



(43) International Publication Date
23 October 2014 (23.10.2014)

(10) International Publication Number
WO 2014/172288 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2014/034037

(22) International Filing Date:
14 April 2014 (14.04.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/866,111 19 April 2013 (19.04.2013) US

(71) Applicant: **RAINDANCE TECHNOLOGIES, INC.**
[US/US]; 749 Middlesex Turnpike, Billerica, MA 01821
(US).

(72) Inventors: **LINK, Darren, Roy**; 8 Buckman Drive, Lexington, MA 02421 (US). **MILLER, Benjamin**; 44 Crane Road, Littleton, MA 01460 (US).

(74) Agents: **MEYERS, Thomas, C.** et al.; Brown Rudnick Llp, One Financial Center, Boston, MA 02111 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: DIGITAL ANALYTE ANALYSIS

(57) Abstract: The invention generally relates to detecting target molecules.



WO 2014/172288 A2

DIGITAL ANALYTE ANALYSIS

Related Application

This application claims the benefit of and priority to U.S. non-provisional application serial number 13/866,111, filed April 19, 2013, the content of which is incorporated by reference herein in its entirety.

Field of the Invention

The invention generally relates to droplet based digital PCR and methods for analyzing a target nucleic acid using the same.

Background

Assays have been developed that rely on analyzing nucleic acid molecules from bodily fluids for the presence of mutations, thus leading to early diagnosis of certain diseases such as cancer. In a typical bodily fluid sample however, any abnormal nucleic acids containing mutations of interest are often present in small amounts (e.g., less than 1%) relative to a total amount of nucleic acid in the bodily fluid sample. This can result in a failure to detect the small amount of abnormal nucleic acid due to stochastic sampling bias.

The advent of PCR and real-time PCR methodologies has greatly improved the analysis of nucleic acids from both throughput and quantitative perspectives. While traditional PCR techniques typically rely on end-point, and sometimes semi-quantitative, analysis of amplified DNA targets via agarose gel electrophoresis, real-time PCR (or qPCR) methods are geared toward accurately quantifying exponential amplification as the reaction progresses. qPCR reactions are monitored either using a variety of highly sequence specific fluorescent probe technologies, or by using non-specific DNA intercalating fluorogenic dyes.

As the need for higher throughput in analyzing multiple targets in parallel continues to escalate in the fields of genomics and genetics, and as the need for more efficient use of sample grows in medically related fields such as diagnostics, the ability to perform and quantify multiple amplifications simultaneously within the same reaction volume (multiplexing) is paramount for both PCR and qPCR. While end-point PCR can support a high level of amplicon multiplexing, such ample capacity for multiplexing probe-based qPCR reactions remains elusive for a number

of reasons. For example, most commercial real-time thermal cyclers only support up to four differently colored fluorophores for detection as a consequence of the limited spectral resolution of common fluorophores, translating into a multiplexing capacity of 4x. Additionally, while optimization of single target primer/probe reactions is now standard practice, combining primers and probes for multiple reactions changes the thermodynamic efficiencies and/or chemical kinetics, necessitating potentially extensive troubleshooting and optimization. Very high multiplexing of greater than 100x has been demonstrated in a “one of many” detection format for pathogen identification using “sloppy” molecular beacons and melting points as fingerprints, however the approach is restricted to applications with a slim likelihood of the presence of multiple simultaneous targets. A half-multiplexing method achieved 19x in a two-step reaction with general multiplexed preamplification in the first step, followed by separate single-plex quantitative PCR in the second step. However a general purpose single-pot solution to qPCR multiplexing does not yet exist.

Digital PCR (dPCR) is an alternative quantitation method in which dilute samples are divided into many separate reactions. See for example, Brown et al. (U.S. patent numbers 6,143,496 and 6,391,559) and Vogelstein et al. (U.S. patent numbers 6,440,706, 6,753,147, and 7,824,889), the content of each of which is incorporated by reference herein in its entirety. The distribution from background of target DNA molecules among the reactions follows Poisson statistics, and at so called “terminal dilution” the vast majority of reactions contain either one or zero target DNA molecules for practical intents and purposes. In another case, at so called “limiting dilution” some reactions contain zero DNA molecules, some reactions contain one molecule, and frequently some other reactions contain multiple molecules, following the Poisson distribution. It is understood that terminal dilution and limiting dilution are useful concepts for describing DNA loading in reaction vessels, but they have no formal mathematical definition, nor are they necessarily mutually exclusive. Ideally, at terminal dilution, the number of PCR positive reactions (PCR(+)) equals the number of template molecules originally present. At limiting dilution, Poisson statistics are used to uncover the underlying amount of DNA. The principle advantage of digital compared to qPCR is that it avoids any need to interpret the time dependence of fluorescence intensity—an analog signal—along with the main underlying uncertainty of non-exponential amplification during early cycles.

Summary

The invention generally relates to the manipulation of nucleic acid in droplets, and in particular, nucleic acid amplification and detection. In one aspect, the invention provides a droplet that contains a single nucleic acid template and a plurality of primer pairs specific for multiple target sites on the template. The single nucleic acid template can be DNA (e.g., genomic DNA, cDNA, etc.) or RNA. The template is amplified in the droplet for detection; and may preferably be amplified using a plurality of primer pairs as described herein.

The ability to amplify and detect single nucleic acids in droplets enables digital PCR, detection, counting, and differentiation among nucleic acids, especially those present in heterogeneous samples. Thus, the invention applies to digital amplification techniques and, in specific embodiments enables multiplex PCR in droplets. For example, multiplexing primers in droplets enables the simultaneous increase in the number of PCR droplets while keeping the amount of input DNA the same or lower and generate the same or greater amplicon yield. This results in an overall increase in the amount of PCR positive droplets and amplicon yield without the consumption of more DNA. Even though the number of PCR primer pairs per droplet is greater than one, there is only one template molecule per droplet, and thus, in some implementations, there is only one primer pair per droplet that is being utilized at one time. As such, the advantages of droplet PCR for eliminating bias from either allele specific PCR or competition between different amplicons is maintained. However, as described below in relation to detection of haplotypes, other implementations advantageously allow detection of multiple loci on a single template using multiple primer pairs, preferably designed to minimize bias.

In further aspects of the invention, a plurality of primer pairs encompassed within a droplet each include a bridge section flanked by a first targeting arm and a second targeting arm. Each primer pair is specific for a different target site on a single template nucleic acid also disposed within the droplet. The first targeting arm hybridizes to a first region of the target site and the second targeting arm hybridizes to a second targeting site. The bridge section is not complementary to the single template nucleic acid within the droplet. The droplet also includes a plurality of probes in which each member of the plurality of probes corresponds to a primer and is designed to hybridize to a complement of the bridge region of the corresponding primer. Preferably, the bridge section of each primer of a primer pair is the same, and thus corresponds to the same probe for detection. According to certain embodiments, each member of the

plurality of probes designed to hybridize to a bridge region includes a detectable label. The plurality of probes may include one or more groups of probes at varying concentrations. Members of the one or more groups of probes may each have the same detectable label. Alternatively, each member of the plurality of probes includes a unique detectable label. Typically, the detectable label is a fluorescent label.

Microfluidic droplets for multiplex analysis according to the invention contain a plurality of probes that hybridize to amplicons produced in the droplets. Preferably, the droplet contains two or more probes, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 500, or more probes. Certain members of the plurality of probes include a detectable label. Members of the plurality of probes can each include the same detectable label, or a different detectable label. The detectable label is preferably a fluorescent label. The plurality of probes can include one or more groups of probes at varying concentrations. The one or more groups of probes can include the same detectable label which will vary in intensity upon detection, due to the varying probe concentrations. The droplets of the invention can further contain one or more reagents for conducting a polymerase chain reaction, such as a DNA or RNA polymerase, and/or dNTPs.

The present invention additionally relates to a method for detecting a plurality of targets in a biological sample using digital PCR in microfluidic droplets. The sample may be a human tissue or body fluid. Exemplary body fluids pus, sputum, semen, urine, blood, saliva, and cerebrospinal fluid.

One or more droplets are formed, each containing a single nucleic acid template and a heterogeneous mixture of primer pairs and probes, each specific for multiple target sites on the template. For example, a first fluid (either continuous, or discontinuous as in droplets) containing a single nucleic acid template (DNA or RNA) is merged with a second fluid (also either continuous, or discontinuous as in droplets) containing a plurality of primer pairs and a plurality of probes, each specific for multiple targets sites on the nucleic acid template to form a droplet containing the single nucleic acid template and a heterogeneous mixture of primer pairs and probes. The second fluid can also contain reagents for conducting a PCR reaction, such as a polymerase and dNTPs.

Certain members of the plurality of probes include a detectable label. Members of the plurality of probes can each include the same detectable label, or a different detectable label.

The detectable label is preferably a fluorescent label. The plurality of probes can include one or more groups of probes at varying concentrations. The one or more groups of probes can include the same detectable label which varies in intensity upon detection, due to the varying probe concentrations.

The first and second fluids can each be in droplet form. Any technique known in the art for forming droplets may be used with methods of the invention. An exemplary method involves flowing a stream of the sample fluid containing the nucleic acid template such that it intersects two opposing streams of flowing carrier fluid. The carrier fluid is immiscible with the sample fluid. Intersection of the sample fluid with the two opposing streams of flowing carrier fluid results in partitioning of the sample fluid into individual sample droplets containing the first fluid. The carrier fluid may be any fluid that is immiscible with the sample fluid. An exemplary carrier fluid is oil. In certain embodiments, the carrier fluid includes a surfactant, such as a fluorosurfactant. The same method may be applied to create individual droplets from the second fluid containing the primer pairs (and, in some implementations, the amplification reagents). Either the droplets containing the first fluid, the droplets containing the second fluid, or both, may be formed and then stored in a library for later merging, aspects of certain implementations of which are described in U.S. Patent Application Serial No. 12/504,764, hereby incorporated herein in its entirety for all purposes. Once formed, droplets containing the first and second fluids can be merged to form single droplets containing the single nucleic acid template and heterogeneous mixture of primer pairs and probes. Merging can be accomplished, for example, in the presence of an electric field. Moreover, it is not required that both fluids be in the form of droplets when merging takes place. One exemplary method for merging of fluid portions with droplets is taught, for example, in co-pending U.S. Patent Application No. 61/441,985, filed on February 11, 2011.

The nucleic acid template in each of the merged/formed droplets is amplified, e.g., by thermocycling the droplets under temperatures/conditions sufficient to conduct a PCR reaction. The resulting amplicons in the droplets can then be analyzed. For example, the presence of absence of the plurality of targets in the one or more droplets is detected optically, e.g., by the detectable label on the plurality of probes.

The invention further relates to methods for analyzing a target nucleic acid. More particularly, methods of the invention are able to detect polymerase errors that occur during a

PCR reaction and are able to exclude from analysis amplification products that are a result of a polymerase error. Methods of the invention are particularly useful in digital PCR where a polymerase error may result in a partitioned section of sample being incorrectly identified as containing a mutant allele, i.e., a false positive. Such false positives greatly impact the validity and precision of digital PCR results. Methods of the invention are able to uniquely detect multiple targets with the same optical color. Methods of the invention are particularly useful in digital PCR where it is desirable to identify multiple different target molecules that may be present in the starting test fluid.

Methods of the invention involve forming sample droplets containing target nucleic acid. Ideally, methods of the invention comprise forming droplets for digital PCR. Preferred digital PCR droplets contain one copy of a nucleic acid to be amplified, although they may contain multiple copies of the same nucleic acid sequence. Any technique known in the art for forming sample droplets may be used with methods of the invention. One exemplary method involves flowing a stream of sample fluid including nucleic acids such that it intersects two opposing streams of flowing carrier fluid. The carrier fluid is immiscible with the sample fluid. Intersection of the sample fluid with the two opposing streams of flowing carrier fluid results in partitioning of the sample fluid into individual sample droplets. The carrier fluid may be any fluid that is immiscible with the sample fluid. An exemplary carrier fluid is oil. In certain embodiments, the carrier fluid includes a surfactant, such as a fluorosurfactant.

The targets are then amplified in the droplets. Any method known in the art may be used to amplify the target nucleic acids either linearly or exponentially. A preferred method is the polymerase chain reaction (PCR). For purposes of the invention, any amplification technique commonly known in the art may be implemented such as rolling circle amplification, isothermal amplification, or any combination of amplification methods using loci specific primers, nested-primers, or random primers (such primers, and/or primers used for PCR, are included in the term “amplification reagents”). Once amplified, droplets containing amplicon from the target and amplicon from a variant of the target are excluded. One method to exclude droplets that contain a heterogeneous population of amplicons from droplets that contain a homogeneous population of amplicons includes hybridizing detectably-labeled probes to the amplicons, flowing the droplets through a microfluidic channel, and excluding those droplets in which both amplicon from the target and amplicon from a variant of the target are detected.

Once droplets containing a heterogeneous population of amplicons are excluded, droplets that contain a homogeneous population of amplicons are analyzed. Any analytical technique known in the art may be used. In certain embodiments, analyzing the droplets involves determining a number of droplets that contain only wild-type target, and determining a number of droplets that contain only a variant of the target. Generally, the presence of droplets containing only the variant is indicative of a disease, such as cancer. The variant may be an allelic variant. An exemplary allelic variant is a single nucleotide polymorphism. The variant may also be a specific haplotype. Haplotypes refer to the presence of two or more variants on the same nucleic acid strand. Haplotypes can be more informative or predictive than genotypes when used to determine such things as the presence or severity of disease, response to drug therapy or drug resistance of bacterial or viral infections. Because each droplet contains only one template strand it is an ideal vessel for the determination of haplotypes. The detection of two or more variants in a single droplet that contains a single intact nucleic acid strand identifies the haplotype of the variants on that strand. The presence of two or more markers in the same droplet can be identified by such methods as the presence of dyes of multiple colors or the increase in the intensity of a single dye or a combination of both. Any method that allows the identification of multiple variants in a single droplet enables the determination of a samples haplotype.

In accordance with some implementations of the invention, a method is provided for analyzing a target nucleic acid that includes compartmentalizing a first fluid into portions, each portion containing a single target nucleic acid; amplifying the target in the portions; excluding portions containing amplicon from the target and amplicon from a variant of the target; and analyzing target amplicons.

In other aspects, the invention generally provides methods for detecting a recurrence of a cancer in a patient. Those methods may involve forming sample droplets containing a single target nucleic acid derived from a patient sample, flowing the sample droplets through a channel, amplifying the target in the droplets, detecting amplified target in the droplets, excluding droplets including a heterogeneous population of amplicons, and analyzing non-excluded droplets to determine the presence of mutant alleles indicative of recurrence. In certain embodiments, the analyzing step includes capturing amplicon obtained from the droplets using labeled capture probes. The sample may be a human tissue or body fluid. Exemplary body

fluids are pus, sputum, semen, urine, blood, saliva, stool, and cerebrospinal fluid. In other aspects of the invention generally provide a method for forensic identification of low levels of target nucleic acid in an environment having multiple other sources of nucleic acid. Such methods may also be practiced using fluids compartmentalized in containers other than or in addition to droplets.

As described, aspects of the invention require determining whether a droplet contains an amplifiable product. One technique for determining whether a droplet contains amplifiable product includes flowing droplets in a mono-dispersed fashion past a detector (e.g. one droplet passing the detector at a time), and detecting the signals of each droplet separately. Droplets having signals above background are droplets that include amplifiable product. A problem with this mono-dispersed detection technique is that it requires a large data file because the signal from every droplet is obtained and recorded. In addition, the mono-disperse flow of droplets past the detector takes time and is inefficient. In order to solve the problems associated with mono-dispersed detection, methods of the invention provide a non-mono-dispersed technique for determining whether a droplet contains amplifiable product. In this alternative technique, a plurality of droplets is simultaneously flowed past the detector, in which the plurality of droplets is not mono-dispersed within a portion of the channel. The detector is configured to only obtain and record signals above background noise. As a result, only droplets having amplifiable product are counted and detected as they pass the detector.

