina and Thermo Fisher Scientific. This information was discovered by Dr. Zirvi after The Case was settled in light of public disclosures by Illumina of this secret joint project happening at a time before the start of the Markman hearing in The Case. Upon information and belief, the defendant lawyers failed to disclose this conflict of interest and act in accordance with this conflict of interest.

e. Upon information and belief, defendant Matthew Pearson (of Aiken Gump, outside counsel for Thermo Fisher at the time) falsely stated to Cornell inventors and Dr. Zirvi that the unredacted First Amendment Agreement between Illumina and P.E. Biosystems (now Thermo Fisher) was not relevant to The Case or the inventors. Upon information and belief, the unredacted First Amendment Agreement was not provided to the Cornell inventors or Dr. Zirvi at any time before, during or after the Markman Hearing of The Case. when it clearly states Tag Sequences are defined as described in WO97/31256, which Plaintiff is a co-inventor of.

f. Upon information and belief, new Council of Record for The Case, Roger Chin and Douglas Lumish of Latham Watkins, were simultaneously representing Thermo Fisher in The Other Case (*Illumina, Inc. v. Life Technologies Corporation et al*, 3_11-cv-03022). Upon Information and belief defendants Chin and Lumish failed to advise their client Dr. Zirvi that (1) they had a conflict of interest; and (2) the unredacted First Amendment Agreement was produced as part of the Zip Code Discovery and admits that the term "Tag Sequences" are defined as described in Dr. Zirvi's co-invention WO97/31256, which is a clear admission against interest by defendant Illumina and evidences a failure of zealous representation of Dr. Zirvi and failure to disclose a conflict of interest between Thermo Fisher and Dr. Zirvi.

g. Upon information and belief, in-house counsel defendant Rip Finst independently and without outside counsel filed an incorrect document with the Court in The Case *Cornell et al v Illumina*, Case 1:10-cv-00433-LPS-MPT. Upon information and belief, never before and never after in The Case did Rip Finst file a similar document in the Court without outside counsel. Upon information and belief, this was also done without informing or consulting the Cornell inventors or Dr. Zirvi, which was the procedure for other documents filed by outside counsel in The Case.

h. Lastly, upon information and belief and without consulting their clients, the Cornell inventors and Dr. Zirvi, the defendant lawyers settle The Case on terms that provided no compensation to Cornell or Dr. Zirvi. Upon information and belief, at no point in time was Dr. Zirvi advised of the settlement negotiations, asked to participate or be represented in the settlement negotiations, advised of any secret quid pro quos as consideration for the settlement of The Case or ever advised of any attorney conflicts of interest on the part of the defendant attorneys.

8. In my opinion and upon the information provided to me, the clear conflict of interest of the lawyer defendants as attorneys for The Case and The Other Case, obfuscation of The Case discovery information in light of the discovery evidence from The Other Case, the failure to act for the interests of Dr. Zirvi and misleading statements are a deviation from the standard of care, lack of zealous representation, and conflicts of interest attorneys are required to abide by.

9. I have no financial interest in the outcome of this case. But, in view of the Thermo Fisher (Life Technologies at the time) lobbying activity to add a retroactive provision to the America Invents Act (which if done at my other country of citizenship, the German Parliament, would have been a violation of the Foreign Corrupt Practices Act), I would like to see justice served here.

10. Accordingly, in my professional opinion, within reasonable probability, the legal services rendered by the Thermo Fisher in-house counsel Rip Finst, Matthew Pearson of Akin Gump, Roger Chin of Latham & Watkins and Douglas Lumish of Latham & Watkins are well below the acceptable professional standards of care and ethical requirements.

Signed: Jeffrey B. Oster Dated: March 24 2023

Jeffrey B. Oster, Ph.D., J.D.
WSBA # 17709

Notary:

State of Washington
County of King
Signed or attested before me on 3/24/2023
by Jeffrey B. Oster

Signature Nathan Campbell
(Printed name) Notary
Title
My appointment expires 06/17/2024

Notary Public
State of Washington
NATHAN CAMPBELL
LICENSE # 20108757
MY COMMISSION EXPIRES
JUNE 17, 2024

EXHIBIT 6

LATHAM & WATKINS LLP

Cornell et al. v. Illumina 10-cv-433 (D. Del.)

Case Assessment

PRIVILEGED & CONFIDENTIAL

February 17, 2017

Patents and Accused Products

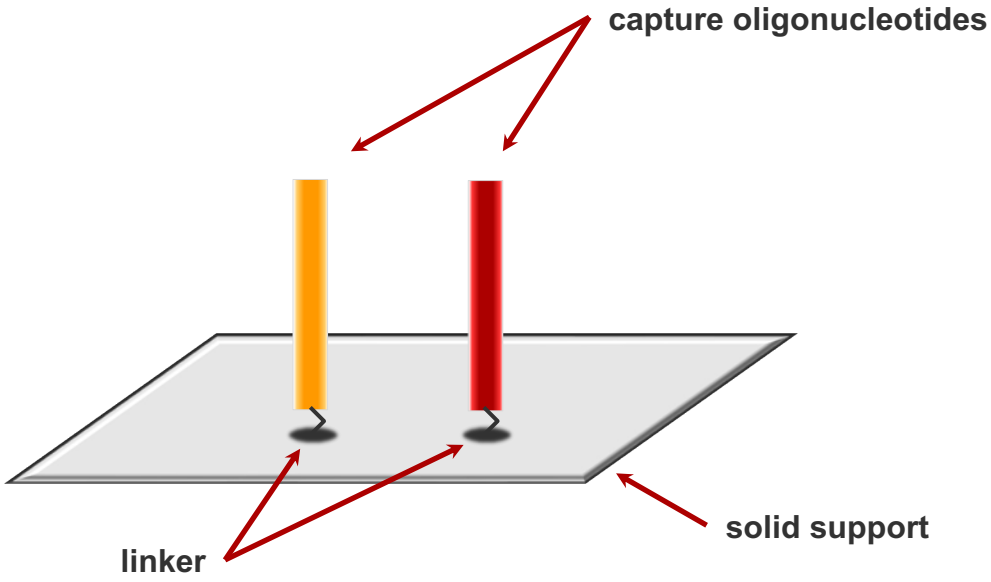
Array Patents

- Infinium
- GoldenGate/DASL

LDR-PCR Patents

- GoldenGate/DASL

Array Patents



1. A device comprising:

   a **solid support** having an array of positions each suitable for attachment of an oligonucleotide probe;

   a **linker** suitable for coupling an oligonucleotide probe to the solid support and attached to the solid support at each of the array positions; and

   an array of **capture oligonucleotide probes** on the solid support at the array positions, said capture oligonucleotide probes each having greater than sixteen nucleotides and able to bind to complementary nucleic acids at uniform hybridization conditions, wherein each capture oligonucleotide probe of the array differs in sequence from its adjacent capture oligonucleotide probe, when aligned to each other by at least 25% of the nucleotides.

Array Patents – Post-Markman

1. A device comprising:

a solid support having an array of positions each suitable for attachment of an oligonucleotide probe;

a linker suitable for coupling an oligonucleotide probe to the solid support and attached to the solid support at each of the array positions; and

an array of capture oligonucleotide probes on the solid support at the array positions, said capture oligonucleotide probes each having greater than sixteen nucleotides and able to bind to complementary nucleic acids at uniform hybridization conditions, wherein each capture oligonucleotide probe of the array differs in sequence from its adjacent capture oligonucleotide probe, when aligned to each other by at least 25% of the nucleotides.

MARKMAN

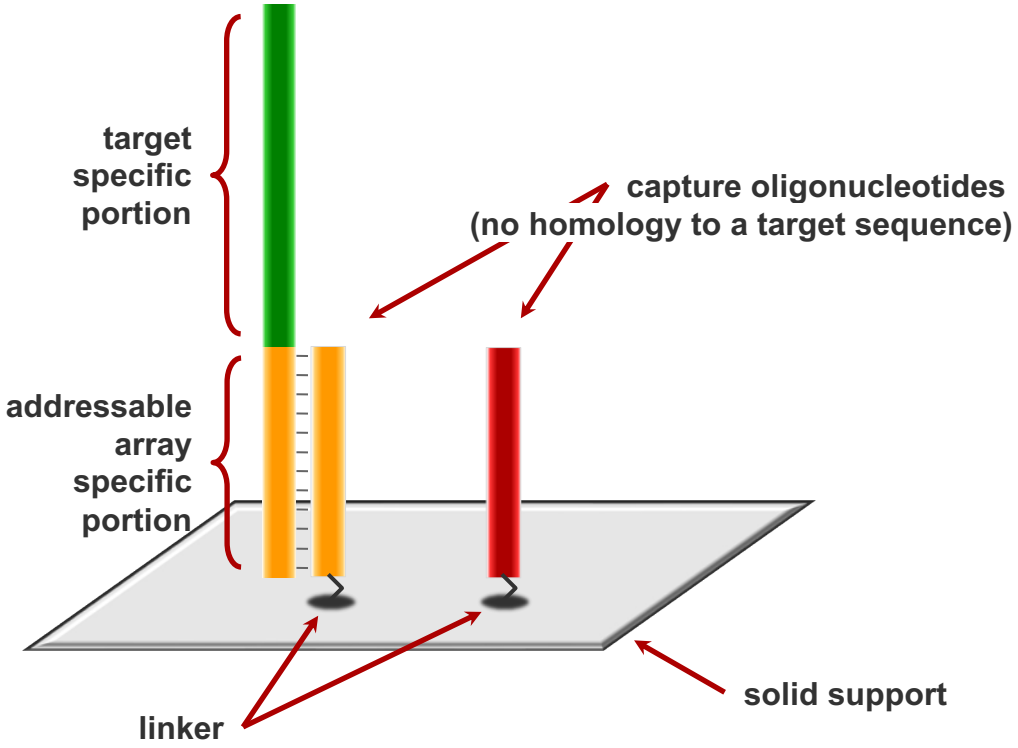
1. A device comprising:

a **solid-phase substrate** having an array of positions each **capable of being connected** of an oligonucleotide probe;

a **feature that connects an oligonucleotide to a solid support** suitable for coupling an oligonucleotide probe to the solid support and connected to the solid support at each of the array positions; and

an array of **oligonucleotide probes which has no homology to a target sequence and is complementary to the addressable array specific portion of an oligonucleotide containing an addressable array specific portion and a target specific portion** on the solid support at the array positions, said capture oligonucleotide probes each having greater than sixteen nucleotides and able to bind to complementary nucleic acids at uniform hybridization conditions, wherein **each capture oligonucleotide probe of the array differs in sequence from every adjacent capture oligo-nucleotide probe, when aligned to each other by at least 25% of the nucleotides**.