According to one embodiment, the invention provides a method for counting droplets. The method includes providing a plurality of droplets, in which a subset of droplets includes nucleic acid from a sample. The nucleic acid is then amplified within the droplets. A signal above background is detected for only droplets containing amplified nucleic acid. From the detected droplets having signals above background, a ratio of normal to abnormal nucleic acid is determined. This ratio is indicative of a condition. In certain embodiments, at least one of the plurality of droplets includes a first detectable target and at least one other droplet includes a second detectable target. The first detectable target includes a normal nucleic acid and the second detectable target includes an abnormal nucleic acid. In certain embodiments, the ratio is determined by counting a number X of droplets containing the first detectable target and a number Y of droplets containing a second detectable target.

In further aspects of the invention, methods of determining the nucleic acid make-up of a sample are provided. More specifically, these methods can be used to detect contiguous, intact sequences of RNA or DNA within a sample as well as fragments of those sequences. For certain experiments, samples comprising a higher proportion of contiguous, intact sequences are better candidates for additional testing than samples comprising mostly fragmented sequences. These methods may include partitioning a nucleic acid sample of different lengths into a plurality of different portions, where each portion includes, on average, a single nucleic acid molecule. A first and second primer pairs and a first and second detectably labeled probes are also introduced into the partitioned portions, where the first and second primer pairs are specific for first and second locations on the nucleic acid. If the nucleic acid contains the first and second locations spaced apart from each other, when the first probe hybridizes to the first location and the second probe hybridizes to the second location, an amplicon is created spanning a length. Thus, by amplifying the nucleic acid in the partitioned portions and detecting amplicons in the partitioned portions, the presence of signal from both probes indicates the presence of a nucleic acid that is contiguous between the first and second locations. By making this measurement for the plurality of portions, it is possible to determine a nucleic acid make-up of the sample based on the detecting step. In certain embodiments, the determining step may include comparing relative amounts of contiguous nucleic acid to relative amounts of non-contiguous nucleic acid. In other embodiments, the determining step may include comparing relative amounts of contiguous nucleic acid or non-contiguous nucleic acid to a total amount of nucleic acid.

Methods in accordance with the invention also encompass the use of a single primer pair. The method includes providing a fluid comprising the sample nucleic acid and a plurality of one or more primer pairs, wherein each primer pair has at least one unique related probe and is selected to be complementary to one or more sequences of known length. The method also includes partitioning the fluid into a plurality of partitions, wherein at least a first portion of the partitions comprise one molecule of the nucleic acid sample having sequences complementary to one or more of the primer pairs, and at least one related probe, and a second portion of the partitions comprise no molecules of the sample nucleic acid having sequences complementary to one or more of the primer pairs. The method further includes conducting a PCR reaction in the partitions, thereby changing a fluorescent property of the first portion of the partitions, detecting the fluorescent property of each partition, and determining the number of occurrences in the

sample nucleic acid of one or more sequences of known length based on the detecting step. In some aspects of the invention, the method further includes comparing a first number of occurrences of a first sequence of known length to a second number of occurrences of a second sequence of a second known length.

Additional embodiments of the invention may also contemplate the use of a single primer pair as well as rely on something other than a probe for detecting the amplified sequence. In certain embodiments, the method comprises partitioning a sample comprising nucleic acid of different lengths into a plurality of partitioned portions, wherein each portion comprises, on average a single nucleic acid molecule. The method further includes introducing at least one primer pair, in which each primer of the pair is specific for a first and second location on the nucleic acid, the first and second locations being spaced apart from each other. The method further includes amplifying the nucleic acid in the partitioned portions, detecting the amplicons in the partitioned portions, and determining a nucleic acid make-up of the sample based on the results of the detecting step.

Methods in accordance with the invention also encompass the analysis of cell-free nucleic acids in a biological sample. In some embodiments, the biological sample can be blood, saliva, sputum, urine, semen, transvaginal fluid, cerebrospinal fluid, sweat, breast milk, breast fluid (e.g., breast nipple aspirate), stool, a cell or a tissue biopsy.

The methods of the invention can also be used to evaluate the quality of cell-free nucleic acids, for example cell-free DNA or RNA, which can be obtained from a biological sample. The invention allows the cell-free nucleic acid to be evaluated for quality, e.g., continuity, prior to amplification and sequencing. Thus, a cell-free nucleic acid sample can be partitioned into samples comprising nucleic acids of different lengths, primer pairs can be introduced along with appropriate probes, the nucleic acids amplified, and the make-up, e.g., the continuity of the cell-free nucleic acid sample can be determined.

Brief Description of the Drawings

Figure 1 depicts a droplet formation device.

Figure 2 depicts a portion of the droplet formation device of Figure 1.

Figures 3A-3C depicts an exemplary microfluidic system for droplet generation and readout. Figure 3A depicts the droplet generation chip; Figure 3B depicts the droplet spacing for readout; and Figure 3C depicts a cartoon of droplet readout by fluorescence.

Figures 4A-4C depicts the serial dilution of template DNA quantified by dPCR. Figure 4A shows droplet fluorescence during readout for the most concentrated sample. Each discrete burst of fluorescence corresponded to an individual droplet. Two different groups of droplets were evident: PCR(+) droplets peaking at ~0.8 V and PCR(-) droplets at ~0.1 V; Figure 4B shows a histogram of the peak fluorescence intensities of droplets from the complete data trace in (a). PCR(+) and PCR(-) droplets appeared as two very distinct populations centered at 0.78 and 0.10 V, respectively; Figure 4C shows the serial dilution of template DNA. Open circles: measured occupancies; solid line: the best fit to Eqn 2 ($A = 0.15$, $f = 4.8$, $R^2 = 0.9999$).

Figure 5A is a schematic representation of a droplet having 5 sets of primers for PCR amplification of a template sequence and 5 probes, each labeled with a fluorescent dye, that binds specifically to the amplified sequences; Figure 5B is a time trace of fluorescence intensity detected from droplets after PCR amplification; Figure 5C is a scatter plot showing clusters representing droplets that contain specific amplified sequences (TERT, RNaseP, E1a, SMN1 and SMN2). As shown in Figure 5A, five Taqman assays are conducted in a droplet. This provides for expression analysis and precise quantification of copy number while only requiring 1/5 the DNA of 5 opti-plex reaction. Each reaction has a unique location in the 2D scatter plot shown in Figure 5C.

Figure 6A is a schematic representation of a droplet having 5 sets of primers for PCR amplification of a template sequence and 5 probes, each labeled with a fluorescent dye, that binds specifically to the amplified sequences; Figure 6B is a scatter plot showing clusters representing droplets that contain specific amplified sequences (TERT, 815A, RNaseP, E1a, and 815G); Figure 6C is a table showing the copy number of specific sequences shown in Figure 6B. In Figure 6A, off-axis populations are generated by making blends of probes, which expands optical space available for multiplexing assays.

Figures 7A-E are a schematic depicting one-color detection of a genetic sequence with a microfluidic device.

Figures 8A-D are a schematic depicting two-color detection of two genetic sequences with a microfluidic device.

Figures 9A-D are a schematic depicting two-color detection of three genetic sequences with a microfluidic device.

Figure 10 shows two dot plots depicting clusters of genetic sequences detected through fluorescence intensity. Left panel is a dot plot showing four clusters. Block for SMN1 sequence was present. Top left: microdroplets containing the reference sequence (SMARCC1); bottom left: microdroplets not containing any sequence; bottom middle: microdroplets containing sequence for SMN1; and bottom right: microdroplets containing sequence for SMN2. Right panel is a dot plot showing four clusters. No block for SMN1 sequence was present. Top left: microdroplets containing the reference sequence (SMARCC1); bottom left: microdroplets not containing any sequence; bottom middle: microdroplets containing sequence for SMN1; and bottom right: microdroplets containing sequence for SMN2. The shift of the bottom middle cluster in right panel as compared to left panel confirms that fluorescence intensity provides a very sensitive measurement for the presence of a sequence.

Figures 11A and 11B depicts histograms of a duplex gene copy number assay using only one type of fluorophore by digital PCR; Figure 11Aa depicts a histogram of droplet peak fluorescence intensities; Figure 11B shows a comparison of gene copy numbers measured by monochromatic dPCR.

Figures 12A-12C are a schematic for tuning the intensity of a detectable label to a particular target with a microfluidic device. In Figures 12A-12C, relative intensity of signals from multiple targets of the same color can be generated in a variety of ways. Examples include: FIG. 12A) using a single base mismatch of the probe and target to distinguish different targets; FIG. 12B) blend identical probes with and without fluorophores; and FIG. 12C) blend identical probes with two or more different color fluorophores.

Figure 13 is a line graph depicting the linear dependence of droplet fluorescence intensity on probe concentration (Line, best linear fit ($y = -0.092x + 0.082$, $R^2 = 0.995$)).

Figures 14A-B depict a 5-plex dPCR assay for spinal muscular atrophy with only two fluorophores. Figure 14A is a 2D histogram of droplet fluorescence intensities, shown as a heat map, for the 5-plex assay against the synthetic model chromosome for validation. The six well resolved droplet populations corresponded to the five individual assays plus the empty droplets; Figure 14B shows the results of the SMA pilot study.

Figures 15A-B depict a 9-plex dPCR assay for spinal muscular atrophy with only two fluorophores, showing the process of optimizing droplet intensities. Figures 15A and 15B show 2-D histograms of droplet fluorescence intensity, shown as heat maps with hotter colors representing higher droplet counts, for the 9-plex assay against the synthetic model chromosome FIG. 15A= Before optimization; FIG. 15B = after optimization).

Figure 16 depicts an optical schematic for combining optical labels with multiplexing.

Figure 17 depicts a dPCR assay combining multiplexing with optical labels using co-flow microfluidics. The contributions from all droplets are shown, that is, from three different triplex assays. (Both panels) 2-D histograms shown as heat maps with hotter colors representing higher droplet counts. (Left panel) histogram of optical labels, i.e. fluorescence intensities of droplets measured at wavelengths for the two fluorophores comprising the optical labels. (Right panel) assay histogram, i.e. fluorescence intensities of droplets measured at wavelengths suitable for FAM detection (x-axis), and VIC detection (y-axis). Both histograms were compensated for spectral overlap by standard techniques.

Figures 18A-C show single assay selections using optical labels. Selections were taken from all of the droplets from Figure 17. Each of the three different selections in Figures A-C were for optical labels encoding the same assay (TERT, SMN1, and SMN2). Histograms are as described in Figure 17. (Left histograms, optical labels) Superimposed lines demark the bounding box for selecting a single optical label. (Right histograms, assay) Only droplets containing the selected optical label are displayed.

Figures 19A-C show single assay selections using optical labels. Selections were taken from all of the droplets from Figure 17. Each of the three different selections in Figures A-C was for optical labels encoding the same assay (TERT, c.5C from SMN1, and BCKDHA). Histograms are as described in Figure 17. (Left histograms, optical labels) Superimposed lines demark the bounding box for selecting a single optical label. (Right histograms, assay) Only droplets containing the selected optical label are displayed.

Figures 20A-C show single assay selections using optical labels. Selections were taken from all of the droplets from Figure 17. Each of the three different selections in Figures 20A-C was for optical labels encoding the same assay (TERT, c.88G from SMN1, and RNaseP). Histograms are as described in Figure 17. (Left histograms, optical labels) Superimposed lines

demarcate the bounding box for selecting a single optical label. (Right histograms, assay) Only droplets containing the selected optical label are displayed.

Figures 21A-J depict a dPCR assay combining multiplexing with optical labels using droplet merging.

Figure 22 is a schematic showing haplotype detection in droplets.

Figure 23 is a schematic showing several loci of interest on a pair of alleles and fragments of those alleles resulting from enzymatic digestion.

Figure 24 depicts a multiplex dPCR assay to determine the nucleic acid make-up of a sample containing nucleic acids shown schematically in Figure 23.

Figure 25 is a schematic showing a nucleic acid and a pair of primers, each primer specific for a separate location on the nucleic acid, and probe specific for a region of interest on the nucleic acid.

Figure 26 depicts dPCR assay conducted with a single pair of primers and a nucleic acid shown schematically in Figure 25.

Figure 27A depicts detection of droplets flowing single-file and spaced apart configuration through a channel.

Figure 27B depicts detection of droplets flowing in a non-single file configuration through a channel.

Figure 27C depicts detection of droplets flowing past the detector which adjacent and in contact with each other.

Figure 28A depicts threshold detection of droplets flowing past a detector in a mono-dispersed configuration.

Figure 28B depicts threshold detection of droplets flowing past a detector in a non-mono-dispersed configuration.

Figure 29 is a schematic showing an amplification strategy that inserts sequence adaptors and barcodes onto amplicons.

Figure 30A exemplifies fluorescent polarization detectors placed adjacent to each other for tracking polarized light emitted from a droplet over time according to one embodiment.

Figure 30B exemplifies fluorescent polarization detectors arranged to simultaneously obtain polarized light emitted from a droplet according to one embodiment.

Figure 31 depicts another configuration for detecting fluorescence polarization according to certain embodiments.

Figure 32 depicts another configuration for detecting fluorescence polarization according to certain embodiments.

Detailed Description

The invention provides materials and methods for analysis of biomolecules. In one aspect, the invention provides for digital analysis in droplets, such as microfluidic droplets. The invention allows digital PCR to be conducted and provides for significantly reduced or eliminated errors.

Ideally, the sensitivity of digital PCR is limited only by the number of independent amplifications that can be analyzed, which has motivated the development of several ultra-high throughput miniaturized methods allowing millions of single molecule PCR reactions to be performed in parallel (discussed in detail elsewhere). In a preferred embodiment of the invention, digital PCR is performed in aqueous droplets separated by oil using a microfluidics system. In another preferred embodiment, the oil is a fluorinated oil such as the Fluorinert oils (3M). In a still more preferred embodiment the fluorinated oil contains a surfactant, such as PFPE-PEG-PFPE triblock copolymer, to stabilize the droplets against coalescence during the amplification step or at any point where they contact each other. Microfluidic approaches allow the rapid generation of large numbers (e.g. 10^6 or greater) of very uniformly sized droplets that function as picoliter volume reaction vessels (see reviews of droplet-based microfluidics). But as will be described, the invention is not limited to dPCR performed in water-in-oil emulsions, but rather is general to all methods of reaction compartmentalization for dPCR. In the description that follows, the invention is described in terms of the use of droplets for compartmentalization, but it is understood that this choice of description is not limiting for the invention, and that all of the methods of the invention are compatible with all other methods of reaction compartmentalization for dPCR.

Methods of the invention involve novel strategies for performing multiple different amplification reactions on the same sample simultaneously to quantify the abundance of multiple different DNA targets, commonly known to those familiar with the art as “multiplexing”. Methods of the invention for multiplexing dPCR assays promise greater plexity—the number of

simultaneous reactions—than possible with existing qPCR or dPCR techniques. It is based on the singular nature of amplifications at terminal or limiting dilution that arises because most often only a single target allele is ever present in any one droplet even when multiple primers/probes targeting different alleles are present. This alleviates the complications that otherwise plague simultaneous competing reactions, such as varying arrival time into the exponential stage and unintended interactions between primers.