Array Patents – Post-Markman



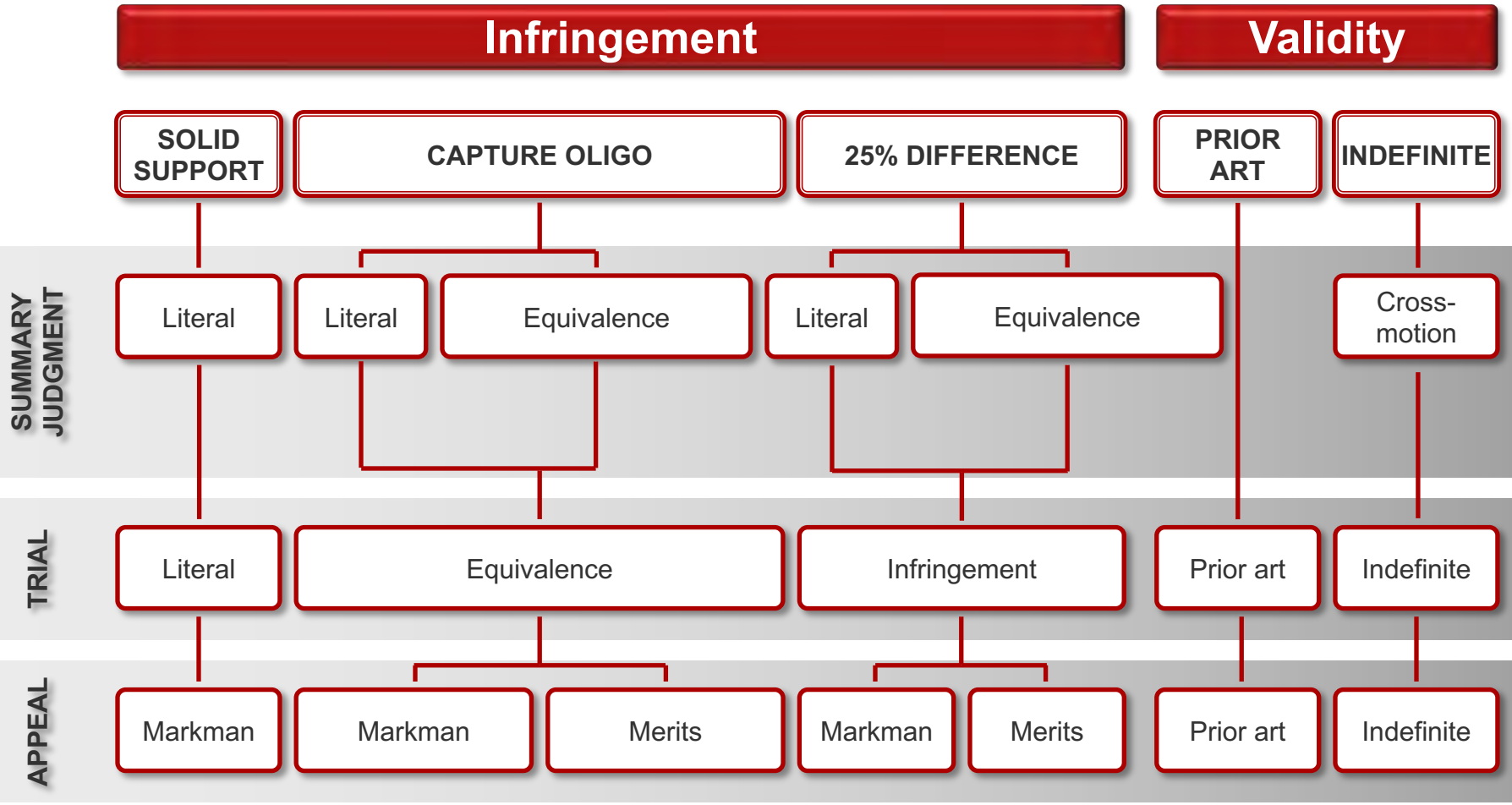
1. A device comprising:

   a solid-phase substrate having an array of positions each capable of being connected of an oligonucleotide probe;

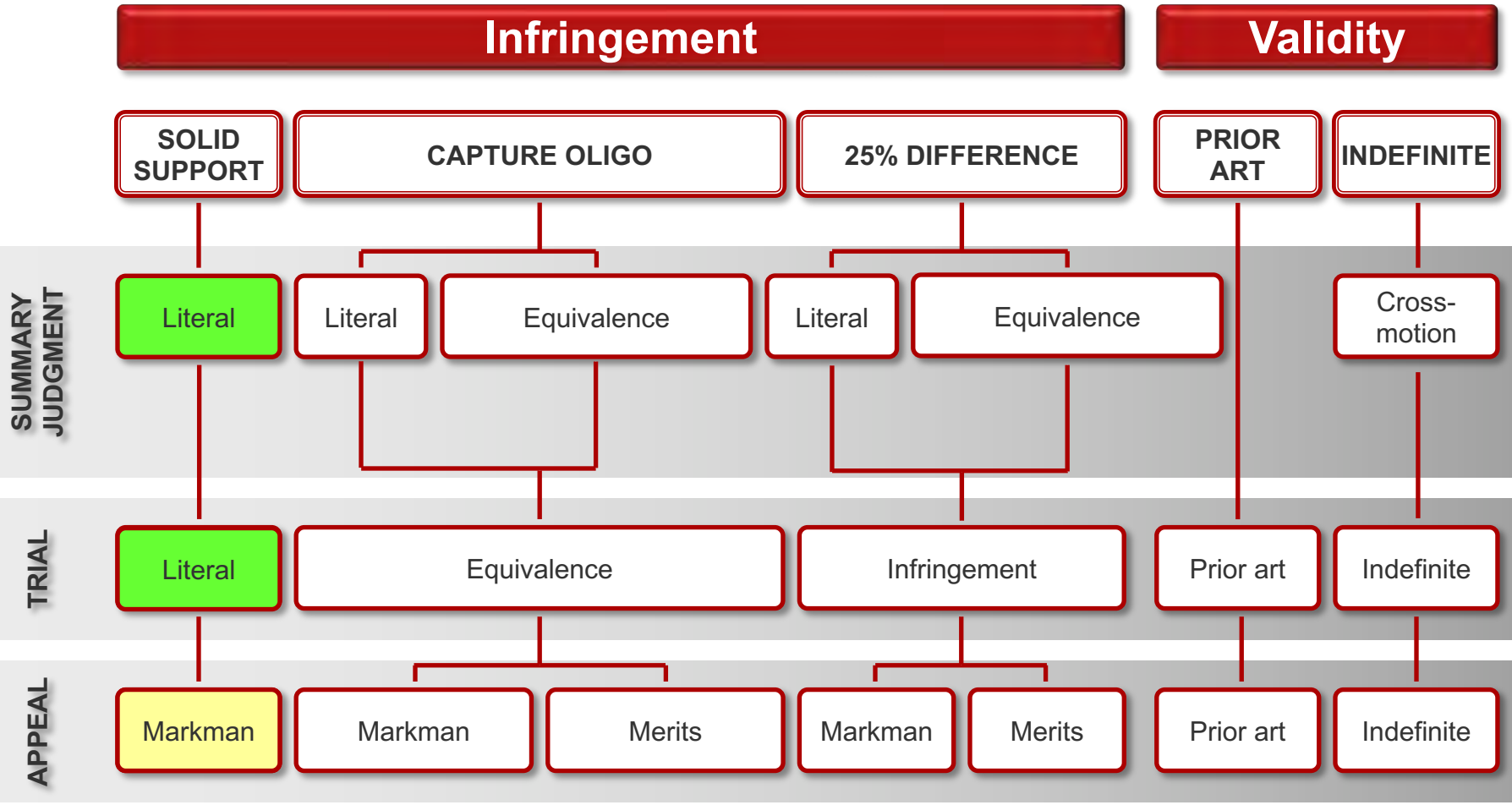
   a feature that connects an oligonucleotide to a solid support suitable for coupling an oligonucleotide probe to the solid support and connected to the solid support at each of the array positions; and

   an array of oligonucleotide probes which has **no homology to a target sequence** and is **complementary to** the addressable array specific portion of **an oligonucleotide containing** an addressable array specific portion and **a target specific portion** on the solid support at the array positions, said capture oligonucleotide probes each having greater than sixteen nucleotides and able to bind to complementary nucleic acids at uniform hybridization conditions, wherein each capture oligonucleotide probe of the array differs in sequence from **every** adjacent capture oligo-nucleotide probe, when aligned to each other by at least 25% of the nucleotides.

Infinium – Array Patents



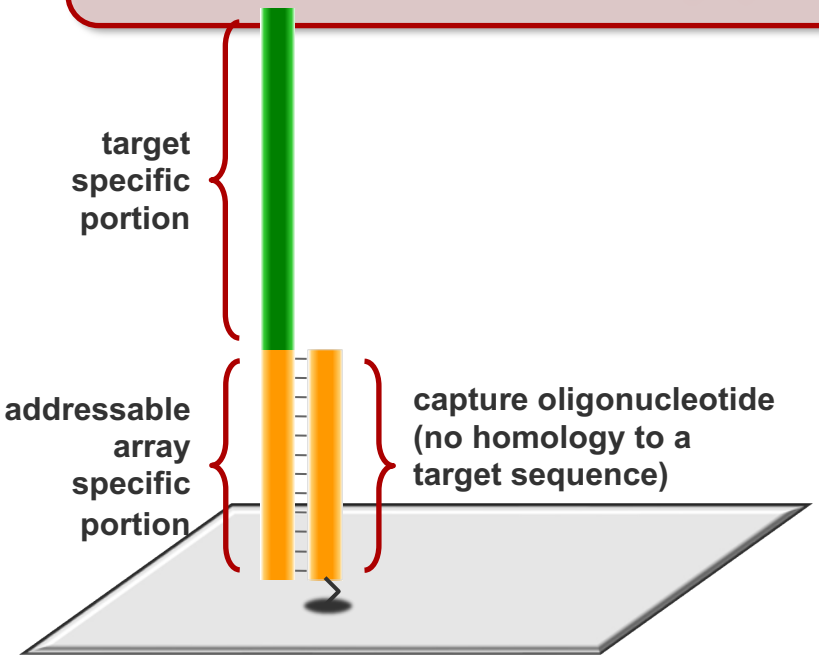
Infinium – Array Patents



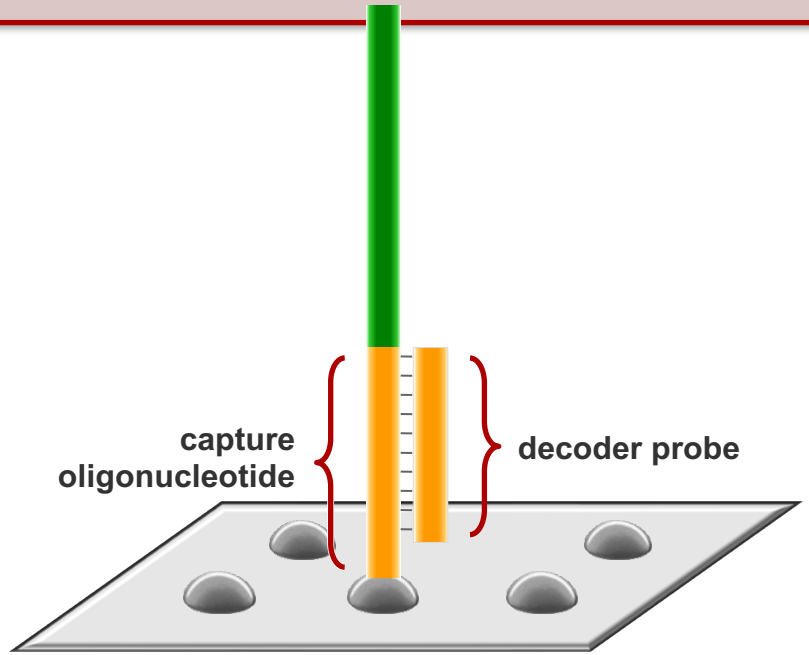
Infinium – Array Patents

CAPTURE OLIGONUCLEOTIDE

"Oligonucleotide probe which has **no homology to a target sequence** and is complementary to the addressable array specific portion of an oligonucleotide containing an addressable array specific portion and a **target specific portion**"



ARRAY PATENT



ILLUMINA INFINIUM ARRAY

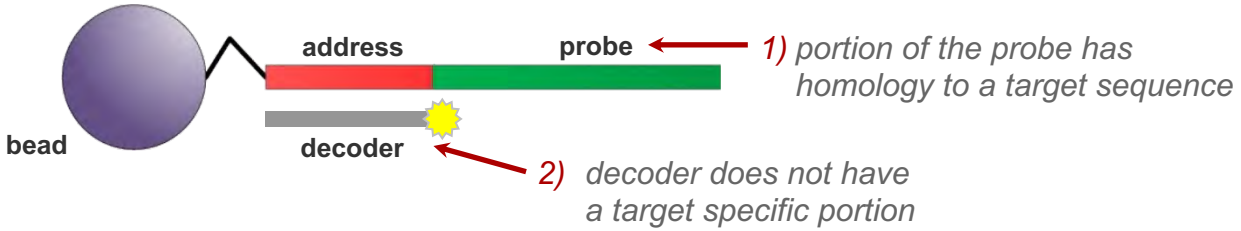
Infinium – Array Patents

CAPTURE OLIGONUCLEOTIDE

"Oligonucleotide probe which has **no homology to a target sequence** and is complementary to the addressable array specific portion of an oligonucleotide containing an addressable array specific portion and a **target specific portion**"

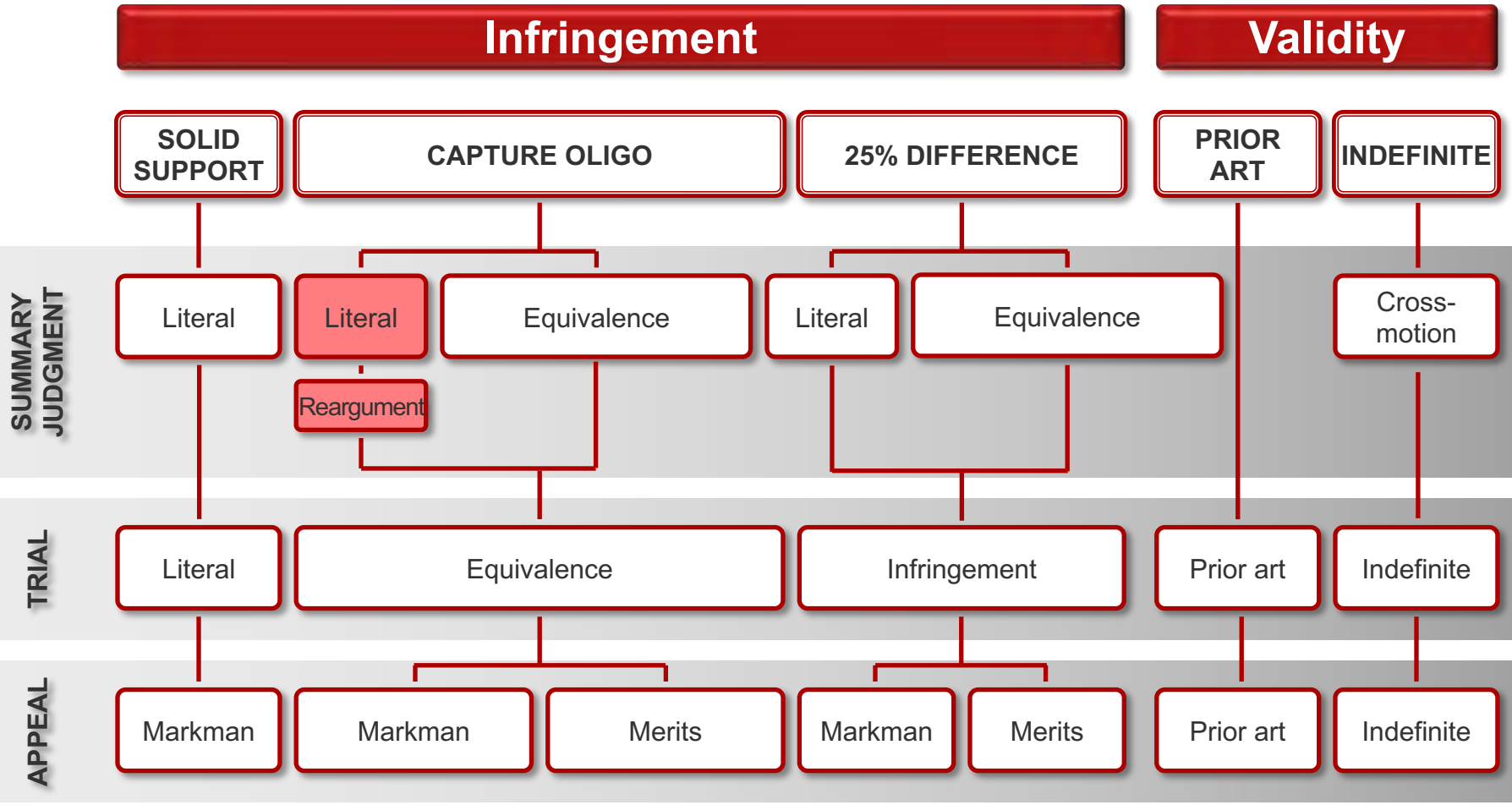
LITERAL INFRINGEMENT

- **Literal infringement is not asserted under the current claim construction**



- **Motion for reargument is best shot at literal infringement, but reargument is granted very rarely in the District of Delaware**

Infinium – Array Patents



Infinium – Array Patents

CAPTURE OLIGONUCLEOTIDE

"Oligonucleotide probe which has **no homology to a target sequence** and is complementary to the addressable array specific portion of an oligonucleotide containing an addressable array specific portion and a **target specific portion**"

DOCTRINE OF EQUIVALENTS

- **Illumina advances two *legal* restrictions on the doctrine of equivalents:**
  1) **Argument-based estoppel:** a competitor would reasonably believe that the applicant had surrendered the relevant subject matter
  2) **Claim vitiation:** no reasonable jury could determine two elements to be equivalent
- **If we avoid legal restriction, there is a reasonable *factual* argument for equivalence**

Infinium – Array Patents

CAPTURE OLIGONUCLEOTIDE

"Oligonucleotide probe which has **no homology to a target sequence** and is complementary to the addressable array specific portion of an oligonucleotide containing an addressable array specific portion and a **target specific portion**"

DOCTRINE OF EQUIVALENTS

ESTOPPEL

VITIATION

Argument-based estoppel: a competitor would reasonably believe that the applicant had surrendered the relevant subject matter (must be clear and unmistakable)

The present invention "avoids all of the aforementioned problems associated with typical hybridization arrays (i.e., target-capture probe cross-hybridization and false-positive/negative signal generation)."

Infinium – Array Patents

CAPTURE OLIGONUCLEOTIDE

"Oligonucleotide probe which has **no homology to a target sequence** and is complementary to the addressable array specific portion of an oligonucleotide containing an addressable array specific portion and a **target specific portion**"

DOCTRINE OF EQUIVALENTS

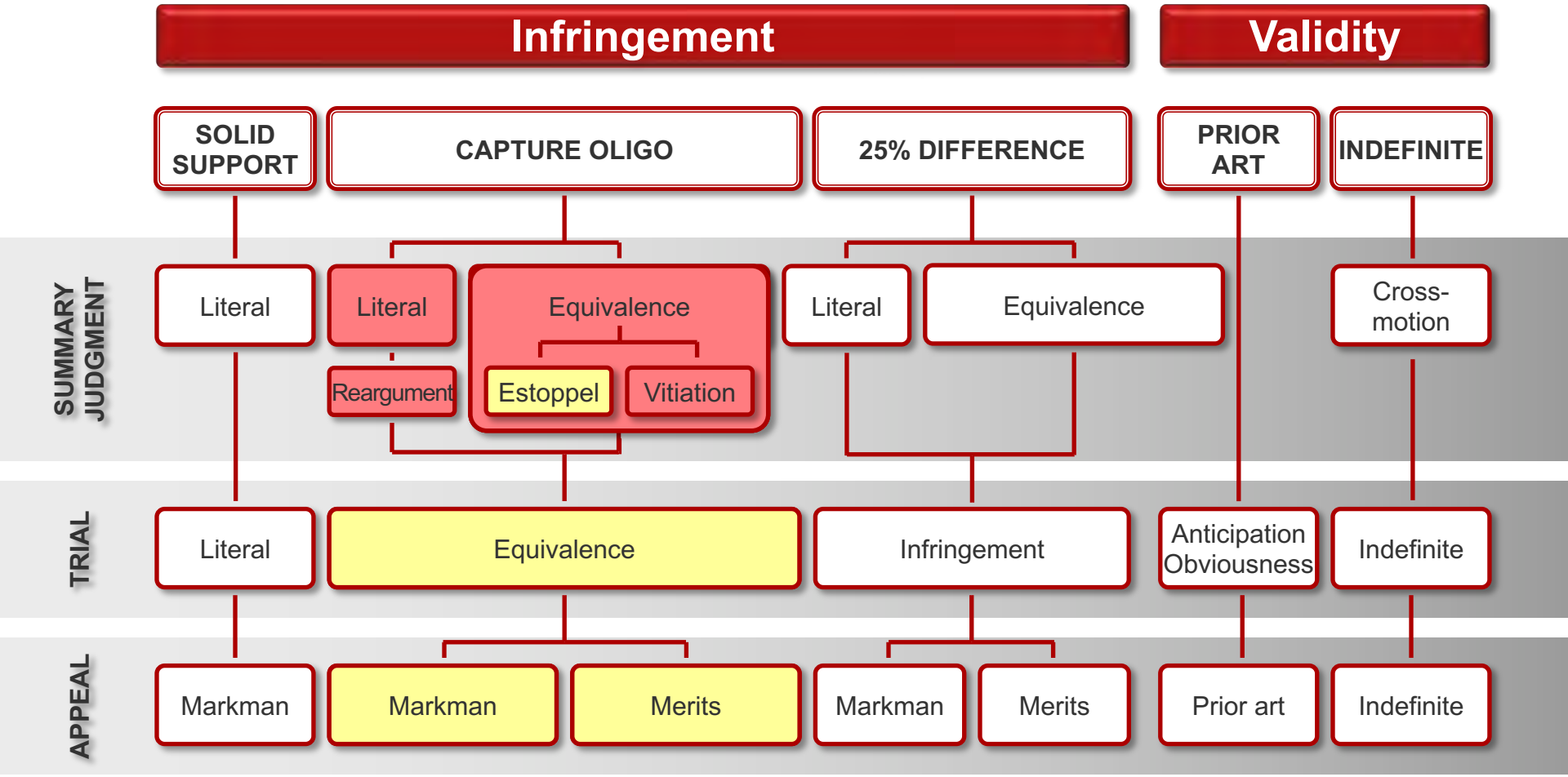
ESTOPPEL

VITIATION

Claim vitiation: no reasonable jury could determine two elements to be equivalent

*Illumina will argue that "an oligonucleotide containing an addressable array specific portion **and a target specific portion**" is the antithesis of "an oligonucleotide containing an addressable array specific portion **without a target specific portion**"*

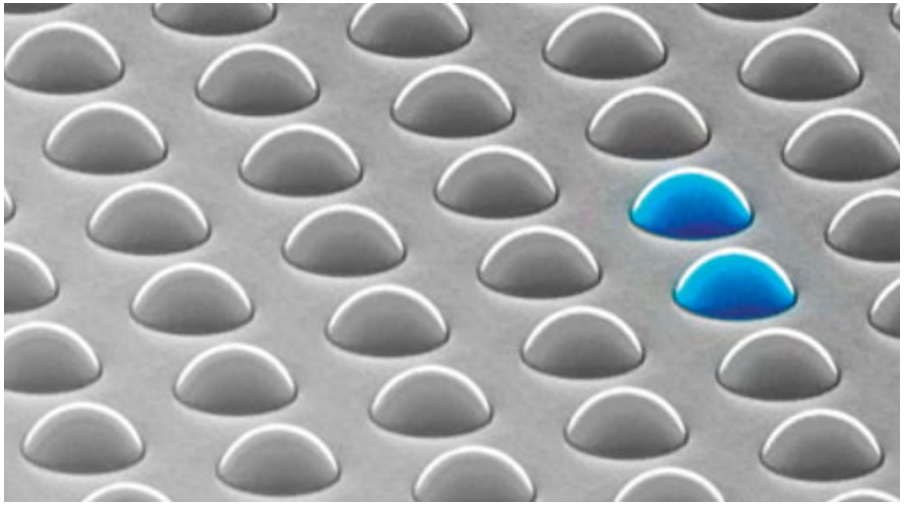
Infinium – Array Patents

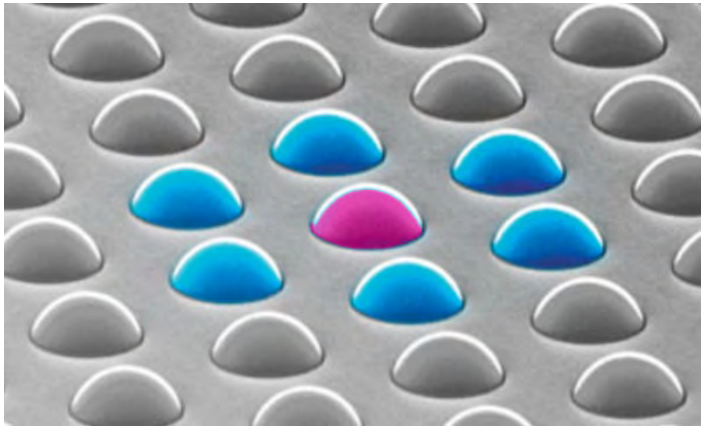


Infinium – Array Patents

25% difference

"**each** capture oligonucleotide probe of the array **differs** in sequence **from every** adjacent capture oligonucleotide probe, when aligned, by at least 25% in nucleotide sequence"