In one aspect, the invention provides materials and methods for improving amplicon yield while maintaining the sensitivity and specificity in droplet based digital PCR. More specifically, the invention provides droplets containing a single nucleic acid template and multiplexed PCR primers and methods for detecting a plurality of targets in a biological sample by forming such droplets and amplifying the nucleic acid templates using droplet based digital PCR.

Reactions within microfluidic droplets yield very uniform fluorescence intensity at the end point, and ultimately the intensity depends on the efficiency of probe hydrolysis. Thus, in another aspect of the methods of the invention, different reactions with different efficiencies can be discriminated on the basis of end point fluorescence intensity alone even if they have the same color. Furthermore, in another method of the invention, the efficiencies can be tuned simply by adjusting the probe concentration, resulting in an easy-to-use and general purpose method for multiplexing. In one demonstration of the invention, a 5-plex TaqMan[®] dPCR assay worked “right out of the box”, in contrast to lengthy optimizations that typify qPCR multiplexing to this degree. In another aspect of the invention, adding multiple colors increases the number of possible reactions geometrically, rather than linearly as with qPCR, because individual reactions can be labeled with multiple fluorophores. As an example, two fluorophores (VIC and FAM) were used to distinguish five different reactions in one implementation of the invention.

Methods of the invention are able to detect polymerase errors that occur during an amplification reaction and are able to exclude from analysis those products that are a result of polymerase errors. In essence, methods of the invention increase the sensitivity of digital PCR by identifying amplification products that are false positives, and excluding those products from analysis.

Methods of the invention involve forming sample droplets containing a single target nucleic acid, amplifying the target in the droplets, excluding droplets containing amplicon from the target and amplicon from a variant of the target, and analyzing target amplicons.

Nucleic Acid Target Molecules

Nucleic acid molecules include deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA). Nucleic acid molecules can be synthetic or derived from naturally occurring sources. In one embodiment, nucleic acid molecules are isolated from a biological sample containing a variety of other components, such as proteins, lipids and non-template nucleic acids. Nucleic acid template molecules can be obtained from any cellular material, obtained from an animal, plant, bacterium, fungus, or any other cellular organism. In certain embodiments, the nucleic acid molecules are obtained from a single cell. Biological samples for use in the present invention include viral particles or preparations. Nucleic acid molecules can be obtained directly from an organism or from a biological sample obtained from an organism, e.g., from blood, urine, cerebrospinal fluid, seminal fluid, saliva, sputum, stool and tissue. Any tissue or body fluid specimen may be used as a source for nucleic acid for use in the invention. Nucleic acid molecules can also be isolated from cultured cells, such as a primary cell culture or a cell line. The cells or tissues from which template nucleic acids are obtained can be infected with a virus or other intracellular pathogen. A sample can also be total RNA extracted from a biological specimen, a cDNA library, viral, or genomic DNA. In certain embodiments, the nucleic acid molecules are bound as to other target molecules such as proteins, enzymes, substrates, antibodies, binding agents, beads, small molecules, peptides, or any other molecule and serve as a surrogate for quantifying and / or detecting the target molecule.

Generally, nucleic acid can be extracted from a biological sample by a variety of techniques such as those described by Maniatis, et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y., pp. 280-281 (1982). Nucleic acid molecules may be single-stranded, double-stranded, or double-stranded with single-stranded regions (for example, stem- and loop-structures).

Droplet Formation

Methods of the invention involve forming sample droplets where some droplets contain

zero target nucleic acid molecules, some droplets contain one target nucleic acid molecule, and some droplets may or may not contain multiple nucleic acid molecules (corresponding to limiting or terminal dilution, respectively, as defined above). In the preferred embodiment, the distribution of molecules within droplets obeys the Poisson distribution. However, methods for non-Poisson loading of droplets are known to those familiar with the art, and include but are not limited to active sorting of droplets, such as by laser-induced fluorescence, or by passive one-to-one loading. The description that follows assumes Poisson loading of droplets, but such description is not intended to exclude non-Poisson loading, as the invention is compatible with all distributions of DNA loading that conform to limiting or terminal dilution.

The droplets are aqueous droplets that are surrounded by an immiscible carrier fluid. Methods of forming such droplets are shown for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163), Stone et al. (U.S. patent number 7,708,949 and U.S. patent application number 2010/0172803), Anderson et al. (U.S. patent number 7,041,481 and which reissued as RE41,780) and European publication number EP2047910 to Raindance Technologies Inc. The content of each of which is incorporated by reference herein in its entirety.

Figure 1 shows an exemplary embodiment of a device **100** for droplet formation. Device **100** includes an inlet channel **101**, and outlet channel **102**, and two carrier fluid channels **103** and **104**. Channels **101**, **102**, **103**, and **104** meet at a junction **105**. Inlet channel **101** flows sample fluid to the junction **105**. Carrier fluid channels **103** and **104** flow a carrier fluid that is immiscible with the sample fluid to the junction **105**. Inlet channel **101** narrows at its distal portion wherein it connects to junction **105** (See Figure 2). Inlet channel **101** is oriented to be perpendicular to carrier fluid channels **103** and **104**. Droplets are formed as sample fluid flows from inlet channel **101** to junction **105**, where the sample fluid interacts with flowing carrier fluid provided to the junction **105** by carrier fluid channels **103** and **104**. Outlet channel **102** receives the droplets of sample fluid surrounded by carrier fluid.

The sample fluid is typically an aqueous buffer solution, such as ultrapure water (e.g., 18 mega-ohm resistivity, obtained, for example by column chromatography), 10 mM Tris HCl and 1 mM EDTA (TE) buffer, phosphate buffer saline (PBS) or acetate buffer. Any liquid or buffer that is physiologically compatible with nucleic acid molecules can be used. The carrier fluid is one that is immiscible with the sample fluid. The carrier fluid can be a non-polar solvent, decane

(e.g., tetradecane or hexadecane), fluorocarbon oil, silicone oil or another oil (for example, mineral oil).

In certain embodiments, the carrier fluid contains one or more additives, such as agents which increase, reduce, or otherwise create non-Newtonian surface tensions (surfactants) and/or stabilize droplets against spontaneous coalescence on contact. Surfactants can include Tween, Span, fluorosurfactants, and other agents that are soluble in oil relative to water. In some applications, performance is improved by adding a second surfactant, or other agent, such as a polymer or other additive, to the sample fluid. Surfactants can aid in controlling or optimizing droplet size, flow and uniformity, for example by reducing the shear force needed to extrude or inject droplets into an intersecting channel. This can affect droplet volume and periodicity, or the rate or frequency at which droplets break off into an intersecting channel. Furthermore, the surfactant can serve to stabilize aqueous emulsions in fluorinated oils from coalescing.

In certain embodiments, the droplets may be coated with a surfactant or a mixture of surfactants. Preferred surfactants that may be added to the carrier fluid include, but are not limited to, surfactants such as sorbitan-based carboxylic acid esters (e.g., the "Span" surfactants, Fluka Chemika), including sorbitan monolaurate (Span 20), sorbitan monopalmitate (Span 40), sorbitan monostearate (Span 60) and sorbitan monooleate (Span 80), and perfluorinated polyethers (e.g., DuPont Krytox 157 FSL, FSM, and/or FSH). Other non-limiting examples of non-ionic surfactants which may be used include polyoxyethylenated alkylphenols (for example, nonyl-, p-dodecyl-, and dinonylphenols), polyoxyethylenated straight chain alcohols, polyoxyethylenated polyoxypropylene glycols, polyoxyethylenated mercaptans, long chain carboxylic acid esters (for example, glyceryl and polyglycerol esters of natural fatty acids, propylene glycol, sorbitol, polyoxyethylenated sorbitol esters, polyoxyethylene glycol esters, etc.) and alkanolamines (e.g., diethanolamine-fatty acid condensates and isopropanolamine-fatty acid condensates).

In certain embodiments, the carrier fluid may be caused to flow through the outlet channel so that the surfactant in the carrier fluid coats the channel walls. In one embodiment, the fluorosurfactant can be prepared by reacting the perfluorinated polyether DuPont Krytox 157 FSL, FSM, or FSH with aqueous ammonium hydroxide in a volatile fluorinated solvent. The solvent and residual water and ammonia can be removed with a rotary evaporator. The surfactant

can then be dissolved (e.g., 2.5 wt %) in a fluorinated oil (e.g., Fluorinert (3M)), which then serves as the carrier fluid.

One approach to merging sample fluids, using a device called a lambda injector, involves forming a droplet, and contacting the droplet with a fluid stream, in which a portion of the fluid stream integrates with the droplet to form a mixed droplet. In this approach, only one phase needs to reach a merge area in a form of a droplet. Further description of such method is shown in the co-owned and co-pending U.S. patent application to Yurkovetsky, et al. (U.S. patent application serial number 61/441,985), the content of which is incorporated by reference herein in its entirety.

According to a method for operating the lambda injector, a droplet is formed as described above. After formation of the sample droplet from the first sample fluid, the droplet is contacted with a flow of a second sample fluid stream. Contact between the droplet and the fluid stream results in a portion of the fluid stream integrating with the droplet to form a mixed droplet.

The droplets of the first sample fluid flow through a first channel separated from each other by immiscible carrier fluid and suspended in the immiscible carrier fluid. The droplets are delivered to the merge area, i.e., junction of the first channel with the second channel, by a pressure-driven flow generated by a positive displacement pump. While droplet arrives at the merge area, a bolus of a second sample fluid is protruding from an opening of the second channel into the first channel. Preferably, the channels are oriented perpendicular to each other. However, any angle that results in an intersection of the channels may be used.

The bolus of the second sample fluid stream continues to increase in size due to pumping action of a positive displacement pump connected to channel, which outputs a steady stream of the second sample fluid into the merge area. The flowing droplet containing the first sample fluid eventually contacts the bolus of the second sample fluid that is protruding into the first channel. Contact between the two sample fluids results in a portion of the second sample fluid being segmented from the second sample fluid stream and joining with the first sample fluid droplet to form a mixed droplet. In certain embodiments, each incoming droplet of first sample fluid is merged with the same amount of second sample fluid.

In certain embodiments, an electric charge is applied to the first and second sample fluids. Description of applying electric charge to sample fluids is provided in Link et al. (U.S. patent application number 2007/0003442) and European Patent Number EP2004316 to

Raindance Technologies Inc., the content of each of which is incorporated by reference herein in its entirety. Electric charge may be created in the first and second sample fluids within the carrier fluid using any suitable technique, for example, by placing the first and second sample fluids within an electric field (which may be AC, DC, etc.), and/or causing a reaction to occur that causes the first and second sample fluids to have an electric charge, for example, a chemical reaction, an ionic reaction, a photocatalyzed reaction, etc.

The electric field, in some embodiments, is generated from an electric field generator, i.e., a device or system able to create an electric field that can be applied to the fluid. The electric field generator may produce an AC field (i.e., one that varies periodically with respect to time, for example, sinusoidally, sawtooth, square, etc.), a DC field (i.e., one that is constant with respect to time), a pulsed field, etc. The electric field generator may be constructed and arranged to create an electric field within a fluid contained within a channel or a microfluidic channel. The electric field generator may be integral to or separate from the fluidic system containing the channel or microfluidic channel, according to some embodiments.

Techniques for producing a suitable electric field (which may be AC, DC, etc.) are known to those of ordinary skill in the art. For example, in one embodiment, an electric field is produced by applying voltage across a pair of electrodes, which may be positioned on or embedded within the fluidic system (for example, within a substrate defining the channel or microfluidic channel), and/or positioned proximate the fluid such that at least a portion of the electric field interacts with the fluid. The electrodes can be fashioned from any suitable electrode material or materials known to those of ordinary skill in the art, including, but not limited to, silver, gold, copper, carbon, platinum, copper, tungsten, tin, cadmium, nickel, indium tin oxide ("ITO"), etc., as well as combinations thereof. In some cases, transparent or substantially transparent electrodes can be used.

The electric field facilitates rupture of the interface separating the second sample fluid and the droplet. Rupturing the interface facilitates merging of bolus of the second sample fluid and the first sample fluid droplet. The forming mixed droplet continues to increase in size until it a portion of the second sample fluid breaks free or segments from the second sample fluid stream prior to arrival and merging of the next droplet containing the first sample fluid. The segmenting of the portion of the second sample fluid from the second sample fluid stream occurs as soon as the shear force exerted on the forming mixed droplet by the immiscible carrier fluid

overcomes the surface tension whose action is to keep the segmenting portion of the second sample fluid connected with the second sample fluid stream. The now fully formed mixed droplet continues to flow through the first channel.

In other embodiments, the rupture of the interface can be spontaneous, or the rupture can be facilitated by surface chemistry. The invention is not limited in regard to the method of rupture at the interface, as rupture can be brought about by any means.

In the context of PCR, in a preferred embodiment, the first sample fluid contains nucleic acid templates. Droplets of the first sample fluid are formed as described above. Those droplets will include the nucleic acid templates. In certain embodiments, the droplets will include only a single nucleic acid template, and thus digital PCR can be conducted. The second sample fluid contains reagents for the PCR reaction. Such reagents generally include Taq polymerase, deoxynucleotides of type A, C, G and T, magnesium chloride, and forward and reverse primers, all suspended within an aqueous buffer. The second fluid also includes detectably labeled probes for detection of the amplified target nucleic acid, the details of which are discussed below. A droplet containing the nucleic acid is then caused to merge with the PCR reagents in the second fluid as described above, producing a droplet that includes Taq polymerase, deoxynucleotides of type A, C, G and T, magnesium chloride, forward and reverse primers, detectably labeled probes, and the target nucleic acid. In another embodiment, the first fluid can contain the template DNA and PCR master mix (defined below), and the second fluid can contain the forward and reverse primers and the probe. The invention is not restricted in any way regarding the constituency of the first and second fluidics for PCR or digital PCR. For example, in some embodiments, the template DNA is contained in the second fluid inside droplets.

Target Amplification

Methods of the invention further involve amplifying the target nucleic acid in each droplet. Amplification refers to production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction or other technologies well known in the art (e.g., Dieffenbach and Dveksler, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y. [1995]). The amplification reaction may be any amplification reaction known in the art that amplifies nucleic acid molecules, such as polymerase chain reaction, nested polymerase chain reaction, ligase chain reaction (Barany F. (1991) PNAS 88:189-193; Barany F.

(1991) PCR Methods and Applications 1:5-16), ligase detection reaction (Barany F. (1991) PNAS 88:189-193), strand displacement amplification, transcription based amplification system, nucleic acid sequence-based amplification, rolling circle amplification, and hyper-branched rolling circle amplification.

In certain embodiments, the amplification reaction is the polymerase chain reaction. Polymerase chain reaction (PCR) refers to methods by K. B. Mullis (U.S. patent numbers 4,683,195 and 4,683,202, hereby incorporated by reference) for increasing concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. The process for amplifying the target sequence includes introducing an excess of oligonucleotide primers to a DNA mixture containing a desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The primers are complementary to their respective strands of the double stranded target sequence.

To effect amplification, primers are annealed to their complementary sequence within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one cycle; there can be numerous cycles) to obtain a high concentration of an amplified segment of a desired target sequence. The length of the amplified segment of the desired target sequence is determined by relative positions of the primers with respect to each other and by cycling parameters, and therefore, this length is a controllable parameter.

Methods for performing PCR in droplets are shown for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163), Anderson et al. (U.S. patent number 7,041,481 and which reissued as RE41,780) and European publication number EP2047910 to Raindance Technologies Inc. The content of each of which is incorporated by reference herein in its entirety.