Infinium – Array Patents

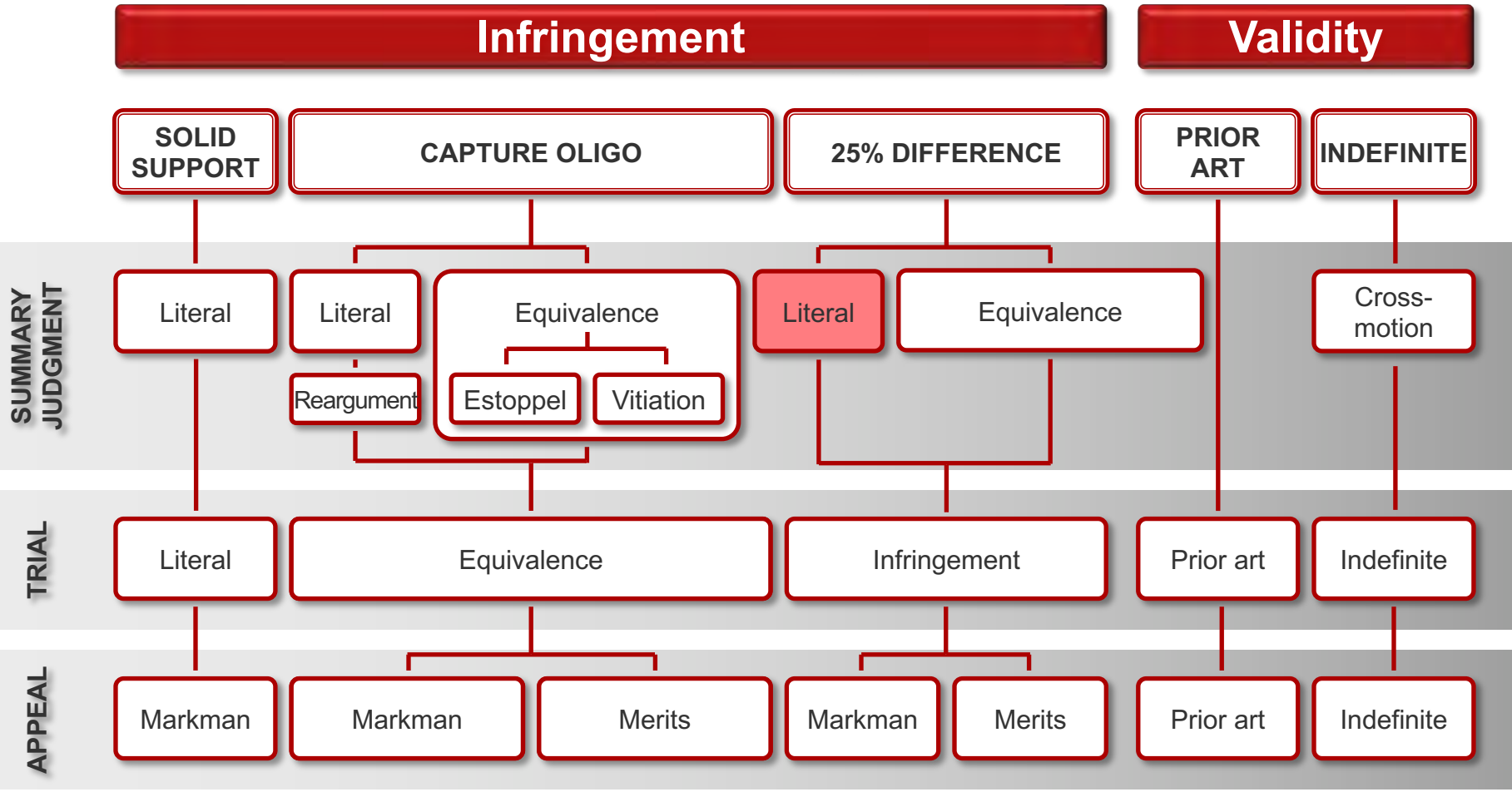
25% DIFFERENCE

"**Each** capture oligonucleotide probe of the array **differs** in sequence **from every** adjacent capture oligonucleotide probe, when aligned, by at least 25% in nucleotide sequence."

LITERAL INFRINGEMENT

- **Likely no literal infringement because a small portion of adjacent beads have oligonucleotide probes with sequences that differ by less than 25%**
- **"Margin of error" argument is plausible, but likely uphill battle**
- **Potential argument that smaller regions within the Infinium BeadChip could qualify as "the array" in which all oligonucleotide probes differ from their neighbors**

Infinium – Array Patents



Infinium – Array Patents

25% DIFFERENCE

"**Each** capture oligonucleotide probe of the array **differs** in sequence **from every** adjacent capture oligonucleotide probe, when aligned, by at least 25% in nucleotide sequence."

DOCTRINE OF EQUIVALENTS

- **Illumina advances three *legal* restrictions on the doctrine of equivalents:**
  1) **Amendment-based estoppel:** an amendment made for reasons of patentability bars the doctrine of equivalents unless unrelated to the equivalent
  2) **Argument-based estoppel:** a competitor would reasonably believe that the applicant had surrendered the relevant subject matter
  3) **Claim vitiation:** no reasonable jury could determine two elements to be equivalent
- **If we avoid legal restriction, there is a reasonable *factual* argument for equivalence**

Infinium – Array Patents

25% DIFFERENCE

"**Each** capture oligonucleotide probe of the array **differs** in sequence **from every** adjacent capture oligonucleotide probe, when aligned, by at least 25% in nucleotide sequence."

DOCTRINE OF EQUIVALENTS

ESTOPPEL	VITIATION

Amendment-based estoppel: an amendment made for reasons of patentability bars the doctrine of equivalents unless unrelated to the equivalent

The claim limitation was added in response to a restriction requirement, with little explanation. An "unexplained" narrowing of the claims is presumed to be for reasons of patentability and may preclude application of the doctrine of equivalents.

Infinium – Array Patents

25% DIFFERENCE

"**Each** capture oligonucleotide probe of the array **differs** in sequence **from every** adjacent capture oligonucleotide probe, when aligned, by at least 25% in nucleotide sequence."

DOCTRINE OF EQUIVALENTS

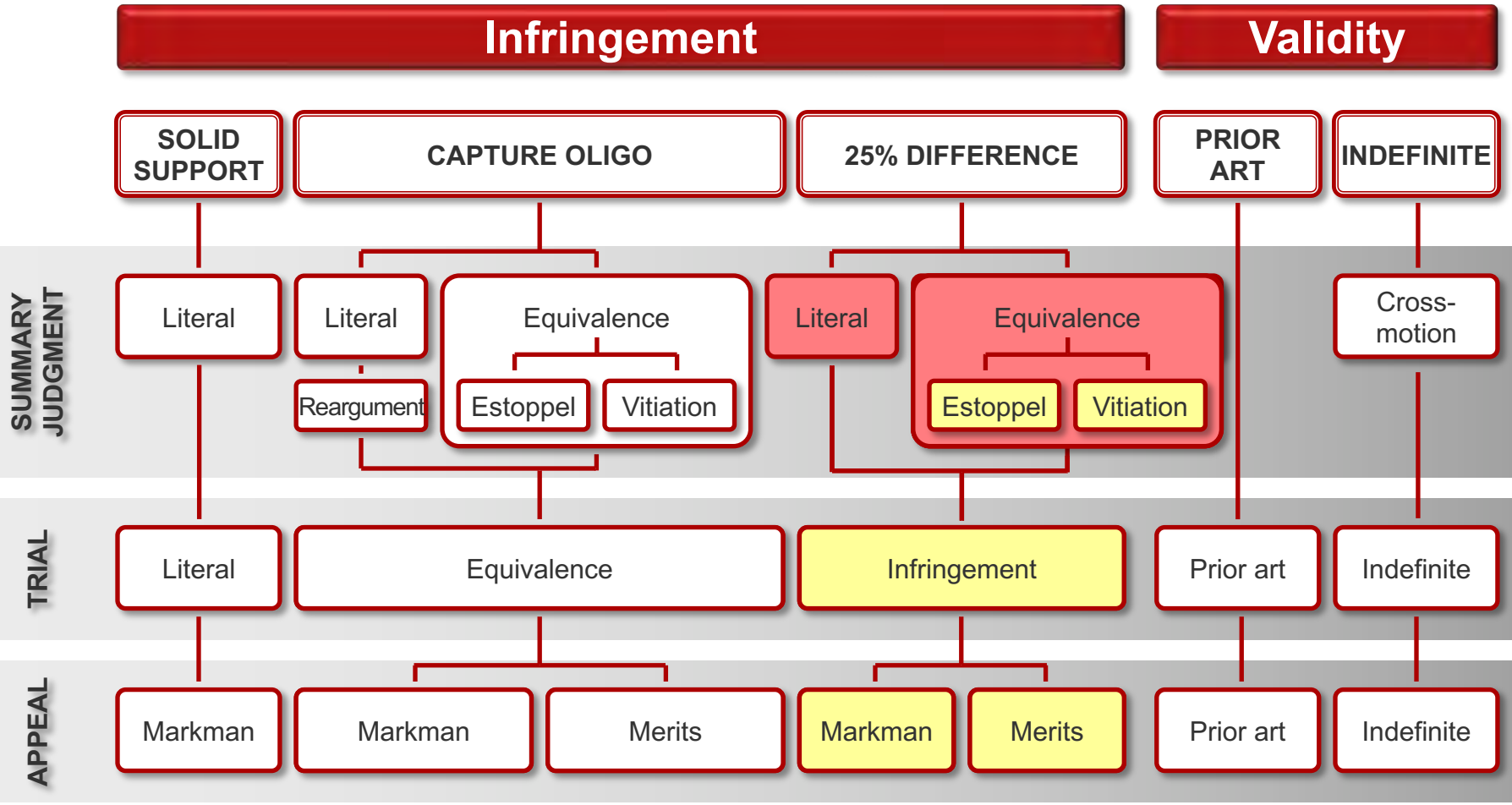
ESTOPPEL

VITIATION

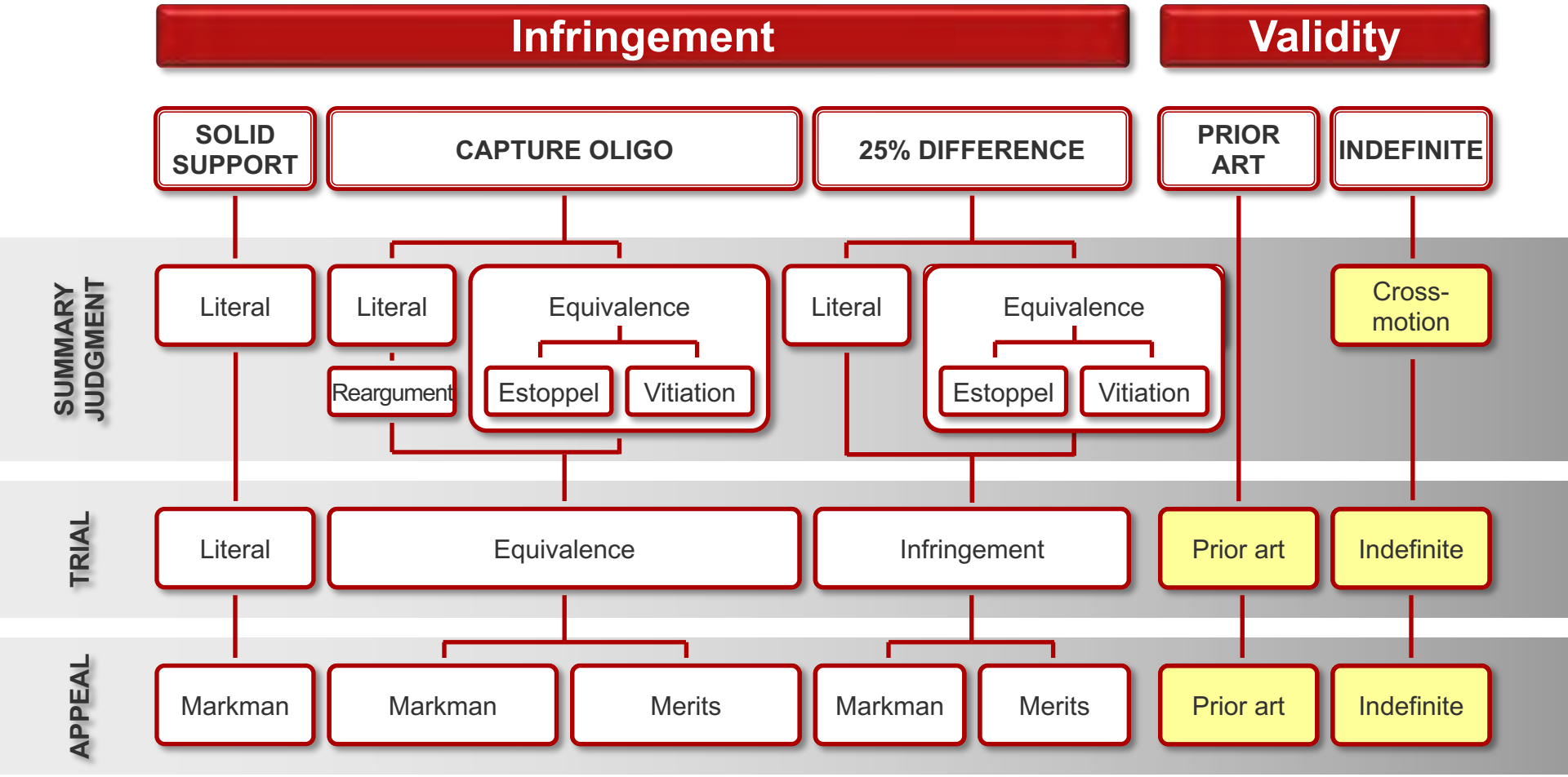
Claim vitiation: no reasonable jury could determine two elements to be equivalent