The sample droplet may be pre-mixed with a primer or primers, or the primer or primers may be added to the droplet. In some embodiments, droplets created by segmenting the starting sample are merged with a second set of droplets including one or more primers for the target nucleic acid in order to produce final droplets. The merging of droplets can be accomplished using, for example, one or more droplet merging techniques described for example in Link et al.

(U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc.

In embodiments involving merging of droplets, two droplet formation modules are used. In one embodiment, a first droplet formation module produces the sample droplets consistent with limiting or terminal dilution of target nucleic acid. A second droplet formation or reinjection module inserts droplets that contain reagents for a PCR reaction. Such droplets generally include the “PCR master mix” (known to those in the art as a mixture containing at least Taq polymerase, deoxynucleotides of type A, C, G and T, and magnesium chloride) and forward and reverse primers (known to those in the art collectively as “primers”), all suspended within an aqueous buffer. The second droplet also includes detectably labeled probes for detection of the amplified target nucleic acid, the details of which are discussed below. Different arrangements of reagents between the two droplet types is envisioned. For example, in another embodiment, the template droplets also contain the PCR master mix, but the primers and probes remain in the second droplets. Any arrangement of reagents and template DNA can be used according to the invention.

Primers can be prepared by a variety of methods including but not limited to cloning of appropriate sequences and direct chemical synthesis using methods well known in the art (Narang et al., *Methods Enzymol.*, 68:90 (1979); Brown et al., *Methods Enzymol.*, 68:109 (1979)). Primers can also be obtained from commercial sources such as Operon Technologies, Amersham Pharmacia Biotech, Sigma, and Life Technologies. The primers can have an identical melting temperature. The lengths of the primers can be extended or shortened at the 5' end or the 3' end to produce primers with desired melting temperatures. Also, the annealing position of each primer pair can be designed such that the sequence and, length of the primer pairs yield the desired melting temperature. The simplest equation for determining the melting temperature of primers smaller than 25 base pairs is the Wallace Rule ($T_d = 2(A+T) + 4(G+C)$). Another method for determining the melting temperature of primers is the nearest neighbor method. Computer programs can also be used to design primers, including but not limited to Array Designer Software (Arrayit Inc.), Oligonucleotide Probe Sequence Design Software for Genetic Analysis (Olympus Optical Co.), NetPrimer, and DNAsis from Hitachi Software Engineering. The T_M (melting or annealing temperature) of each primer is calculated using software programs such as Oligo Design, available from Invitrogen Corp.

According to certain aspects, primers suitable for use in methods of the invention each include a first targeting arm coupled to a second targeting arm via a bridge section (e.g. the first and second targeting arms flank the bridge section). The first and second targeting arms of a primer are designed to hybridize (at least partially) to different nucleic acids on a single nucleic acid template. That is, the first targeting arm targets a first region and the second targeting arm targets a second region. The first and second regions may be separated by one or more nucleotides. The bridge section includes a sequence that does not hybridize to any naturally occurring sequence. Accordingly, when bound, the bridge section of the primer do not hybridize to nucleic acid on the template nucleic acid. Each droplet includes primer pairs that hybridize to different nucleic acid templates. For example, a single droplet may include primer pairs in which a set of the primer hybridize to sites on chromosome 21, a set of the primer hybridize to sites on chromosome 10, a set of the primer hybridize to sites on chromosome 3, a set of the primer hybridize to sites on chromosome 2, etc. Any number of different sets of primer pairs may be put into droplets. For example, some droplets may include sets of primer pairs against two or more different chromosomes, three or more different chromosomes, five or more different chromosomes, ten or more different chromosomes, twenty or more different chromosomes, or all 24 chromosomes.

The target site may include a few nucleotides to about 100 or more nucleotides on the single nucleic acid template. For example, primers can be designed to capture targets having lengths in the range of up to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more bases (or base pairs). It should be appreciated that primers can be used to target single-stranded or double-stranded template nucleic acid. It is to be appreciated that the length of the target site of the template nucleic acid may be selected based upon multiple considerations (e.g. constraints on the droplet size). In one example, where analysis of the target involves sequencing, e.g., with a next-generation sequencer, the target length should typically match the sequencing read-length so that shotgun library construction is not necessary. However, it should be appreciated that captured nucleic acids may be sequenced using any suitable sequencing technique as aspects of the invention are not limited in this respect.

The targeting arms of primers according to certain aspects may be of the same length or of different lengths. The targeting arms may consist of any number of nucleotides (e.g., in the ranges of 1-200 bases) configured to hybridize to specific regions of the target template. The

targeting arms of each primer of the plurality of primers within a droplet are either, for example, chromosome-specific or mutation-specific.

The bridge section (flanked by the first and second targeting arms) is formed from a nucleotide sequence that is not complementary to and does not hybridize to the single nucleic acid template. The bridge section may include one or more nucleotides, which may be the same or different from the number of nucleotides spanning between the first region and the second region of the target site. For multiplex reactions, the bridge section of a primer corresponds to a unique probe that is designed to hybridize to a complement of the bridge section. This allows a primer hybridized to a target site to be detected via the probe hybridized to a complement of the bridge section after the amplification reaction. For primer pairs, the bridge section of each primer of a primer pair may be the same or different. Preferably, each primer of a primer pair has the same bridge section so that a corresponding probe will bind to either primer for detection of a target site specific to the primer pair.

Probes designed to hybridize to a complement of a primer bridge region allow the target template nucleic acid to be detected. Preferably, the droplet contains two or more probes, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 60, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 500, or more probes, each being specific to a complement of a bridge region. A probe may be designed to hybridize to the complement of the bridge region or a portion thereof. Each probe (or member) of a plurality of probes in a droplet may include a detectable label. Probes of the plurality of probes may be grouped together and at varying concentrations within the droplet. Members (probes) of each group may include the same detectable label. Alternatively, each probe may include a unique detectable label. Types of detectable labels suitable for use with probes specific to bridge regions of a primer and other probes for use in methods of the invention are described hereinafter.

In one embodiment, the droplet formation modules are arranged and controlled to produce an interdigitation of sample droplets and PCR reagent droplets flowing through a channel. Such an arrangement is described for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc.

A sample droplet is then caused to merge with a PCR reagent droplet, producing a droplet that includes the PCR master mix, primers, detectably labeled probes, and the target nucleic acid.

Droplets may be merged for example by: producing dielectrophoretic forces on the droplets using electric field gradients and then controlling the forces to cause the droplets to merge; producing droplets of different sizes that thus travel at different velocities, which causes the droplets to merge; and producing droplets having different viscosities that thus travel at different velocities, which causes the droplets to merge with each other. Each of those techniques is further described in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc. Further description of producing and controlling dielectrophoretic forces on droplets to cause the droplets to merge is described in Link et al. (U.S. patent application number 2007/0003442) and European Patent Number EP2004316 to Raindance Technologies Inc.

In another embodiment, called simple droplet generation, a single droplet formation module, or a plurality of droplet formation modules are arranged to produce droplets from a mixture already containing the template DNA, the PCR master mix, primers, and detectably labeled probes. In yet another embodiment, called co-flow, upstream from a single droplet formation module two channels intersect allowing two flow streams to converge. One flow stream contains one set of reagents and the template DNA, and the other contains the remaining reagents. In the preferred embodiment for co-flow, the template DNA and the PCR master mix are in one flow stream, and the primers and probes are in the other. However, the invention is not limited in regard to the constituency of either flow stream. For example, in another embodiment, one flow stream contains just the template DNA, and the other contains the PCR master mix, the primers, and the probes. On convergence of the flow streams in a fluidic intersection, the flow streams may or may not mix before the droplet generation nozzle. In either embodiment, some amount of fluid from the first stream, and some amount of fluid from the second stream are encapsulated within a single droplet. Following encapsulation, complete mixing occurs.

Once final droplets have been produced by any of the droplet forming embodiments above, or by any other embodiments, the droplets are thermal cycled, resulting in amplification of the target nucleic acid in each droplet. In certain embodiments, the droplets are collected off-chip as an emulsion in a PCR thermal cycling tube and then thermally cycled in a conventional thermal cycler. Temperature profiles for thermal cycling can be adjusted and optimized as with any conventional DNA amplification by PCR.

In certain embodiments, the droplets are flowed through a channel in a serpentine path between heating and cooling lines to amplify the nucleic acid in the droplet. The width and depth of the channel may be adjusted to set the residence time at each temperature, which can be controlled to anywhere between less than a second and minutes.

In certain embodiments, the three temperature zones are used for the amplification reaction. The three temperature zones are controlled to result in denaturation of double stranded nucleic acid (high temperature zone), annealing of primers (low temperature zones), and amplification of single stranded nucleic acid to produce double stranded nucleic acids (intermediate temperature zones). The temperatures within these zones fall within ranges well known in the art for conducting PCR reactions. See for example, Sambrook et al. (Molecular Cloning, A Laboratory Manual, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).

In certain embodiments, the three temperature zones are controlled to have temperatures as follows: 95⁰C (T_H), 55⁰C (T_L), 72⁰C (T_M). The prepared sample droplets flow through the channel at a controlled rate. The sample droplets first pass the initial denaturation zone (T_H) before thermal cycling. The initial preheat is an extended zone to ensure that nucleic acids within the sample droplet have denatured successfully before thermal cycling. The requirement for a preheat zone and the length of denaturation time required is dependent on the chemistry being used in the reaction. The samples pass into the high temperature zone, of approximately 95⁰C, where the sample is first separated into single stranded DNA in a process called denaturation. The sample then flows to the low temperature, of approximately 55⁰C, where the hybridization process takes place, during which the primers anneal to the complementary sequences of the sample. Finally, as the sample flows through the third medium temperature, of approximately 72⁰C, the polymerase process occurs when the primers are extended along the single strand of DNA with a thermostable enzyme. Methods for controlling the temperature in each zone may include but are not limited to electrical resistance, peltier junction, microwave radiation, and illumination with infrared radiation.

The nucleic acids undergo the same thermal cycling and chemical reaction as the droplets passes through each thermal cycle as they flow through the channel. The total number of cycles in the device is easily altered by an extension of thermal zones or by the creation of a continuous

loop structure. The sample undergoes the same thermal cycling and chemical reaction as it passes through N amplification cycles of the complete thermal device.

In other embodiments, the temperature zones are controlled to achieve two individual temperature zones for a PCR reaction. In certain embodiments, the two temperature zones are controlled to have temperatures as follows: 95°C (T_H) and 60°C (T_L). The sample droplet optionally flows through an initial preheat zone before entering thermal cycling. The preheat zone may be important for some chemistry for activation and also to ensure that double stranded nucleic acid in the droplets are fully denatured before the thermal cycling reaction begins. In an exemplary embodiment, the preheat dwell length results in approximately 10 minutes preheat of the droplets at the higher temperature.

The sample droplet continues into the high temperature zone, of approximately 95°C , where the sample is first separated into single stranded DNA in a process called denaturation. The sample then flows through the device to the low temperature zone, of approximately 60°C , where the hybridization process takes place, during which the primers anneal to the complementary sequences of the sample. Finally the polymerase process occurs when the primers are extended along the single strand of DNA with a thermostable enzyme. The sample undergoes the same thermal cycling and chemical reaction as it passes through each thermal cycle of the complete device. The total number of cycles in the device is easily altered by an extension of block length and tubing.

In another embodiment the droplets are created and/or merged on chip followed by their storage either on the same chip or another chip or off chip in some type of storage vessel such as a PCR tube. The chip or storage vessel containing the droplets is then cycled in its entirety to achieve the desired PCR heating and cooling cycles.

In another embodiment the droplets are collected in a chamber where the density difference between the droplets and the surrounding oil allows for the oil to be rapidly exchanged without removing the droplets. The temperature of the droplets can then be rapidly changed by exchange of the oil in the vessel for oil of a different temperature. This technique is broadly useful with two and three step temperature cycling or any other sequence of temperatures.

The invention is not limited by the method used to thermocycle the droplets. Any method of thermocycling the droplets may be used.

Target Detection

After amplification, droplets are flowed to a detection module for detection of amplification products. For embodiments in which the droplets are thermally cycled off-chip, the droplets require re-injection into either a second fluidic circuit for read-out—that may or may not reside on the same chip as the fluidic circuit or circuits for droplet generation—or in certain embodiments the droplets may be re-injected for read-out back into the original fluidic circuit used for droplet generation. The droplets may be individually analyzed and detected using any methods known in the art, such as detecting the presence or amount of a reporter. Generally, the detection module is in communication with one or more detection apparatuses. The detection apparatuses can be optical or electrical detectors or combinations thereof. Examples of suitable detection apparatuses include optical waveguides, microscopes, diodes, light stimulating devices, (e.g., lasers), photo multiplier tubes, and processors (e.g., computers and software), and combinations thereof, which cooperate to detect a signal representative of a characteristic, marker, or reporter, and to determine and direct the measurement or the sorting action at a sorting module. Further description of detection modules and methods of detecting amplification products in droplets are shown in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc.

In certain embodiments, amplified target are detected using detectably labeled probes. In particular embodiments, the detectably labeled probes are optically labeled probes, such as fluorescently labeled probes. Examples of fluorescent labels include, but are not limited to, Atto dyes, 4-acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic acid; acridine and derivatives: acridine, acridine isothiocyanate; 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS); 4-amino-N-[3-vinylsulfonyl]phenyl]naphthalimide-3,5 disulfonate; N-(4-anilino-1-naphthyl)maleimide; anthranilamide; BODIPY; Brilliant Yellow; coumarin and derivatives; coumarin, 7-amino-4-methylcoumarin (AMC, Coumarin 120), 7-amino-4-trifluoromethylcoumarin (Coumarin 151); cyanine dyes; cyanosine; 4',6-diaminidino-2-phenylindole (DAPI); 5'5"-dibromopyrogallol-sulfonaphthalein (Bromopyrogallol Red); 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin; diethylenetriamine pentaacetate; 4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid; 4,4'-diisothiocyanatostilbene-2,2'-

disulfonic acid; 5-[dimethylamino]naphthalene-1-sulfonyl chloride (DNS, dansylchloride); 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC); eosin and derivatives; eosin, eosin isothiocyanate, erythrosin and derivatives; erythrosin B, erythrosin, isothiocyanate; ethidium; fluorescein and derivatives; 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF), 2',7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein, fluorescein, fluorescein isothiocyanate, QFITC, (XRITC); fluorescamine; IR144; IR1446; Malachite Green isothiocyanate; 4-methylumbelliferoneortho cresolphthalein; nitrotyrosine; pararosaniline; Phenol Red; B-phycoerythrin; o-phthalaldehyde; pyrene and derivatives: pyrene, pyrene butyrate, succinimidyl 1-pyrene; butyrate quantum dots; Reactive Red 4 (Cibacron. Brilliant Red 3B-A) rhodamine and derivatives: 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride rhodamine (Rhod), rhodamine B, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101, sulfonyl chloride derivative of sulforhodamine 101 (Texas Red); N,N,N',N'tetramethyl-6-carboxyrhodamine (TAMRA); tetramethyl rhodamine; tetramethyl rhodamine isothiocyanate (TRITC); riboflavin; rosolic acid; terbium chelate derivatives; Cy3; Cy5; Cy5.5; Cy7; IRD 700; IRD 800; La Jolla Blue; phthalocyanine; and naphthalocyanine. Preferred fluorescent labels are FAM and VICTM (from Applied Biosystems). Labels other than fluorescent labels are contemplated by the invention, including other optically-detectable labels.

In certain aspects, the droplets of the invention contain a plurality of detectable probes that hybridize to amplicons produced in the droplets. Members of the plurality of probes can each include the same detectable label, or a different detectable label. The plurality of probes can also include one or more groups of probes at varying concentration. The groups of probes at varying concentrations can include the same detectable label which vary in intensity, due to varying probe concentrations. In some embodiments, the droplets of the invention contain a plurality of barcodes that hybridize to amplicons produced in the droplets or are incorporated into the amplicons. The barcodes may be used in lieu of fluorescent probes, to detect the presence of a target sequence, or the barcodes can be used in addition to fluorescent probes, to track a multitude of sample sources. A detectable barcode-type label can be any barcode-type label known in the art including, for example, barcoded magnetic beads (e.g., from Applied Biocode, Inc., Santa Fe Springs, CA), and nucleic acid sequences. Nucleic acid barcode sequences typically include a set of oligonucleotides ranging from about 4 to about 20

oligonucleotide bases (e.g., 8-10 oligonucleotide bases) and uniquely encode a discrete library member without containing significant homology to any sequence in the targeted sample.

The barcode sequence generally includes features useful in sequencing reactions. For example, the barcode sequences are designed to have minimal or no homopolymer regions, i.e., 2 or more of the same base in a row such as AA or CCC, within the barcode sequence. The barcode sequences are also designed so that they are at least one edit distance away from the base addition order when performing base-by-base sequencing, ensuring that the first and last base do not match the expected bases of the sequence. In certain embodiments, the barcode sequences are designed to be correlated to a particular subject, allowing subject samples to be distinguished. Designing barcodes is shown U.S. Pat. 6,235,475, the contents of which are incorporated by reference herein in their entirety.

In some instances, the primers used in the invention (including, e.g., primers having targeting arms flanked with a bridge section) may include barcodes such that the barcodes will be incorporated into the amplified products. For example, the unique barcode sequence could be incorporated into the 5' end of the primer, or the barcode sequence could be incorporated into the 3' end of the primer. In some embodiments, the barcodes may be incorporated into the amplified products after amplification. For example, a suitable restriction enzyme (or other endonuclease) may be introduced to a sample, e.g., a droplet, where it will cut off an end of an amplification product so that a barcode can be added with a ligase.

Attaching barcode sequences to nucleic acids is shown in U.S. Pub. 2008/0081330 and PCT/US09/64001, the content of each of which is incorporated by reference herein in its entirety. Methods for designing sets of barcode sequences and other methods for attaching barcode sequences are shown in U.S. patent numbers 6,138,077; 6,352,828; 5,636,400; 6,172,214; 6,235,475; 7,393,665; 7,544,473; 5,846,719; 5,695,934; 5,604,097; 6,150,516; RE39,793; 7,537,897; 6,172,218; and 5,863,722, the content of each of which is incorporated by reference herein in its entirety.

In a separate embodiment the detection can occur by the scanning of droplets confined to a monolayer in a storage device that is transparent to the wavelengths or method of detection. Droplets stored in this fashion can be scanned either by the movement of the storage device by the scanner or the movement of the scanner over the storage device.

The invention is not limited to the TaqMan assay, as described above, but rather the invention encompasses the use of all fluorogenic DNA hybridization probes, such as molecular beacons, Solaris probes, scorpion probes, and any other probes that function by sequence specific recognition of target DNA by hybridization and result in increased fluorescence on amplification of the target sequence.

Digital PCR Performance in Droplets

Digital PCR performance in the emulsion format was validated by measuring a serial dilution of a reference gene, branched chain keto acid dehydrogenase E1 (BCKDHA). Mixtures of the PCR master mix, 1× primers and probe for BCKDHA, and varying concentrations of a mixture of human genomic DNA (1:1 NA14091 and NA13705) were compartmentalized into over one million 5.3 pL droplets in a water-in-fluorinated oil emulsion using the droplet generation microfluidic chip. The emulsion was thermally cycled off-chip and afterwards the fluorescence of each droplet was analyzed by fluorescence in the readout chip (see Figs. 3A-3C).

An exemplary microfluidic system for droplet generation and readout is depicted in Figs. 3A-3C. The microfluidic system for droplet generation and readout. As shown in Fig. 3A (droplet generation chip), a continuous aqueous phase containing the PCR master mix, primers, and probes, and template DNA flowed into the fluidic intersection from the left, and the carrier oil entered from the top and bottom. An emerging bolus of aqueous liquid was imaged inside the intersection just prior to snapping off into a discrete 4 pL droplet as the fluidic strain began to exceed the surface tension of the aqueous liquid. The steady train of droplets leaving the intersection toward the right was collected off chip as a stable emulsion for thermal cycling. Figure 3B depicts the droplet spacing for readout. Flows were arranged as in FIG. 3A, except instead of a continuous phase, the emulsion from (a) was injected from the left into the intersection after thermal cycling. The oil drained from the emulsion during off-chip handling, hence the emulsion appeared tightly packed in the image before the intersection. The oil introduced in the intersection separated the droplets and the fluorescence of each droplet was measured at the location marked by the arrow. Figure 3C depicts a cartoon of droplet readout by fluorescence. The relatively infrequent PCR(+) droplets (light gray) flow along with the majority of PCR(-) droplets (dark gray) toward the detector. The droplets were interrogated sequentially by laser induced fluorescence while passing through the detection region.

In a serial dilution the average number of target DNA molecules per droplet—called the “occupancy” from this point forward—should decrease in direct proportion to the DNA concentration. The occupancy was calculated from Poisson statistics using the following equation well known to those experienced in the art:

$$\text{occupancy} = \ln \left(\frac{P+N}{N} \right), \quad (1)$$

where P and N are the numbers of PCR(+) and PCR(−) droplets respectively.

Droplets were analyzed by fluorescence while flowing through the readout chip to count the numbers of PCR(+) and PCR(−) droplets (see Fig. 3C). As each droplet passed the detection zone (marked with an arrow in Fig. 3B), a burst of fluorescence was observed. To account for small run-to-run differences in the fluorescence intensity that can occur due to different chip positioning, etc., each set of data was scaled such that the average fluorescence intensity of the empty droplets was 0.1 V. Fig. 4A shows a very short duration of a typical trace of fluorescence bursts from individual droplets for the sample with the highest DNA concentration in the series. PCR(+) and PCR(−) droplets were easily discriminated by fluorescence intensity. The two large bursts of fluorescence peaking at ~0.8 V arose from the PCR(+) droplets, whereas the smaller bursts due to incomplete fluorescence quenching in the PCR(−) droplets peaked at ~0.1 V. A histogram of peak intensities from the complete data set revealed two clear populations centered at 0.10 and 0.78 V (Fig. 4B), demonstrating that the trend evident in the short trace in Fig. 4A was stable over much longer periods of time. Integration over the two populations in Fig. 4B yielded a total of 197,507 PCR(+) and 1,240,126 PCR(−) droplets. Hence the occupancy was 0.15 for this sample by Eqn. 1, corresponding to the expected occupancy of 0.18 based on the measured DNA concentration of 110 ng/μL. The occupancy was measured for each sample in the serial dilution and fit to the dilution equation:

$$\text{occupancy}(n) = \frac{A}{f^n}, \quad (2)$$

where n is the number of dilutions, A is the occupancy at the starting concentration ($n = 0$), and f is the dilution factor. The linear fit was in excellent agreement with the data, with an R^2 value of 0.9999 and the fitted dilution factor of 4.8 in close agreement with the expected value of 5.0.

Multiplexing Primers in a Digital PCR Reaction

Droplet based digital PCR technology, as described in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163), Anderson et al. (U.S. patent number

7,041,481 and which reissued as RE41,780) and European publication number EP2047910 to Raindance Technologies Inc, (the contents of each of which are incorporated by reference herein in their entireties) utilizes a single primer pair per library droplet. This library droplet is merged with a template droplet which contains all the PCR reagents including genomic DNA except for the primers. After merging of the template and the primer library droplets the new droplet now contains all the reagents necessary to perform PCR. The droplet is then thermal cycled to produce amplicons. In one embodiment, the template DNA is diluted in the template mix such that on average there is less than one haploid genome per droplet.

Having only one haploid genome (i.e., one allele) per droplet gives droplet PCR advantages over standard singleplex or multiplex PCR in tubes or microwells. For example, in traditional PCR, both alleles are present in the reaction mix so if there is a difference in the PCR efficiency between alleles, the allele with the highest efficiency will be over represented. Additionally, there can be variances in the sequence to which the PCR primers hybridize, despite careful primer design. A variance in the primer hybridization sequence can cause that primer to have a lower efficiency for hybridization for the allele that has the variance compared to the allele that has the wild type sequence. This can also cause one allele to be amplified preferentially over the other allele if both alleles are present in the same reaction mix.

These issues are avoided in droplet based PCR because there is only one template molecule per droplet, and thus one allele per droplet. Thus, even if primer variance exists that reduces the PCR efficiency for one allele, there is no competition between alleles because the alleles are separated and thus uniformly amplified.

Optimization of traditional multiplexing of standard PCR primers in tubes or wells is known to be difficult. Multiple PCR amplicons being generated in the same reaction can lead to competition between amplicons that have differing efficiencies due to differences in sequence or length. This results in varying yields between competing amplicons which can result in non uniform amplicon yields. However, because droplet based digital PCR utilizes only one template molecule per droplet, even if there are multiple PCR primer pairs present in the droplet, only one primer pair will be active. Since only one amplicon is being generated per droplet, there is no competition between amplicons, resulting in a more uniform amplicon yield between different amplicons.

A certain amount of DNA is required to generate either a specific quantity of DNA and/or a specific number of PCR positive droplets to achieve sufficient sequencing coverage per base. Because only a percentage of the droplets are PCR positive, approximately 1 in 3 in the standard procedure, it takes more DNA to achieve the equivalent PCR yield per template DNA molecule. The number of PCR positive droplets and thus the amplicon yield can be increased by adding more genomic DNA. For instance, increasing the amount of genomic DNA twofold while maintaining the number of droplets constant will double the amplicon yield. However there is a limit to the amount of genomic DNA that can be added before there is a significant chance of having both alleles for a gene in the same droplet, thereby eliminating the advantage of droplet PCR for overcoming allele specific PCR and resulting in allelic dropout.

One way to allow the input of more genomic DNA is by generating more droplets to keep the haploid molecules per droplet ratio constant. For instance doubling the amount of DNA and doubling the amount of droplets increases the amplicon yield by 2x while maintaining the same haploid genome per droplet ratio. However, while doubling the number of droplets isn't problematic, increasing the amount of DNA can be challenging to users that have a limited amount of DNA.

The multiplexing of PCR primers in droplets enables the simultaneous increase in the number of PCR droplets while keeping the amount of input DNA the same or lower to generate an equal or greater amplicon yield. This results in an overall increase in the amount of PCR positive droplets and amplicon yield without the consumption of more DNA.

By way of example, if there is an average of 1 haploid genome per every 4 droplets or $\frac{1}{4}$ of the haploid genome per droplet and one PCR primer pair per droplet, the chances of the correct template being present for the PCR primer in the droplet is 1 out of 4. However, if there are 2 PCR primer pairs per droplet, then there is double the chance that there will be the correct template present in the droplet. This results in 1 out of 2 droplets being PCR positive which doubles the amplicon yield without doubling the input DNA. If the number of droplets containing the 2x multiplexed primers is doubled and the DNA kept constant, then the number of PCR positive droplets drops back to 1 in 4, but the total number of PCR droplets remains the same because the number of droplets have been doubled. If the multiplexing level in each droplet is increased to 4x and the input DNA is the same, the chance of the correct template molecule being present in each droplet doubles. This results in the number of PCR positive

droplets being increased to 1 in 2 which doubles the amount of amplicon yield without increasing the amount of input DNA. Thus, by increasing the multiplexing of PCR primers in each droplet and by increasing the number of droplets overall, the amplicon yield can be increased by 4-fold without increasing the amount of input DNA.

Alternatively, if the amplicon yield is already sufficient, by increasing the multiplexing level for the PCR primers in each droplet, the amount of input genomic DNA can be dropped without sacrificing amplicon yield. For example if the multiplexing level of the PCR primers goes from 1x to 2x, the amount of input genomic DNA can be decreased by 2x while still maintaining the same overall amplicon yield.

Even though the number of PCR primer pairs per droplet is greater than one, there is still only one template molecule per droplet and thus there is only one primer pair per droplet that is being utilized at one time. This means that the advantages of droplet PCR for eliminating bias from either allele specific PCR or competition between different amplicons is maintained.

An example demonstration of droplet-based amplification and detection of multiple target sequences in a single droplet is shown here. Multiple copies of 5 sets of primers (primers for TERT, RNaseP, E1a, SMN1 and SMN2) were encapsulated in a single droplet at various concentrations along with the template DNA and the PCR master mix. Probes that specifically bind to TERT, RNaseP, E1a, SMN1 or SMN2 were also encapsulated in the droplets containing the primers. Probes for TERT, RNaseP and E1a were labeled with the VIC dye and probes for SMN1 and SMN2 were labeled with the FAM dye. The sequences for TERT RNaseP, E1a, SMN1 and SMN2 were amplified by PCR. The PCR was conducted with a standard thermal cycling setting. For example:

95 °C for 10 min

31 cycles

92 °C for 15 s

60 °C for 60 s

At the end of the PCR, the fluorescence emission from each droplet was determined and plotted on a scattered plot based on its wavelength and intensity. Six clusters, each representing droplets having the corresponding fluorescence wavelength and intensity were shown. The TERT, RNaseP and E1a clusters showed the fluorescence of the VIC dye at three distinct intensities and SMN1 and SMN1 clusters showed the fluorescence of the FAM dye at two

distinct intensities (Figure 5). The number of droplets, each having one or more sequences selected from TERT, RNaseP, E1a, SMN1 and SMN2, can be determined from the scattered plot.

In an another demonstration of droplet-based amplification and detection of multiple target sequences in a single droplet, 5 sets of primers (primers for TERT, RNaseP, E1a, 815A and 815G) were encapsulated in a single droplet at various concentrations along with the template DNA, the PCR master mix, and the probes. The five different probes TERT, RNaseP, E1a, 815A and 815G were also encapsulated in the droplets containing the primers. Probes for TERT and 815A were labeled with the VIC dye and probes for 815G were labeled with the FAM dye. For each of RNaseP and E1a, two probes, one labeled with the VIC dye and the other labeled with the FAM dye, were encapsulated.

The droplets containing both the primers and probes were fused with droplets containing the template. PCR reactions were conducted with the fused droplets to amplify the sequences for TERT, RNaseP, E1a, 815A and 815G. The PCR was conducted with a standard thermal cycling setting.

At the end of the PCR, the fluorescence emission from each fused droplet was determined and plotted on a scattered plot based on its wavelength and intensity. Six clusters, each representing droplets having the corresponding fluorescence wavelength and intensity were shown. The TERT and 815A clusters showed the fluorescence of the VIC dye at two distinct intensities; the 815G clusters showed the fluorescence of the FAM dye; and the RNaseP and E1a clusters showed the fluorescence of both the FAM and the VIC dye at distinct intensities (Figure 6). The number of droplets, each having one or more sequences selected from TERT, RNaseP, E1a, 815A and 815G, can be determined from the scattered plot. The copy number of RNaseP, E1a, 815A and 815G in the template were determined by the ratio between the number of droplets having the RNaseP, E1a, 815A and/or 815G sequences and the number of droplets having the TERT sequence (Figure 6).

In yet another exemplary demonstration of multiplexed primer pairs in a droplet-based digital PCR reaction, two droplet libraries were generated: droplet library A was generated where each droplet contained only one primer pair; and droplet library B was generated where the primer pairs were multiplexed at 5x level in each droplet. HapMap sample NA18858 was processed in duplicate with droplet libraries A or B using standard procedures. Two μ g sample

DNA was used for droplet library A and one μg sample DNA was used for the 5x multiplex droplet library B. After PCR amplification, both droplet libraries were broken and purified over a Qiagen MinElute column and then run on an Agilent Bioanalyzer. Samples were sequenced by Illumina on the Illumina GAII with 50 nucleotide reads and the sequencing results were analyzed using the standard sequencing metrics. The results from the 5x multiplexed droplet library B were compared to the singleplex droplet library A using standard metrics shown in the Table below.