*Illumina will argue that an array where **each** oligonucleotide probe differs from **every** adjacent capture oligonucleotide probe is the antithesis of an array where only **some** oligonucleotide probes differ from **some** adjacent capture oligonucleotide probes*

Infinium – Array Patents



Infinium – Array Patents

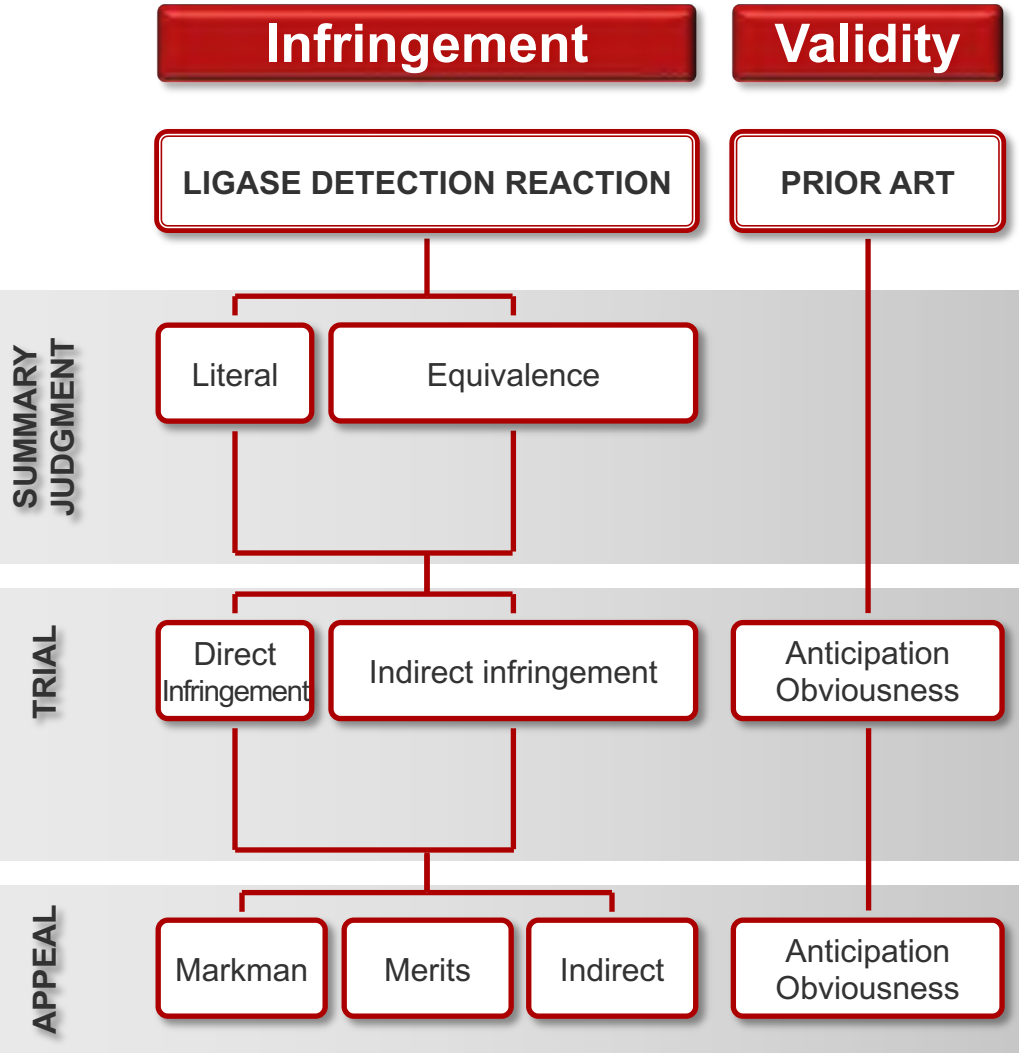


GoldenGate/DASL – LDR-PCR Patents

LDR-PCR Patents

- GoldenGate/DASL

GoldenGate/DASL – LDR-PCR Patents



GoldenGate/DASL – LDR-PCR Patents

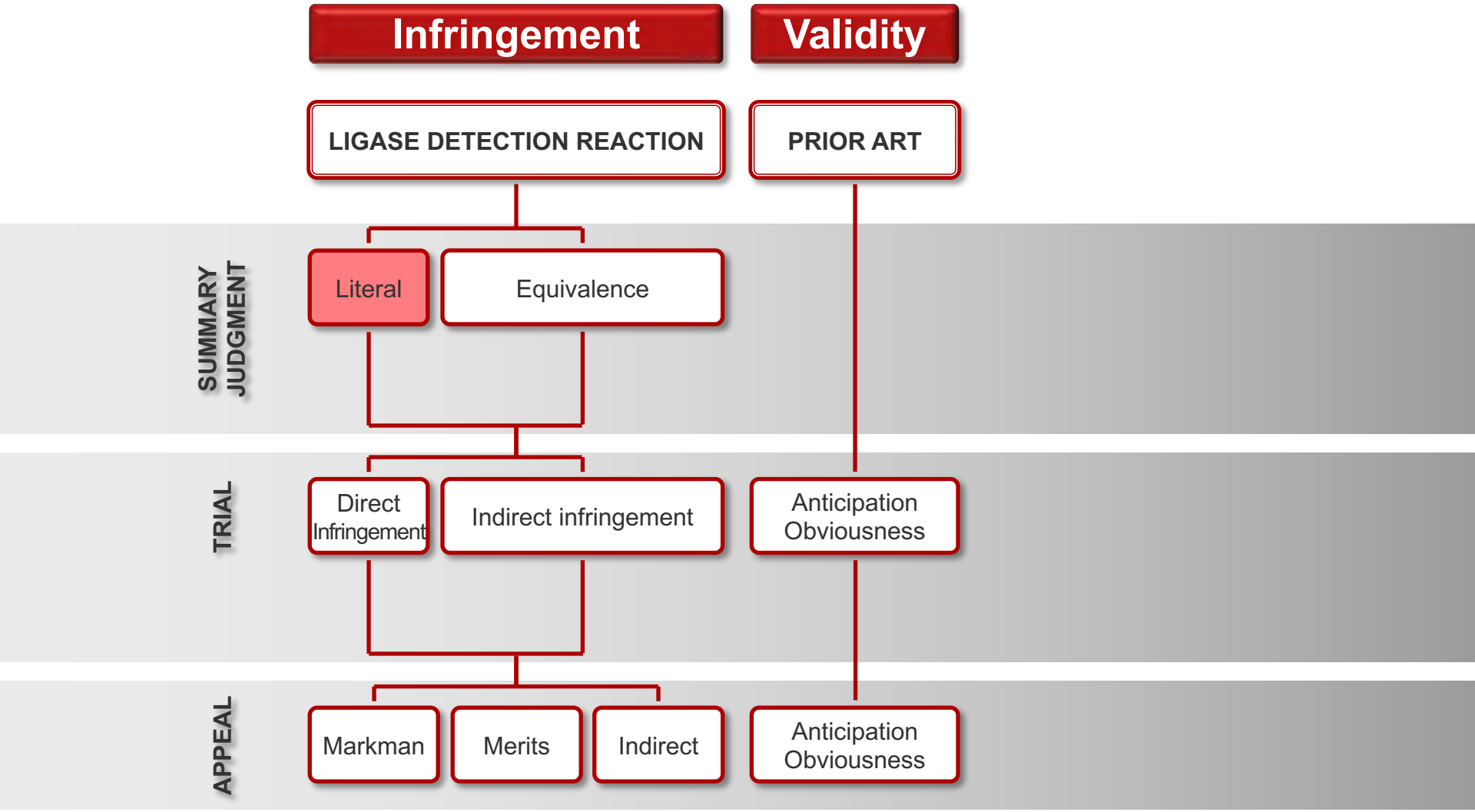
LIGASE DETECTION REACTION

"Reaction which detects the presence of target nucleotide sequence in a sample by using a ligase and one or more sets of two oligonucleotide probes in which the ligase joins the two oligonucleotide probes only if they are **hybridized at adjacent positions** on a target nucleotide sequence."

LITERAL INFRINGEMENT

- **Literal infringement is not asserted because there is a gap between the oligonucleotide probes in GoldenGate/DASL**

GoldenGate/DASL – LDR-PCR Patents



GoldenGate/DASL – LDR-PCR Patents

LIGASE DETECTION REACTION

"Reaction which detects the presence of target nucleotide sequence in a sample by using a ligase and one or more sets of two oligonucleotide probes in which the ligase joins the two oligonucleotide probes only if they are **hybridized at adjacent positions** on a target nucleotide sequence."

DOCTRINE OF EQUIVALENTS

- **Illumina advances two *legal* restrictions on the doctrine of equivalents:**
  1) **Argument-based estoppel:** a competitor would reasonably believe that the applicant had surrendered the relevant subject matter
  2) **Disclosure-dedication:** unclaimed subject matter that is identified in the patent as an alternative to a claim limitation cannot be equivalent
- **If we avoid legal restriction, there is still a difficult *factual* argument for equivalence**

GoldenGate/DASL – LDR-PCR Patents

LIGASE DETECTION REACTION

"Reaction which detects the presence of target nucleotide sequence in a sample by using a ligase and one or more sets of two oligonucleotide probes in which the ligase joins the two oligonucleotide probes only if they are **hybridized at adjacent positions** on a target nucleotide sequence."

DOCTRINE OF EQUIVALENTS

ESTOPPEL

DEDICATION

Argument-based estoppel: a competitor would reasonably believe that the applicant had surrendered the relevant subject matter (must be clear and unmistakable)

"The gap ligase chain reaction process requires an additional step—polymerase extension."

Additional unfavorable statements were made in the prosecution history of the unrelated '521 patent.

GoldenGate/DASL – LDR-PCR Patents

LIGASE DETECTION REACTION

"Reaction which detects the presence of target nucleotide sequence in a sample by using a ligase and one or more sets of two oligonucleotide probes in which the ligase joins the two oligonucleotide probes only if they are **hybridized at adjacent positions** on a target nucleotide sequence."