The results obtained from the 5x multiplexed droplet library B were equivalent or better than what was obtained from droplet library A. The multiplexing of primers delivers the same sequencing results for base coverage, specificity and uniformity that the singleplexing does with the added advantage of reduced input DNA.

Sample	Total reads	Mapped reads	Specificity	Mean base coverage	C1	C20	C100	Base coverage (0.2x of mean)
Library A with sample 1	27431697	99.4%	0.813	1394	99.5%	99.0%	98.2%	92.8%
Library A with sample 2	15147288	99.4%	0.862	819	99.1%	98.2%	87.6%	78.0%
Library B with sample 1	27861378	99.5%	0.847	1472	99.7%	99.3%	97.6%	89.9%
Library B with sample 2	25758406	99.1%	0.837	1321	99.8%	99.4%	97.9%	91.3%

Total reads: total number of sequencing read found within the provided sample data.

Mapped reads (%): percentage of total reads that mapped to the human genome.

Specificity: percentage of mapped reads that include the target. The target includes all amplicon sequences with primer sequences excluded.

Mean base coverage: average base coverage within the target. The target includes all amplicon sequences with primer sequences excluded.

C1: % of target that has at least 1x base coverage. Note: non-unique sequencing reads are mapped randomly.

C20: % of target that has at least 20x base coverage.

C100: % of target that has at least 100x base coverage.

Base coverage (0.2x of mean): % of target that has at least 20% of mean base coverage.

Monochromatic Gene Copy Number Assay

Traditional digital PCR methods involve the use of a single labeled probe specific for an individual target. Figure 7 is a schematic depicting one-color detection of a target sequence using droplet based digital PCR. As shown in Panel A of Figure 7, a template DNA is amplified with a forward primer (F1) and a reverse primer (R1). Probe (P1) labeled with a fluorophore of color 1 binds to the target genetic sequence (target 1). Microdroplets are made of diluted solution of template DNA under conditions of limiting or terminal dilution. Droplets containing the target sequence emit fluorescence and are detected by laser (Panels B and C). The number of microcapsules either containing or not containing the target sequence is shown in a histogram (D) and quantified (E).

Figure 8 is a schematic depicting two-color detection of two genetic sequences with a microfluidic device. As shown in Panel A of Figure 8, a template DNA is amplified with two sets of primers: forward primer (F1) and a reverse primer (R1), and forward primer (F2) and a reverse primer (R2). Probe (P1) labeled with a fluorophore of color 1 binds to the target 1 and probe (P2) labeled with a fluorophore of color 2 binds to the target 2 (Panels B and C). Droplets are made of diluted solution of template DNA under conditions of limiting or terminal dilution. Droplets containing the target sequence 1 or 2 emit fluorescence of color 1 or 2 respectively and are optically detected by laser (Panels B and C). The number of microcapsules containing target 1 or 2 is shown by histogram in Panel D.

Methods of the invention involve performing accurate quantitation of multiple different DNA targets by dPCR using probes with the same fluorophore. Figure 9 is a schematic depicting two-color detection of three genetic sequences with a microfluidic device. As shown in Panel A of Figure 9, a template DNA is amplified with three sets of primers: forward primers (F1, F2 and F3) and reverse primers (R1, R2 and R3). Probes (P1, P2 and P3) are labeled with fluorophores (color 1, color 2 and color 1) and bind to the target genetic sequences (target 1, target 2 and target 3) (Panels B and C). Microdroplets are made of diluted solution of template DNA under conditions of limiting or terminal dilution. Microdroplets containing target sequence 1 or 3 emit fluorescence of color 1 at two different intensities; and microdroplets containing target sequence 2 emit fluorescence of color 2. The number of microdroplets containing target 1, 2 or 3 is shown by histogram in Panel D.

Recent results from the droplet digital PCR (dPCR) shows that multiple independent PCR reactions can be run and separately quantified using the same fluorophore. Specifically, an SMN2 assay yields an unexpected population of droplets with slightly elevated signal in the FAM detection channel.

The results are depicted in Figure 10. The left-side dot plot in Figure 10 depicts the effect of having the SMN1 blocker present in the reaction. The four clusters depicted in the left-side dot plot are as follows: the top left cluster includes microdroplets containing the reference sequence (SMARCC1); the bottom left cluster includes microdroplets not containing any sequence; the bottom middle cluster includes microdroplets containing sequence for SMN1; and the bottom right cluster includes microdroplets containing sequence for SMN2. The dot plot on the right-side of Figure 10 depicts four clusters where no SMN1 blocker was present in the reaction: the top left cluster includes microdroplets containing the reference sequence (SMARCC1); the bottom left cluster includes microdroplets not containing any sequence; the bottom middle cluster includes microdroplets containing sequence for SMN1; and the bottom right cluster includes microdroplets containing sequence for SMN2. The shift of the bottom middle cluster in right panel as compared to left panel confirms that fluorescence intensity provides a very sensitive measurement for the presence of a sequence.

Without intending to be bound by any theory, the simplest explanation is that the cluster arises from weak association of the SMN2 probe to the SMN1 gene despite the presence of a blocker to that gene (a nonfluorescent complementary probe to the SMN1 gene).

One definitive confirmation of SMN1 as the source of the unexpected cluster was an observed dependence of the intensity of this feature on the presence of the SMN1 blocker. A clear shift toward higher FAM fluorescent intensities was observed in the absence of the blocker (Figure 10). In another definitive confirmation the ratio of the SMN1 (putative) population size to the reference size of 0.96 in perfect agreement with expectation (two copies of each) (S_131 sample). Another sample, S_122, with the same number of SMN1 copies yielded a ratio of 0.88 in one run and 0.93 in another, also consistent with the proposed explanation of the unexpected cluster.

Without intending to be bound by any theory, these observations indicate that SMN2 probe binding to SMN1 DNA yields an elevated fluorescent signal. A simple kinetic model explaining this phenomenon assumes that the hybridization of the SMN2 probe to the SMN1 DNA achieves equilibrium at a faster rate than the polymerase fills in the complementary strand. The amount of probe fluorophore that is released in each thermal cycle is therefore proportional to (or even equal to) the number of bound probes. Thus the lower the binding affinity the fewer the number of probe fluorophores that are released. Due to SMN1 sequence mismatch(es) with the SMN2 probe, the affinity of the probe is certainly expected to be lower to SMN1 than SMN2. This model also explains the signal dependence on the SMN1 blocker: the blocker competitively inhibits the SMN2 probe hydrolysis by the polymerase exonuclease activity.

It may also be, however, that the probe hybridization does not reach equilibrium before exonuclease activity. In this case, the association rates would play a more dominant role. Similar logic applies. The binding rate to the matching site is likely to be faster than to the mismatch site, and the blocker would act to decelerate probe binding to the mismatch site. The binding of SMN2 probe to SMN1 DNA might be detectable by conventional bulk qPCR, especially in absence of SMN2, but highly quantitative results like those shown here are very unlikely. Definitely, there is no report of qPCR or any other technique quantifying two different DNA sequence motifs with the same color fluorophore. Sequestration of the individual reactions by single molecule amplification within droplets eliminates any confusion regarding mixed contributions to the signal.

The advantage of quantifying DNA with multiple probes of the same color fluorophore extends beyond the example of two highly homologous sequences shown here. Rather, any plurality of sequences of any degree of similarity or dissimilarity can be quantified so long as the different probes have significantly different binding occupancies to their respective DNA binding sites.

Another advantage of the dPCR approach for multiplexed reactions is that the different reactions do not compete with each other for reagents as they would in a bulk qPCR assay. However, the possibility for unintended cross-reactivity remains. A multiplexed assay can require a more dilute sample. For instance, at 10% occupancy a duplex reaction would have double occupancy 1% of the time. Hence 1 in 10 PCR+ droplets would be doubles, resulting in a final intensity at least as high and possibly higher than the brighter of the two probes. For a simple duplex system the contribution from each probe could be recovered. In this example the total number of PCR+ droplets for probe 1 would be (Probe 1) + (Probe1+Probe2). Higher degrees of multiplexing would require greater dilution. For example, for a 4-plex at 1% occupancy the probability of one probe overlapping any of the other 3 is ~3%, and that error may be too high for some applications. The need for large dilutions strongly favors the large number of dPCR reactions.

In another example of the invention, a single fluorophore (FAM) was used in a gene copy number assay for both the reference and the target DNA. A model system was used with varying concentrations of plasmid DNA to represent a change in the target gene copy number, relative to a reference gene, equivalent to 0–16 copies of the target gene per cell. BCKDHA and SMN2 plasmid DNA served as the reference and target with 1× and 0.5× primers and probes respectively. With a starting ratio of 8:1 SMN2 to BCKDHA, the sample was diluted serially by 2× into a solution of BCKDHA at the same concentration to vary just the amount of SMN2. The resultant samples were emulsified, thermally cycled, and over 10^5 droplets were analyzed for each sample as described in the previous section. The process was repeated in triplicate.

Methods of the invention also include analytical techniques for identification of fluorescence signatures unique to each probe. In this example of the invention, histograms of the droplet fluorescence intensities are shown in Fig. 11a for three different template DNA samples: a no template control (dotted line), BCKDHA only (solid line), and 1:1 BCKDHA to SMN2 (dashed line). For clarity, the histograms are shown both overlapped to highlight the similarity

for certain peaks, and offset from each other to reveal all of the features. In the case of 1:1 BCKDHA to SMN2, three populations were readily apparent: a dominant feature appeared at 0.08 V, and two smaller peaks were evident at 0.27 and 0.71 V. The dominant feature at 0.08 V was assigned to PCR(−) droplets since both small peaks disappeared, but the large one remained, in the no template control. The peak at 0.71 V was assigned to BCKDHA since it was the sole feature arising with the addition of just BCKDHA, and the peak at 0.27 V appeared on subsequent addition of SMN2, completing the assignments. A very small peak appeared at ~0.9 V, not visible on the scale of Fig. 11a, that corresponded to droplets occupied by both genes. As another method of the invention, once the different peaks are identified, droplets within each peak were counted corresponding to each possible state (PCR(+)) for either BCKDHA or SMN2, or both, or PCR(−)), and the gene copy number was then determined from the ratio of occupancies. Gene copy numbers for each sample in the serial dilution are plotted in Fig. 11b against expected values (observed ratios of SMN2 to BCKDHA to expected ratios of SMN2 to BSKDHA), with an excellent linear fit ($y = 1.01x$) across the full range ($R^2 = 0.9997$, slope = 1.01), demonstrating accurate and precise measurement of the equivalent of 0 to 16 copies of SMN2 per cell.

Detection of alternatively spliced transcripts

The same principle can be used to detect and count alternatively spliced transcripts. TaqMan assays can be designed that are specific for each of the exons in an RNA transcript. After the RNA is turned into cDNA it can be encapsulated into a droplet at 1 copy or less per droplet. The droplet would also contain the multiplexed TaqMan assay for each of the exons. Each of the TaqMan assays would contain a different probe but all the probes would have the same fluorescent dye attached. The droplets would be thermocycled to generate signal for each of the TaqMan assays. If there are multiple splice variants in the sample they each will contain a different number of exons depending on the splicing events. The fluorescent intensity of each droplet would be different depending on the number of exons present. By counting the number of droplets with different intensities it would be possible to identify the presence and abundance of different splice variants in a sample.

Copy number variants in a heterogeneous sample

It would be possible to determine if a heterogeneous sample contained components with different copy level numbers. If the copy number variants to be assayed were spaced close enough along the chromosome, the DNA from a sample could be fragmented and encapsulated in droplets at a level of one haploid genomic equivalent or less per droplet. The droplet would also contain a TaqMan assay specific for the copy number variant. The intensity of the signal in each droplet would depend on the number of copy number variants are present for the sample. Counting of the number of droplets of different intensities would indicate things like how many cells in a particular sample had what level of copy number variants.

Tuning TaqMan Probe Fluorescence Intensity

Identifying probes by fluorescence intensity often requires adjusting the brightness of the probes, particularly for higher-plex assays with dense probe patterns. In the previous section the probes for the gene copy number assay yielded very well resolved peaks (Fig. 11a). Clearly room exists to accommodate one or multiple extra probes in the copy number assay within the resolution of the measurement, but a method for adjusting the fluorescence intensity of the new probes is required to avoid interference with the existing assay. One method of the invention involves varying the probe and primer concentrations together as a very simple technique to optimize relative intensities in higher-plex reactions.

Figure 12 is a schematic for tuning the intensity of a detectable label to a particular target with a microfluidic device. As shown in Figure 12A, a template DNA is amplified with two sets of primers: forward primers (F1 and F2) and reverse primers (R1 and R2). Probes (P1 and P2) are labeled with fluorophore of color 1 and bind to target 1 and target 2 respectively. Fluorescence from target 2 is lower in intensity than that from target 1 due to single base mismatch between P2 and target 2. As shown in Figure 12B, template DNA is amplified with two sets of primers: forward primers (F1 and F2) and reverse primers (R1 and R2) (Figure 12B). Fluorescence from target 2 is lower in intensity than that from target 1 due to the presence of a competing probe 2 that is not labeled with the fluorophore. As shown in Figure 12, template DNA is amplified with two sets of primers: forward primers (F1 and F2) and reverse primers (R1 and R2). Probes (P1 and P2) are labeled with fluorophore of color 1 and bind to target 1 and

target 2 respectively. Fluorescence from target 2 is lower in intensity than that from target 1 due to the presence of a competing probe 2 that is labeled with a different fluorophore.

Fig. 13 shows probe fluorescence intensities throughout a serial dilution of the probes and primers for a different reference gene, ribonuclease P (RNaseP), against a constant amount of genomic DNA from the Coriell cell line NA3814 at an occupancy of 0.02 target DNA molecules per droplet. The probe fluorescent intensities varied in direct proportion to probe concentration over a narrow concentration range spanning ~0.15 to 0.4 μM ($R^2 = 0.995$)—roughly centered about the typical probe concentration of 0.2 μM —after compensation for dilution errors and other run-to-run differences such as optical realignments using the intensity of the PCR(–) droplets as a reference. In summary, probe intensities can be varied by dilution over a small but adequate range for the purpose of tuning multiplexed assays without affecting the amplification itself.

Although the example above for adjusting probe fluorescence intensities involves varying probe and primer concentrations together by the same factor, the invention is not limited to this method alone for varying probe intensity. Other methods known to those familiar with the art for varying probe intensities are also considered. Such methods include varying just the probe concentration; varying just the primer concentrations; varying just the forward primer concentration; varying just the reverse primer concentration; varying the probe, forward, and reverse primers concentrations in any way; varying the thermal cycling program; varying the PCR master mix; incorporating into the assay some fraction of probes that lack fluorophores; or incorporating into the assay any hybridization-based competitive inhibitors to probe binding, such as blocking oligomer nucleotides, peptide nucleic acids, and locked nucleic acids. The invention incorporates the use of these methods adjusting probe fluorescence intensity, or any other methods for adjusting probe fluorescence intensity, used either by themselves or in any combination.