DOCTRINE OF EQUIVALENTS

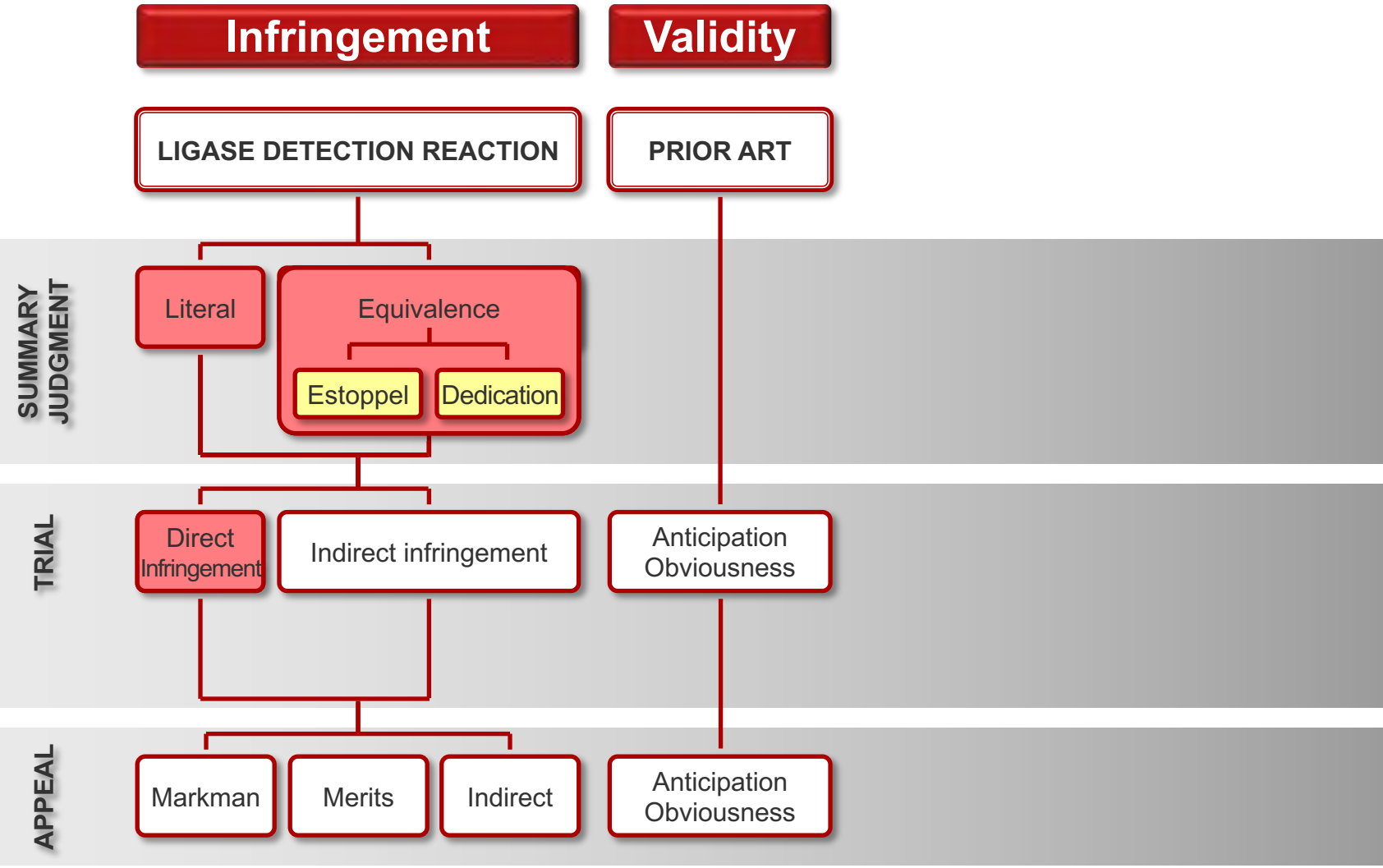
ESTOPPEL

DEDICATION

Disclosure-dedication: unclaimed subject matter that is identified in the patent as an alternative to a claim limitation cannot be equivalent

"Jou et al. … relates to the use of a so called 'gap ligase chain reaction' process to amplify simultaneously selected regions of multiple exons with the amplified products being read on an immunochromatographic strip…"

GoldenGate/DASL – LDR-PCR Patents

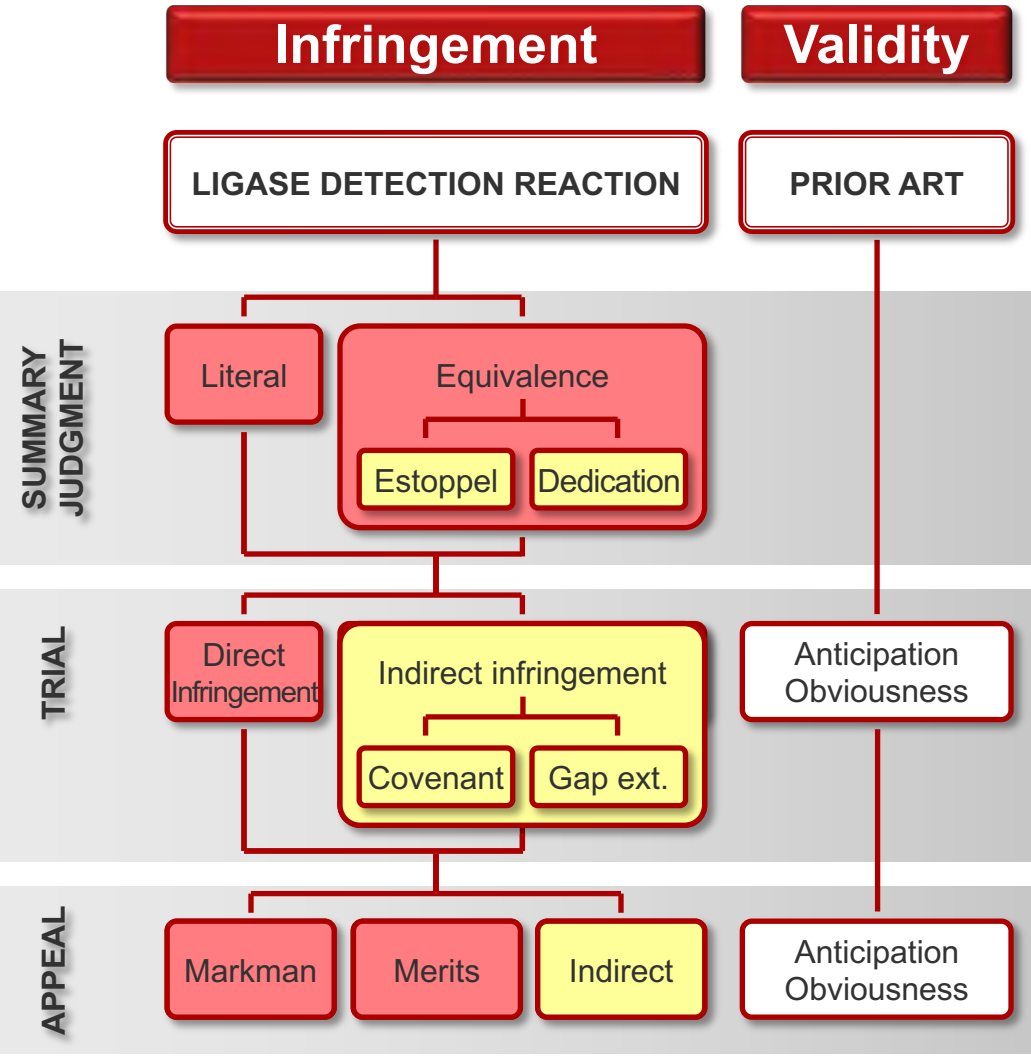


GoldenGate/DASL – LDR-PCR Patents

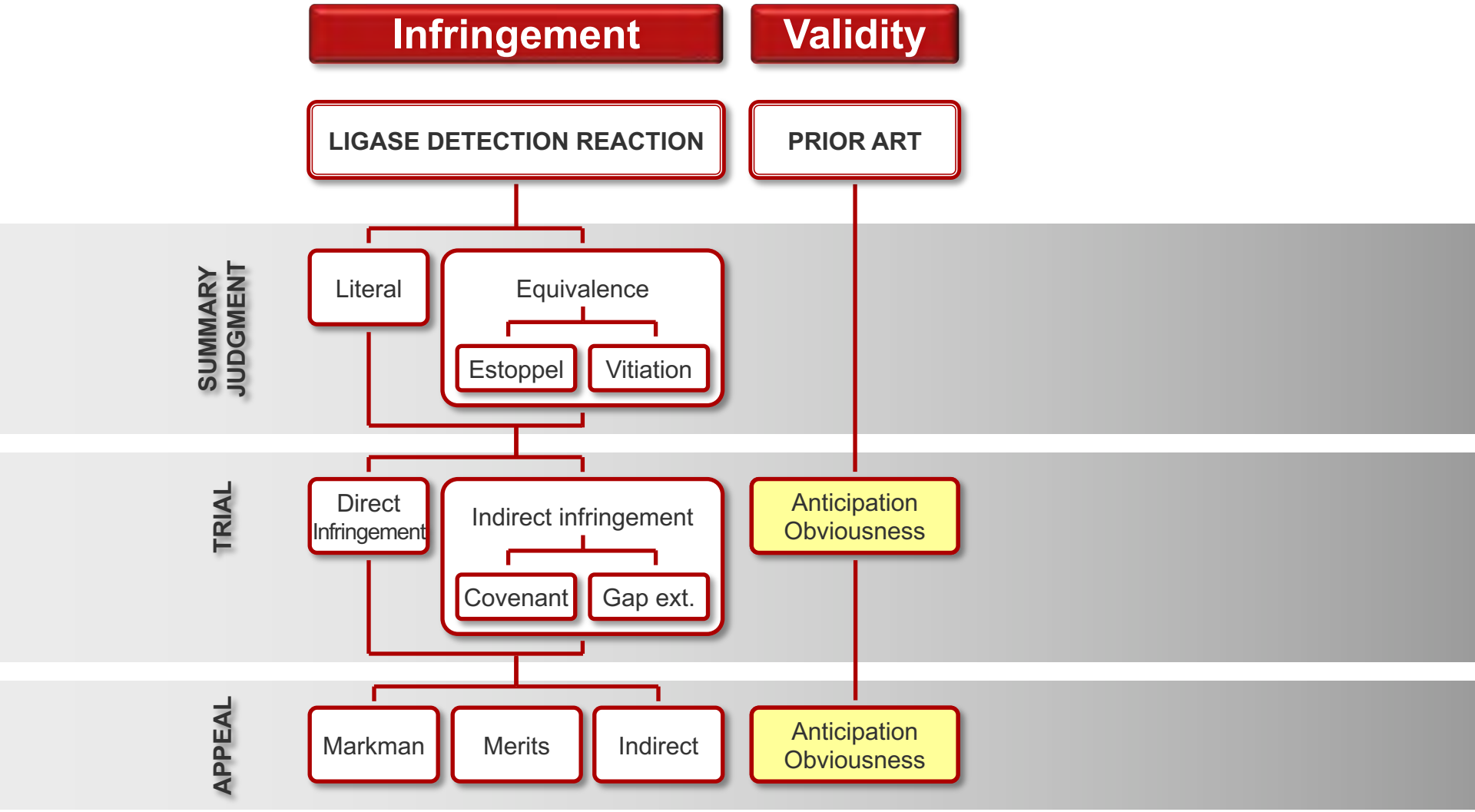
INDIRECT INFRINGEMENT

Indirect infringement: liability for indirect infringement requires that the defendant knew that the acts of infringement by the user constitute patent infringement. A defendant does not indirectly infringe a patent if it has a reasonable belief of noninfringement