Higher-plex Reactions

One method of the invention involves performing higher-plex assays with a single probe color (i.e. fluorophore). As described above, probe fluorescent intensities can be adjusted by a variety of means such that each intensity level uniquely identifies a DNA target. For example, targets T1, T2, T3, and T4 might be uniquely identified by intensity levels I1, I2, I3, and I4. Not

intending to be bound by theory, the maximum number of intensity levels possible for unique identification of targets is related to the resolution of the different intensity levels—that is the spread of intensities for each particular probe compared to the separation between the average intensities of the probes—and it is also related to the intensity of the empty droplets that tends to grow with increasing numbers of probes. The number of intensity levels can be 0, or 1, or 2, or 3, or 4, or up to 10, or up to 20, or up to 50, or up to 100. The number of intensity levels can be higher than 100. In the examples show below, as many as three intensity levels are demonstrated.

Another method of the invention involves performing higher-plex assays using multiple different probe colors (i.e. fluorophores). As above for the monochromatic multiplexing assay, for each color probe, multiple targets can be identified based on intensity. Additionally, multiple colors that are spectrally separable can be used simultaneously. For example, a single droplet might contain four different probes for measuring four different targets. Two probes might be of color A with different intensities (say, A1 and A2), and the other two probes of color B with different intensities (say B1 and B2). The corresponding targets are T1, T2, T3, and T4 for A1, A2, B1, and B2 respectively. If a droplet shows an increase in fluoresce in color A, the droplet therefore contained either targets T1 or T2. Then, based on the fluorescence intensity of color A, the target could be identified as T1 or the target could be identified as T2. If, however, a droplet shows an increase in fluorescence in color B, the droplet therefore contained either targets T3 or T4. Then, based on the fluorescence intensity of color B, the target could be identified as T3 or the target could be identified as T4. Not intending to be bound by theory, the maximum number of different colors possible is limited by spectral overlap between fluorescence emission of the different fluorophores. The maximum number of colors can be 1, or 2, or 3, or 4, or up to 10, or up to 20. The maximum number of colors can be higher than 20. In the demonstrations that follow, the largest number of colors is two.

Another method of the invention involves performing higher-plex assays using multiple different probe colors (i.e. fluorophores), however unlike the strategy above where each target is identified by single type of probe with a unique color and intensity, instead in this method a single target may be identified by multiple probes that constitute a unique signature of both colors and intensities. For example, a single droplet might contain four different probes for measuring three different targets (say, T1, T2, and T3). Two probes might be of color A (say,

A1, and A2), and two probes might be of color B (say, B1 and B2). T1 is measured by probe A1, T2 is measured by probe B1, but T3 is measured by both probes A2 and B2. Thus, when a droplet contains T1 only increased fluorescence appears in color A. When a droplet contains T2 only increased fluorescence appears in color B. However when a droplet contains T3, increased fluorescence appears in both colors A and B.

Generally, without wishing to be constrained by theory, the above three methods for higher-plex dPCR are simplest to implement under conditions of terminal dilution, that is when the probability of multiple different target molecules co-occupying the same droplet is very low compared to the probability of any single target occupying a droplet. With multiple occupancy arises the complexity of simultaneous assays competing within the same reaction droplet, and also complexity of assigning the resulting fluorescence intensity that involves a combination of fluorescence from two different reaction products that may or may not be equal to the sum of the two fluorescence intensities of the individual reaction products. However, methods of the invention can accommodate these complications arising from multiple occupancy.

Methods of the invention for higher-plex reactions also include methods for primer and probe pairing. In the simplest case targets are unlikely to reside on the same DNA fragments, such as when targets are from different cells; or when targets are from different chromosomes within a single cell type; or when targets are distant from each other within a single chromosome such that they become physically separated during DNA fragmentation; or when targets are very close to each other within a chromosome, but nevertheless become separated by targeted cleavage of the DNA, such as by restriction enzyme digestion; or for any other reason. In such cases each probe can be paired with a single set of primers (forward and reverse). However, in other cases the target regions might frequently reside on the same DNA fragments, for example when targets reside within the same codon, or for any other reason. In such cases, a single set of primers might serve for multiple probes (for an example, see Pekin *et al.*).

Higher multiplex reactions can be performed to distinguish the haplotypes of two SNPs. For example, assume that at position one there can be genotypes A or A' and at position two there can be genotypes B or B'. In a diploid genome four unique haplotypes are possible (A,B; A,B'; A',B; and A',B'). If for example A' and B' represent drug resistant mutations for infection, it is often the case that A'B and AB' are less severe and treated differently than A'B' which represents a significant drug resistance that must be treated with extreme care. Digital

PCR with intensity discrimination is ideally suited for identifying low prevalence of A'B' in a background of mixtures of the other three haplotypes. Haplotyping information is also important for construction of haplotypes in HLA. One way that the present example can be constructed is by assay design such that color one is used for A and is of high or low intensity indicative of allele A or A' respectively and color two is used for B and is of high or low intensity respectively indicative of B or B'. Populations of [color1,color2] corresponding to [Low, Low] would be a measure of an allele of AB and [high, low] allele A'B and an allele of [A'B'] will be readily distinguishable as [high, high] even in a background that is predominately a mixture of A'B and AB'. See Fig. 22. In some cases it will be advantageous to start by encapsulating into the droplets long single molecules of nucleic acid that contain both A and B SNP location and in other cases it will be desirable to start by encapsulating single cells, bacteria or other organism within the droplets prior to releasing the nucleic acid from the organism. In still other embodiments the multiplex intensity detection of multiple simultaneous targets can be used as surrogate markers for multiple types of binding interactions or labeling of target materials. This technique is also not limited to single molecule detection and can be used for haplotype detection in single cells (e.g., bacteria, somatic cells, etc.). In single cell analysis, a sorting step may be applied prior to haplotyping.

5-plex Assay for Spinal Muscular Atrophy

An aspect of the invention was reduced to practice in an example demonstration of the quantitation of several genetic markers for spinal muscular atrophy (SMA). SMA was selected for one of the example demonstrations due to both its important clinical significance as well as its complicated genetics. It is the second-most prevalent fatal neurodegenerative disease and affects ~1 in 10,000 live births. SMA is most often caused by homozygous absence of exon 7 within the survival of motor neuron 1 gene (SMN1, reviewed by Wirth *et al.*), however the severity of the condition is modulated by the number of gene copies of SMN2 with prognosis ranging from lethal to asymptomatic over 1–5 copy numbers (reviewed by Elsheikh *et al.*). Hence accurate quantitation of SMN2 copy number is important for clinical prognosis and genetic counseling. Aside from large deletions of SMN1, a number of single point mutations or short deletions / duplications within the same gene also account for ~4% of cases of SMA. In a significant step toward a comprehensive SMA assay, the multiplexed dPCR assay demonstrated

here contains both copy number assays (for SMN1 & 2) and an assay for one of the prevalent SNPs (c.815A>G).

One embodiment of the invention is a 5-plex assay for SMA diagnostics. The 5-plex assay quantifies common genetic variants impacting SMA including two copy number assays for the SMN1 and SMN2 genes with BCKDHA as a reference, and a SNP assay for the c.815A>G mutation. Two differently colored fluorophores, FAM and VIC, were used to uniquely identify each of the assays. The probes for SMN1 and SMN2 contained only FAM, and for c.815A only VIC. However, mixtures of VIC and FAM-labeled probes were used for BCKDHA and c.815G. The use of VIC and FAM fluorophores in this example does not limit the invention, rather the 5-plex assay can be used with any spectrally separable fluorophores compatible with the TaqMan assay, or any other fluorogenic hybridization-based probe chemistries. For validating the assay, a model chromosome was synthesized containing a single target region for each of the different primer/probe pairs. EcoRV restriction sites flanked each target, allowing separation of the fragments.

As another method of the invention, histogram-based data presentation and analysis is incorporated into the invention for identifying and characterizing statistically similar populations of droplets that arise from one probe signature (color and intensity), and for discriminating one population of droplets from the others. Fig. 14A shows a 2-dimensional histogram of droplet fluorescence intensities as a contoured heat map, with hotter colors representing higher occurrences. Standard techniques were used to compensate for spectral overlap of the FAM and VIC signals. Samples were run at 0.006 occupancy per target. Six populations were clearly evident, five for the assay and one for PCR(–) droplets. As one method of the invention, the populations were assigned by selective exclusion of assay components. For example, excluding the SMN2 primers and probe eliminated the population at the bottom right in the histogram, but otherwise the distribution remained unchanged. Assignments are labeled in Fig. 14A. As we have found to be generally true for this method of multiplexing, the assay worked immediately with well resolved or at least distinguishable populations for each target. As another method of the invention, the relative positions of the different populations in the histogram were then adjusted into a regularly spaced rectangular array by tuning the probe concentration as described in the previous section. Usually no more than two iterations are required for optimization.

In another method of the invention, the different populations were sufficiently well resolved to allow droplets within each population to be counted by integration across rectangular boundaries. The boundaries were positioned at mid-sections between neighboring peaks. The methods of the invention are not constrained to rectangular boundaries, or to specific boundary locations between peaks. Rather, any closed or unclosed boundary condition can suffice. Boundary conditions do not need to be “binary” either, in the sense that weighted integrations can also be performed across the boundaries to arrive at droplet counts. The peak position of each cluster varied by no more than 2% from run to run after normalization to the intensity of the empty droplets to account for variations in detection efficiency (data not shown). Hence, once identified, the same boundaries for integration could be reused between samples. The methods of the invention are not limited to fixed boundary positions. Dynamic population identification and boundary selection in between samples or studies is anticipated. Twenty different patient samples from the Coriell cell repositories were analyzed with this assay: 4 afflicted with SMA, 1 SMA carrier, and 15 negative controls. Assay results are shown in Fig. 14B. Gene copy number was calculated as before, as the ratio of occupancies derived from the number of target droplets vs. reference droplets. Like the copy number measurement in Fig. 11, each assay yielded ratios very close to the expected integer values, but when all of the patient data was plotted as actual ratio vs. expected integer ratio a small systematic deviation from the ideal slope of 1 was observed. Measured slopes were 0.92, 0.92, and 0.99 for SMN1, SMN2, and c.815A respectively. For clarity, the data in Fig. 14B was scaled to the ideal slope of 1.

The measured genotypes of the different patients were consistent with their disease conditions (unafflicted, carrier, or afflicted). The patients afflicted with SMA each had zero copies of SMN1 (numbers SMA 1–4 in Fig. 14B), the carrier had just one copy, and the negative controls all had two or three copies (numbers 1-15). Three unrelated individuals (numbers 6, 8, and 9) had three copies of SMN1, occurring at a rate of 20% which is similar to a previous report for healthy individuals. Variability in SMN1 copy number is not surprising since it lies within an unstable region of chromosome 5q13. A larger variety of SMN2 copy numbers was observed. One to two copies were most common in the control group, although one individual had zero copies, a distribution consistent with expectations for normal individuals. The SMA carrier and afflicted patients had elevated copy numbers of SMN2 on average: 5 for the carrier, two afflicted with 3 copies, and the others with 2 copies. The afflicted patients were all diagnosed as

SMA Type I, the most severe form, based on clinical observations according to the Coriell repository. The strong genotype/phenotype correlation between SMN2 copy number and disease severity suggests that the two individuals with three copies of SMN2 might have an improved Type II prognosis, especially for the patient SMA 1 who had survived to three years at the time of sampling, much beyond the typical maximum life expectancy for SMA Type I of 2 years. However there remains reluctance to predict disease outcome based on SMN2 copies alone since other less well characterized or unknown modifying genes may impact prognosis and because not all SMN2 copies may be complete genes. Furthermore some Type I patients have begun surviving longer in newer clinical settings. Hence, with little clinical information regarding the patients available to us, we can conclude that our SMN2 assay results were consistent with broad expectations for disease severity.

The SNP assay revealed that all patients carried the normal c.815A genotype and no instances of c.815G were observed. The mutation is relatively rare and hence was not expected to appear in a small patient panel. Of interest, however, was the presence of an apparent extra gene fragment in two unrelated individuals that was uncovered with the SNP assay. The c.815A>G assay does not discriminate between SMN1 and SMN2 due to their high sequence similarity, and hence the total copies of c.815A and G should equal the sum of the copies of SMN1 and SMN2. This was true for all patients except for healthy patients number 1 and 2, both of whom had one extra copy of c.815A. c.815 lies on exon 6, and the SNP that discriminates between the SMN1 and SMN2 genes lies on exon 7, hence the extra genes may be fragments of SMN1 lacking exon 7. This seems reasonable because the deletion of exon 7 is the common mutation causing 95% of cases of SMA (reviewed by Wirth *et al.*) and it is carried by 1/40 to 1/60 adults. Thus these patients might have been typical carriers of SMA but for the acquisition of at least one compensating healthy copy of SMN1 on the same chromosome.

9-Plex Assay for Spinal Muscular Atrophy

A 9-plex assay for certain SMA related targets was also demonstrated with just two colors (probes containing FAM and VIC fluorophores). Aside from the optimized primer and probe concentrations, assay conditions and experimental procedures were identical to the 5-plex assay above. Figure 15a shows the various droplet populations in 2-D histograms before optimization of probe concentrations. The identity of the different targets is shown on the figure

itself. As one method of the invention, the identification of the different populations was made as before, by selective exclusion and/or addition of one or more assays. Most of the populations were already well resolved, with the exception of the probe for the c.815A genotype that was in close proximity with the cluster corresponding to empty droplets. After three iterations of optimization of probe concentrations, all of the target populations were well resolved from each other, and well resolved from the empty droplets (Figure 15B). Three methods of the invention were highlighted in this demonstration: (1) nine DNA targets were uniquely identified in a two-dimensional histogram, far beyond the capabilities of conventional qPCR; (2) target DNA molecules were distinguished on the basis of some combination of both color and intensity arising from one or multiple probes against the same target; and (3) the relative positions of the target molecules within the histogram were adjusted by varying the probe concentrations to optimize the pattern of colors and intensities for increased resolution amongst the various droplet populations.

As one method of the invention, different droplet populations were identified by selective addition or exclusion of assays in the examples above. However the invention is not limited to this method alone. Rather, any method for population assignments known to those in the art are considered. Methods of the invention include any method that can cause an identifiable displacement, appearance, or disappearance of one or more populations within the histograms including changing the probe and primer concentrations together, either by the same factor or by different factors; changing the probe concentration alone; changing the primer concentrations alone; changing the thermal cycling conditions; and changing the master mix composition. Another method of the invention takes advantage of prior knowledge of the position of an assay within a histogram to assist assignment.

Multiplexing Capacity

The level of multiplexing demonstrated in the preceding SMA example was 9x, significantly exceeding the maximum practicable number with qPCR. Without wishing to be constrained by theory, the two main limitations are the resolution between assays and the increasing fluorescence intensity of empty droplets with higher loading of probes. A method of the invention involves optimizing the pattern of colors and intensities of the different probes for maximum multiplexing while still achieving adequate specificity for each individual reaction.

Although rectangular arrays of droplet populations were demonstrated for the 5- and 9-plex reactions, another desirable pattern is the tight-packed hexagonal array. However the invention is not constrained to any particular array strategy.

Adding extra colors would increase the capability even further, however with some diminishing returns because the fluorescence of the empty droplets would continue to rise. The capacity could be yet further increased with better probes yielding larger differential signals, such as hybrid 5'-nuclease / molecular beacon probes that reduce background by contact quenching yet exhibit the bright signals typical of free unquenched fluorophores. With such improvements multiplexing capacity exceeding 50× can be envisioned.