GoldenGate/DASL – LDR-PCR Patents

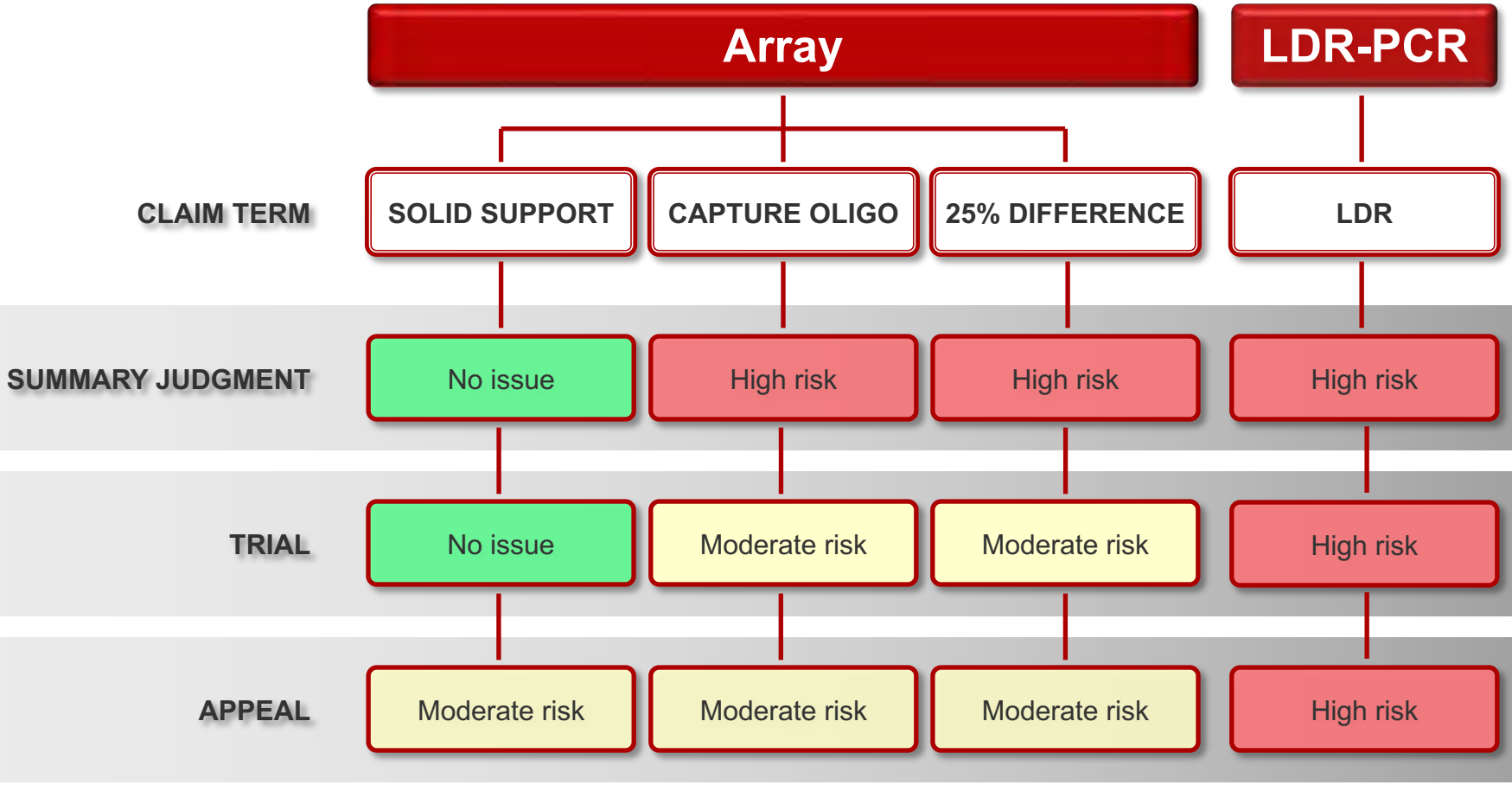


GoldenGate/DASL – LDR-PCR Patents



Procedural Considerations

Procedural Considerations



LATHAM & WATKINS LLP

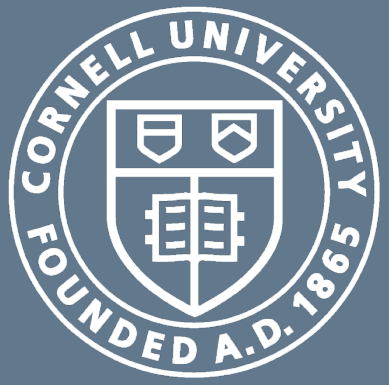


EXHIBIT 7

LATHAM & WATKINS LLP

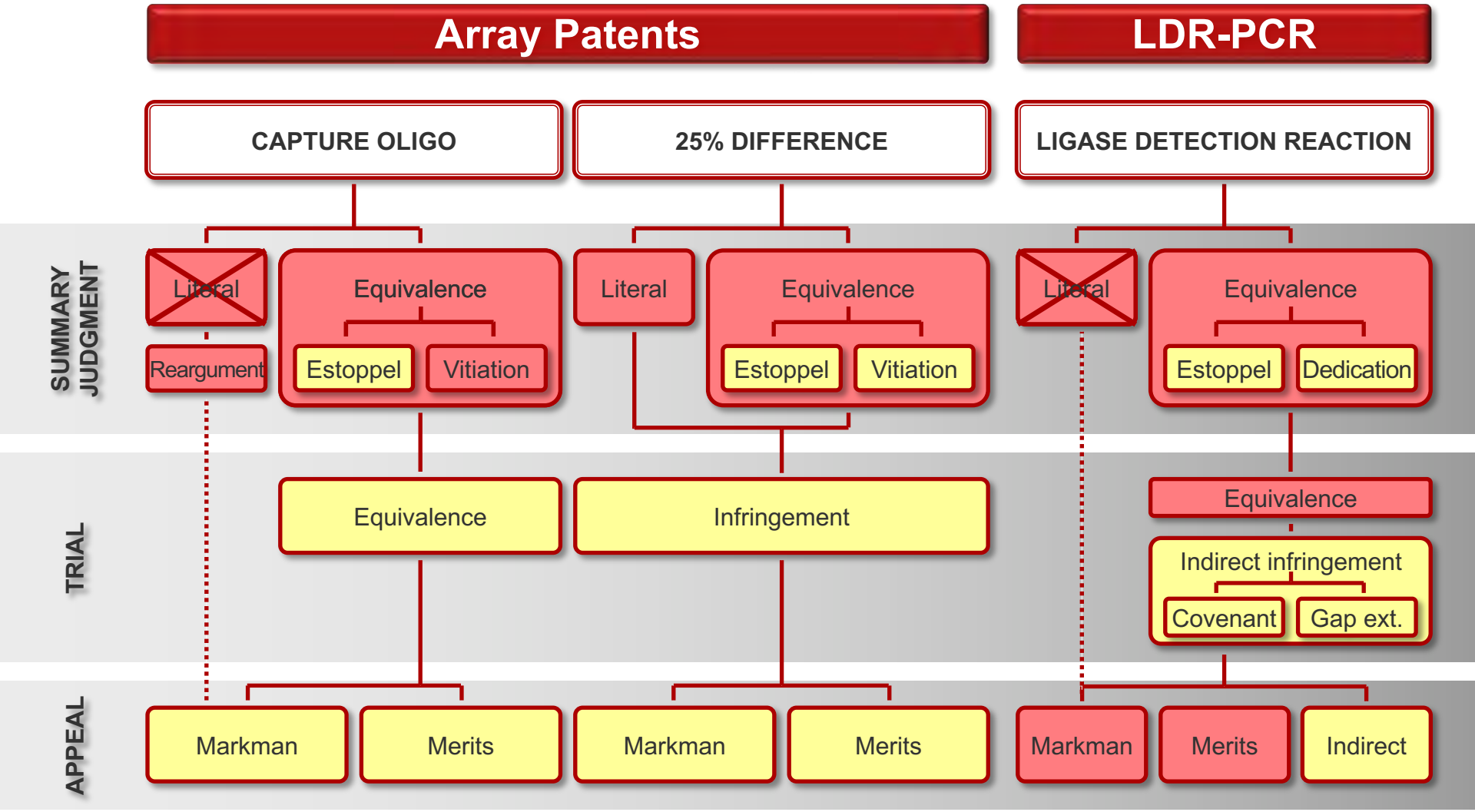
Cornell et al. v. Illumina 10-cv-433 (D. Del.)

Case Assessment

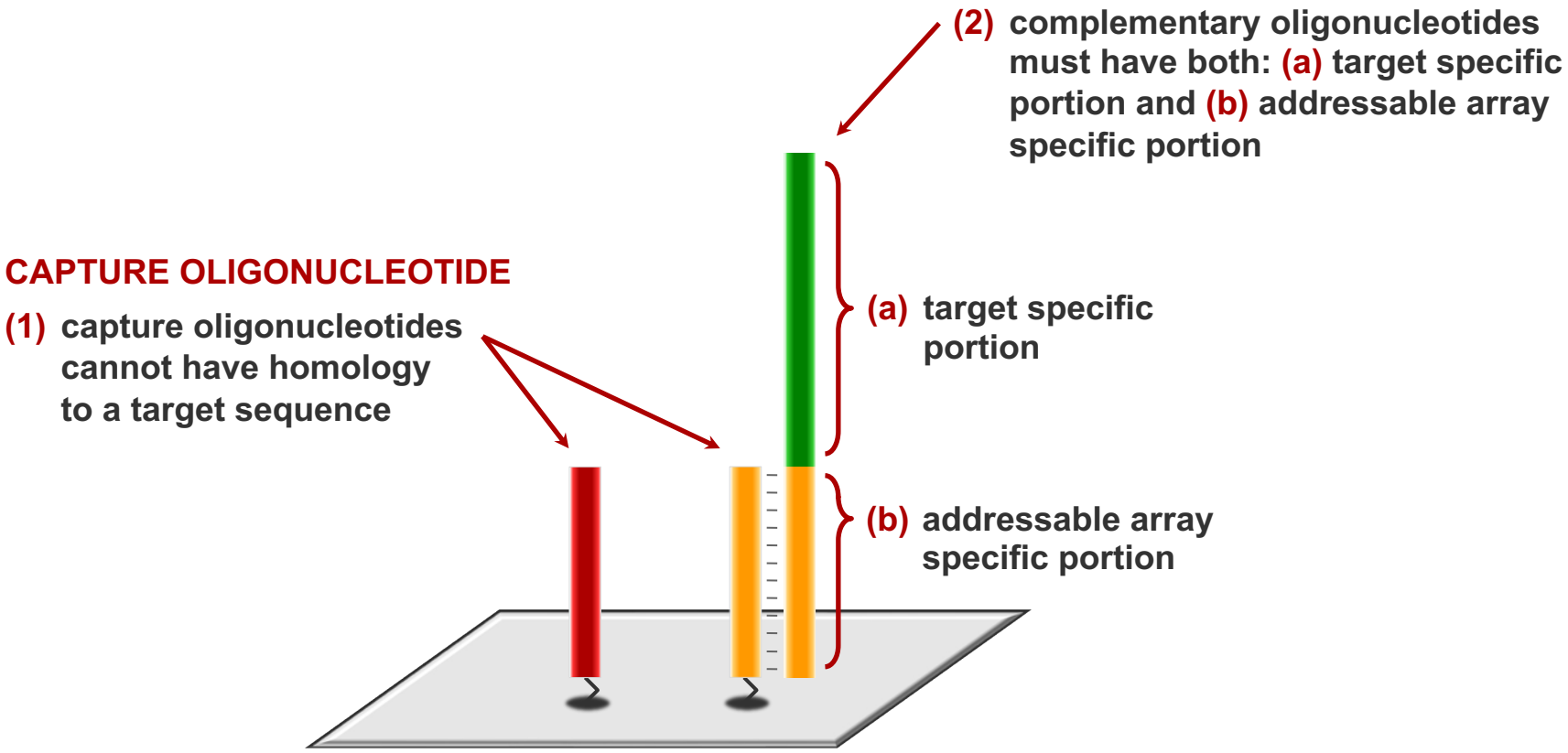
PRIVILEGED & CONFIDENTIAL

February 24, 2017

Infringement Analysis (Post-Markman)



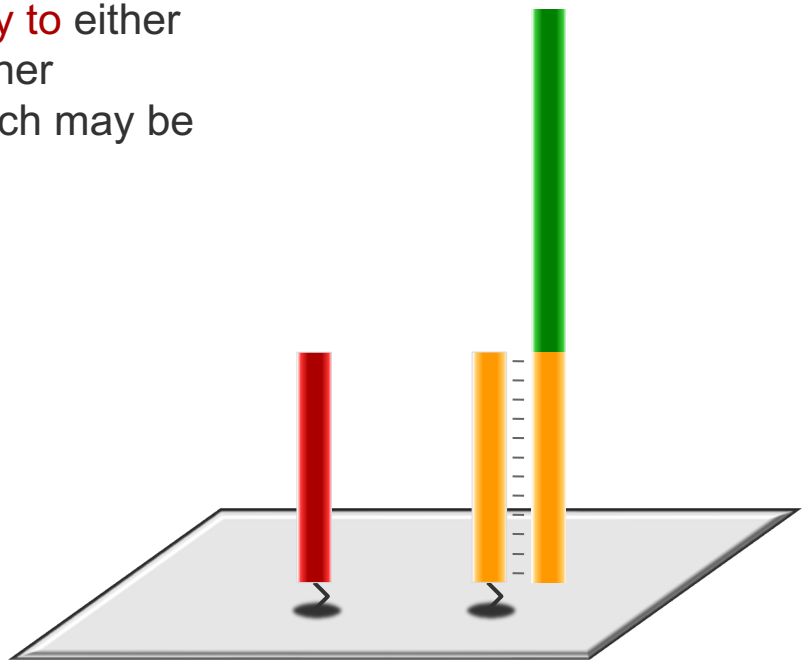
Capture Oligonucleotide (Post-Markman)



Capture Oligonucleotide (Post-Markman)

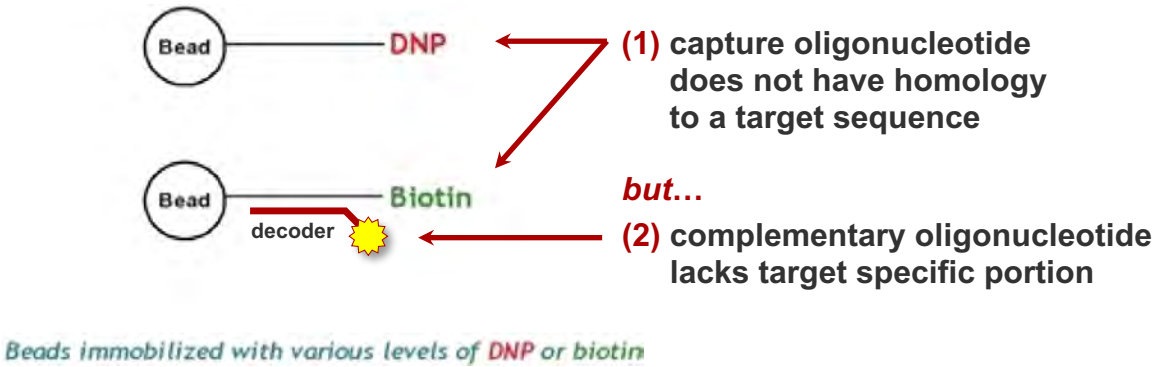
"A capture oligonucleotide sequence does not have any homology to either the target sequence or to other sequences on genomes which may be present in the sample."

('917 patent)



Capture Oligonucleotide (Post-Markman)

Control beads present infringement challenges:

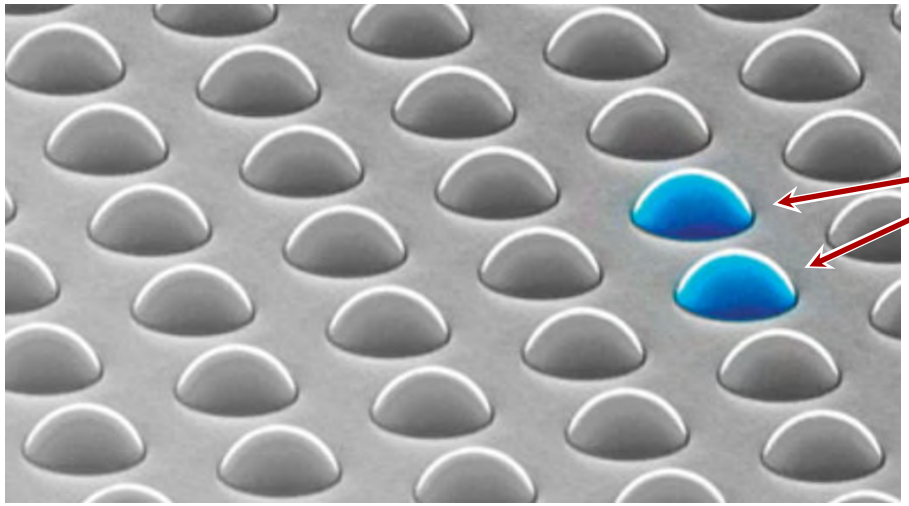


Infinium Staining Controls

Control beads present damages calculation challenges:

To obtain a royalty on the entire market value of the Infinium products, we need to show that the patented feature drives demand for the entire product, and motivates customers to buy the product. Control beads likely do not drive demand for the entire product.

25% Difference (Post-Markman)



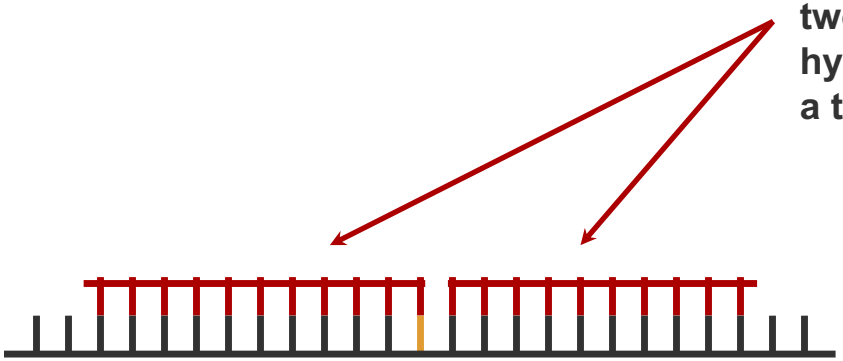
25% DIFFERENCE

each capture oligonucleotide probe of the array must **differ from every** adjacent capture oligonucleotide probe by at least 25%

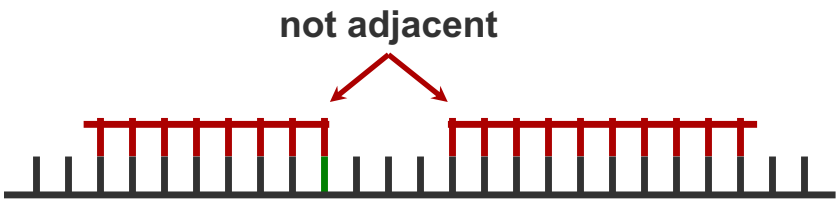
Ligase Detection Reaction (Post-Markman)

LIGASE DETECTION REACTION

two oligonucleotide probes must hybridize at adjacent positions on a target nucleotide sequence



Ligase detection reaction



Gap extension reaction

Ligase Detection Reaction (Post-Markman)

"The claims of the present application clearly refer to an LDR procedure to distinguish single nucleotide differences and require that the oligonucleotide probe be configured to hybridize 'adjacent to one another' on a corresponding target nucleotide sequence. Since they do not involve LCR or filling a gap, these claims are readily distinguishable from Zaun."

('521 patent prosecution history)

Ligase detection reaction

LATHAM & WATKINS LLP

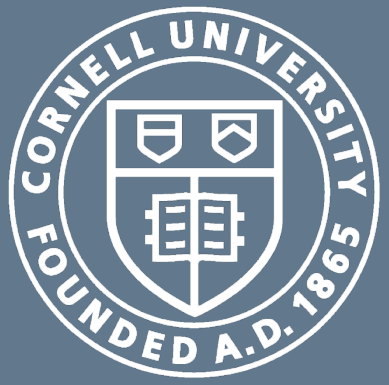


EXHIBIT 8

On Mar 27, 2017, at 11:48 PM, Roger.Chin@lw.com wrote:

PRIVILEGED & CONFIDENTIAL

Francis and Monib,

We had a chance to study the materials that you prepared and organized for us. It is certainly helpful information, and I know that Matt's team developed and explored some similar avenues based on their access to confidential Illumina documentation. However, there are certain limits to our ability to introduce new issues at this juncture in the litigation:

· The technical analysis does not change our views on claim construction and infringement that we discussed on our previous calls. In this case, the doctrine of equivalents case is limited by legal doctrines (estoppel, dedication, vitiation), which are largely separate from the technical points raised.
· The issue of copying plays a limited role in our patent case. While copying can be used as evidence of willfulness, based on our review of the record and discussions with Matt's team, we do not believe there is a viable argument that the commercialized Illumina products actually incorporate the specific Barany zip code sequences. The more general benefits Illumina may have received from copying Barany technology do not change the challenges we face on the infringement issues. In some ways, the patent laws actually encourage building upon others' ideas so long as the later products do not infringe.
· Our current patent case is not the proper vehicle to advance general allegations of IP theft. Those claims are more in the nature of trade secrets misappropriation or breach of contract, which are not part of our case and would have to be separately evaluated.
· We are dealing with a fixed record and new arguments are not easily added at this point. Fact discovery is closed. Expert reports have been submitted and technical experts were already deposed. For example, while our experts discuss the copied 16 sequences, broader arguments about IP theft are not disclosed in their expert reports (for some of the reasons outlined above) and therefore likely cannot be presented at trial.

Roger J. Chin
LATHAM & WATKINS LLP
505 Montgomery Street | Suite 2000 | San Francisco, CA 94111-6538
T: +1.[redacted] | M: +

From: Francis Barany [mailto:barany@med.cornell.edu]
Sent: Thursday, March 23, 2017 11:58 AM
To: Chin, Roger (SF); Matthew Pearson; Elderkin, Dianne; Lumish, Doug (SV); Sobolski, Gregory (SF); Rip.Finst@thermofisher.com; sean.boyle@thermofisher.com; Valerie Cross; Brian Kelly
Subject: Privileged and Confidential: Cornell - Illumina Case; potential obstruction of justice and/or witness tampering; legal team only 6
Importance: High

Dear Roger, Matthew, Dianne, Doug, Greg, Rip, Sean, Valerie, and Brian,

We wish to bring to your attention a potential obstruction of justice and/or witness tampering with the Cornell vs. Illumina case. While not having a key witness available during a trial does occur from time to time, the "disappearance" of two key witnesses is problematic in our view. There is a huge difference between not knowing where a former employee is after a decade, and possibly orchestrating someone's disappearance when subpoenaed.

It appears that Illumina was fully aware of JB Fan's whereabouts since 2015. On multiple occasions, Illumina employees co-authored patents and papers and attended meetings with JB Fan. This occurred both prior to the court ordered Subpoena in July 2015, as well as afterwards. In fact, an Illumina employee registered the website "AnchorDx.com" in April 2015, and is listed as the Chief Technical Officer of AnchorDx, Inc., a company JB Fan started in August 2015, in an attempt to make it appear that he had "fled" to China. He continued to maintain a residence in San Diego and verifiably was present in San Francisco on June 23, 2016. He even listed himself as not only an employee of Illumina in June, 2016, but also a member of Illumina's "Leadership" in an author declaration form for an Abstract he was part of at the 2016 ASCO meeting.

Having analyzed his patent applications and declarations over the years, multiple verifiably misleading or false statements were made in the

prosecution history of many patents as revealed in the Ariosa/Illumina IPR proceedings in 2014. It is felt, that he was, and is, a critical and key witness and co-conspirator in the patent infringement perpetrated by Illumina against Cornell. This combined with the fact that Illumina has likely concealed that they call their "address" sequences on Infinium arrays "Zipcode" in their DLL files for the DMAP Decode File Client software, is powerful evidence that Illumina has deliberately mislead the Court.

We would like to consider options for bringing this additional information to the attention of Judge Stark, with copies to the Illumina attorneys. We believe it is in Cornell's best interest for both settlement talks and/or trial to send this information to Judge Stark and Illumina. We believe that this is the only way to expose that Illumina is infringing on Cornell's IP of the Universal Zipcode Sentrix and Infinium Arrays, and its uses for SNP detection, gene expression studies, methylation studies, NGS sequencing, and other applications. To add insult to injury, ABI shared confidential information with Illumina as part of the JDA, and Illumina took that information, used it to develop their own assays and arrays, and then they tried to re-patent it and call it something else, and then they tried to hide the fact that these were the same arrays invented by Cornell in the first place. The plagiarism was to the degree where they used our exact 16 zipcodes as their first "IllumaCodes". This information will not only strengthen our case for trial, but it will also expose fraudulent claims of Illumina and its current and former employees, leading to a potentially more equitable settlement agreement or jury verdict.

We did not get any response from you about your availability for a webex this week for Tuesday, Wednesday, or Thursday @ 5:00 PM EST. Would you be available to speak about the PowerPoints we sent on Sunday, as well as the contents of this email and follow-up on Friday @ 11:00 AM EST, or Monday or Tuesday @ 5:00 PM EST? Would you kindly let us know if either of these times work for you? If they do not work, would you please suggest alternative times? One of us (FB) will be traveling March 29-31 & April 4-7.

Best,

Francis and Monib

Tel [redacted] or [redacted] or [redacted]
Fax [redacted]
barany@med.cornell.edu

Mobile [redacted]