Combined Multiplexing with Optical Labeling

Using droplet-based microfluidics, multiple targets can also be measured simultaneously by a different method. According to the alternative method, primers and probes can be loaded individually into droplets along with an optical label to uniquely identify the assay. Typically the optical label is a fluorophore, or a combination of different fluorophores, that are spectrally distinct from the probe fluorophore. Various different types of droplets, each containing different assays that are uniquely identified by different optical labels, can be mixed into a “library” of droplets. Then, according to methods of the invention above, library droplets are merged one-to-one with droplets containing template DNA. After thermal cycling, some droplets that contain template DNA will exhibit brighter fluorescence at the emission wavelengths of the probes. The specific target DNA molecules giving rise to these PCR(+) signals are subsequently identified by the optical probes. In one study, the six common mutations in *KRAS* codon 12 were screened in parallel in a single experiment by one-to-one fusion of droplets containing genomic DNA with any one of seven different types of droplets (a seven-member library), each containing a TaqMan probe specific for a different *KRAS* mutation, or wild-type *KRAS*, and an optical code.

In one method of the invention, optical labeling can be combined with the various methods for multiplexing dPCR already incorporated into this invention. For example, a single optical label might code for the entire 5-plex SMA assay, above, instead of just a single assay as in the *KRAS* example above. In this manner, other optical labels might code for different screening assays for newborn infants. According to other methods of the invention, above, a

single DNA sample from an infant could then be analyzed with all of the assays simultaneously by merging droplets containing the DNA one-to-one with library droplets containing the optically encoded assays.

As an example of combining multiplexing with optical labels, a so called 3x3x3 combination multiplex reaction with optical labeling was demonstrated (3x3 optical labeling with two fluorophores, each encoding a triplex assay, for a total of 27-plex). Two fluorophores were employed for optical labeling, Alexa633 and CF680 (excited by a 640 nm laser), with three intensity levels each producing nine total optical labels. As before with the 5- and 9-plex assays for SMA, TaqMan assays were used with FAM and VIC fluorophores (excited by a 488 nm laser). The fluorescence from the FAM and VIC fluorophores were recorded simultaneously with the fluorescence from the optical labels, requiring modifications to the optical layout of the instrumentation described for the SMA assay (the optical schematic for two-laser excitation and 4-color detection is shown in entirety in Figure 16). Also, co-flow microfluidics were used in this example (the use of co-flow based microfluidics for this application is one of the methods of the invention described above). In this case, the template DNA was introduced into the chip in one flow, and the PCR master mix, the primers and probes for one triplex assay, and the unique composition of fluorophores for the optical label were introduced into the chip in another flow simultaneously. The two flow streams converged in a fluidic intersection upstream from the droplet forming module, and thus each droplet formed contained the contents of both flow streams. Methods to implement co-flow microfluidics are well known to those in the art. The droplets were collected, and then the procedure was repeated with the next triplex assay and optical label. The procedure was repeated a total of nine times, once for each pair of assays and optical labels. All of the droplets were collected into a single PCR tube and thermally cycled off chip. The mixture of thermally cycled droplets was reinjected into the same read-out chip as used for the SMA assay, above, and the fluorescence intensities of the assays from all four fluorophores were recorded.

Figure 17 shows the cumulative results from all droplets in the 3x3x3 assay using co-flow microfluidics. The figure shows two 2-D histograms of droplet fluorescence intensities, the histogram on the left from all of the optical labels, and the histogram on the right from the assays. Standard methods were used to compensate for spectral overlap. The histograms are shown as heat maps, with hotter colors designating larger numbers of droplets. Nine different

clusters of droplets were clearly evident in the histogram of the optical labels, corresponding to each of the nine different optical labels: there is a small group of four clusters at the bottom left corner of the histogram, corresponding to optical labels with the lowest fluorescent intensities; and there are five clusters appearing as linear streaks at the higher intensities. The droplet clusters were less distinct in the histogram for the assay, but this was as expected because the droplets shown contained all of the triplex assays. The individual assays became clearly distinct once a single type of assay was selected by using the optical labels, as follows.

Methods of the invention involve selecting individual populations of droplets all containing the same optical labels, or groups of optical labels. In some methods of the invention, boundaries of fluorescence intensity were used to specify populations. In the example shown here, a rectangular boundary was used specifying the minimum and maximum fluorescence intensities for each fluorophore. However the methods of the invention are not restricted to rectangular boundaries. Any boundary, closed or unclosed, can be employed. Furthermore, according to methods of the invention, selections of droplet populations can be made by any method, and is not restricted to threshold-based methods such as boundary selection.

Figure 18A shows the droplet fluorescence intensities for the assay (right histogram) when only one optical label was selected (left histogram). The lines overlaid on the histogram of the optical labels identify the rectangular boundary used to select just the optical label with the lowest fluorescence for both fluorophores. Both histograms showed only the droplets that were selected. After selection, four distinct clusters of droplets appeared in the assay histogram, three for the different assays (in this case, assays for SMN1, SMN2, and TERT, where TERT is another common reference gene) and one for the empty droplets. The copy numbers for SMN1 and SMN2 were measured by the same methods of the invention as described above for the 5-plex SMA assay, with values of 1.8 and 0.94 close to the expected values of 2 and 1, respectively. The same assay was encoded with two other optical labels, and their selections are shown in Figures 18B and C. Similar results were achieved, with an overall measurement of 1.9 ± 0.1 and 0.9 ± 0.1 copies of SMN1 and SMN2 respectively, showing the measurement to be accurate within experimental uncertainty.

Figures 19A, B, and C show optical label selections for a different assay (TERT, c.5C in the SMN1 gene, and BCKDHA (labeled E1a in the figure)). In each case four distinct clusters also appeared, and by the same methods of the invention above, accurate measurements of gene

copy number were made for c.5C and BCKDHA, referenced to TERT, of 2.9 ± 0.1 and 2.0 ± 0.2 compared to 3 and 2, respectively. Figures 20A, B, and C show optical label selections for a third assay (TERT, c.88G in the SMN1 gene, and RNaseP, where RNaseP is a common reference gene). Accurate gene copy numbers of 2.1 ± 0.1 were measured for both c.88G and RNaseP, referenced to TERT, compared to the expected value of 2.

In summary, the demonstration here shows use of nine different optical labels to enable independent measurement of three triplex assays in a single experiment. Although some of the optical labels encoded for redundant assays in this example (there were only three different assays despite having nine optical labels), the invention is not constrained to any particular formatting of assays and optical labels. Embodiments of the invention include formats where all of the assays are the same across all of the optical labels; where none of the assays are the same across all of the optical labels; where some of the assays are the same across all of the optical labels; where some of the assays have greater plexity than others across all of the optical labels; where all of the assays have the same plexity across all of the optical labels; and any other arrangements of assays across all of the optical labels are considered.

Although two different fluorophores were used to create the optical labels in this example, the invention is not constrained to any particular number of fluorophores comprising the optical labels. Embodiments of the invention include optical labels comprised of 1 fluorophore, or 2 fluorophores, or 3 fluorophores, or 4 fluorophores, or up to 10 fluorophores, or up to 20 fluorophores. Optical labels can also comprise more than 20 fluorophores.

Although solely triplex assays were used in the example demonstration here, the invention is not constrained to use of triplex assays with optical labels. Embodiments of the invention include plexities of the following amounts when used with optical labels: single plex, duplex, triplex, 4-plex, up to 10-plex, up to 20-plex, up to 50-plex, and up to 100-plex. Embodiments of the invention also include plexities exceeding 100 when used with optical labels.

Another method of the invention involves the use of droplet merging, instead of co-flow, for combining multiplexing with optical labels. A demonstration using droplet merging was performed with the same 3x3x3 assay as in the preceding example with co-flow. The assays (probes and primers) combined with their unique optical labels were first encapsulated into droplets along with the PCR master mix. Subsequently, according to methods of the invention

described above, a library containing a mixture of droplets from all nine optically labeled assays was merged one-to-one with droplets containing template DNA from the same patient as in the preceding example. As another method of the invention, the droplet merge was performed using a lambda-injector style merge module, as described in U.S. Provisional Application, Serial No. 61/441,985, incorporated by reference herein. Aside from the differences between co-flow and merge, the assays and experimental procedures were identical to those above for the co-flow experiment. Figure 21 shows 2-D histograms of droplet fluorescence intensity for the optical labels and the assays that are similar to those in Figures 17–20. As in the case for co-flow, upon selection of droplets containing individual optical labels, the expected distinct clusters of droplets corresponding to each assay were clearly evident. Furthermore for each assay the measured gene copy number matched or very nearly matched the expected values within experimental uncertainty (See Table 1).

<i>Gene or genotype</i>	<i>Measured copy number</i>	<i>Expected copy number</i>
SMN1	1.98 ± 0.09	2
SMN2	0.99 ± 0.04	1
c.5C in SMN1	3.01 ± 0.06	3
c.88G in SMN1	2.15 ± 0.08	2
BCKDHA	2.00 ± 0.05	2
RNaseP	2.11 ± 0.16	2

Table 1 Gene copy number measurements from the 3x3x3 assay.

Although methods of the invention include using either microfluidics with co-flow or droplet merging, the invention is not limited in this regard. Any fluidic method capable of generating optically labeled droplets that also contain fluorogenic DNA hybridization probes are considered. For example, other embodiments well known in the art are mixing optical labels and assays in the macrofluidic environment before injection into a droplet generating chip; and mixing optical labels and assays thoroughly upstream from the droplet forming module in dedicated mixing modules, such as with a serpentine mixer.

Data Analysis

One method of the invention involves histogram-based data presentation and analysis for identifying and characterizing populations of statistically similar droplets that arise from unique probe signatures (color and intensity), and for discriminating one population of droplets from the others. Another method of the invention involves histogram-based data presentation and analysis for identifying and selecting populations of droplets based on unique signatures from optical labels. Examples of one and two-dimensional histograms have been provided for these methods, but the invention is not limited in this regard. As described above, it is anticipated that greater numbers of colors will be used for both multiplexing and for optical labels. Hence, embodiments of the invention include histograms of dimensionality greater than two, such as 3, or 4, or up to 10, or up to 20. Histograms of dimensionality greater than 20 are also incorporated into the invention.

Another method of the invention involves the selection of droplets within histograms, either for counting, or for assay selection as in the use of optical labels, or for any other purpose. Methods of the invention include selections by boundaries, either closed or unclosed, of any possible shape and dimension. Methods of the invention also include selections of droplets that exhibit fluorescence from single types of fluorophores, or from multiple types of fluorophores, such as arising from multiple probes against a common DNA target.

Polymerase Error Correction

For applications requiring very high sensitivity, such as searching for rare mutations amidst an abundance of wild-type DNA, false positive results can arise from errors from the DNA polymerase itself. For example, during one of the early thermal cycles the polymerase might synthesize the mutant strand of DNA from a wild-type template. This type of error is most likely to occur when the difference between the mutant and the wild-type is very small, such as single nucleotide polymorphism (SNP). In this method of the invention, each droplet contains only a single target nucleic acid, if any at all. In the preferred embodiment, this is accomplished under the conditions of terminal dilution. Droplets that contain amplification products that are a wild-type of the target are detected based on emission from the fluorophore that is released from the probe that hybridizes to the wild-type of the target. Droplets that contain the variant of the target are detected based on emission from the fluorophore that is released from the probe that hybridizes to the variant of the target. Since each droplet starts with only a single nucleic acid

molecule, the resultant amplification products in each droplet are either homogeneous for the target or homogenous for the variant of the target.

However, certain droplets will contain a heterogeneous mixture of both target and target variant due to polymerase errors during the PCR reaction. Error rates in PCR vary according to the precise nucleic acid sequence, the thermostable enzyme used, and the in vitro conditions of DNA synthesis. For example, the error frequency (mutations per nucleotide per cycle) during PCR catalyzed by the thermostable *Thermus aquaticus* (*Taq*) DNA polymerase vary more than 10-fold, from 2×10^{-4} to $<1 \times 10^{-5}$. Eckert et al. (Genome Res. 1:17-24, 1991), the content of which is incorporated by reference herein in its entirety. Polymerase-mediated errors at a frequency of 1 mutation per 10,000 nucleotides per cycle are an important consideration for any PCR application that begins with a small amount of starting material (e.g., less than a total of 10,000 nucleotides of target DNA) or that focuses on individual DNA molecules in the final PCR population.

The proportion of DNA molecules that contain sequence changes is a function of the error rate per nucleotide per cycle, the number of amplification cycles and the starting population size. The population of altered DNA molecules arises during PCR from two sources: (1) new errors at each PCR cycle; and (2) amplification of DNA molecules containing errors from previous cycles. The formula $f = np/2$ describes the average mutation frequency (f) for PCR amplification as a function of the polymerase error rate per nucleotide per cycle (p) and the number of cycles (n), assuming that p is constant at each cycle. Due to the exponential nature of PCR, the occurrence of an early error can increase the final error frequency above the average described by $f = np/2$, because the variant DNA molecule will be amplified with each cycle, resulting in populations with a larger than average number of variants.

A polymerase error that converts a wild-type of the target to a variant of the target during an early round of amplification results in a heterogeneous population of target and target variant in a droplet, and may lead to a droplet being incorrectly identified as containing a variant of the target, i.e., a false positive. Such false positives greatly impact the validity and precision of digital PCR results.

Methods of the invention are able to detect which droplets contain a heterogeneous population of molecules and are able to exclude those droplets from analysis. As droplets containing amplified product flow in a channel through the detector module, the module is able

to detect the fluorescent emission in each droplet. Droplets that produce only a single signal are classified as droplets that contain a homogeneous population of target. Since probes that hybridize to the wild-type of the target have a different fluorophore attached than probes that hybridize to a variant of the wild-type of the target, methods of the invention can classify each droplet as containing either a homogeneous population of amplicons of the target or a homogeneous population of amplicons of the variant of the target.

Droplets that produce two signals are classified as droplets that contain a heterogeneous population of molecules. Since each droplet started with at most a single target nucleic acid, a droplet that includes amplification products that are both amplicons of the target and amplicons of a variant of the target are droplets in which the variant of the target was produced by a polymerase error during the PCR reaction, most likely a polymerase error during an early cycle of the PCR reaction. Such droplets are detected and excluded from analysis.

Polarization Detection

As discussed, droplets containing amplified products may be detected by detecting the fluorescent emission of the detectable probe (such as a TaqMan probe) coupled to a target. TaqMan probes consist of a fluorescent molecule covalently attached to the 5'-end of the oligonucleotide probe and a quencher at the 3'-end. Several different fluorophores (e.g. 6-carboxyfluorescein, acronym: *FAM*, or tetrachlorofluorescein, acronym: *TET*) and quenchers (e.g. tetramethylrhodamine, acronym: *TAMRA*, or dihydrocyclopyrroloindole tripeptide minor groove binder, acronym: *MGB*) are available and suitable for use in methods of the invention. The quencher molecule quenches the fluorescence emitted by the fluorescent molecule when excited by, e.g., a PCR cycler light source via FRET (Fluorescence Resonance Energy Transfer). As long as the fluorescent molecule and the quencher are in proximity, quenching inhibits fluorescence signals. While the fluorescence signal is inhibited, intact probes still emit a background or dim fluorescent signal.

TaqMan probes are designed such that they anneal within a DNA region amplified by a specific set of primers. As the Taq polymerase extends the primer and synthesizes the nascent strand, the 5' to 3' exonuclease activity of the polymerase degrades the probe that has annealed to the template. Degradation of the probe releases the fluorescent molecule from it and breaks the close proximity to the quencher, thus relieving the quenching effect and allowing fluorescence of

the fluorescent molecule. Hence, fluorescence detected in the real-time PCR thermal cycler is directly proportional to the fluorescent molecule released (cleaved) and the amount of DNA template present in the PCR.

However, for certain applications that include a large volume of probes within a droplet, the fluorescence of the released (cleaved) fluorescent molecule of the amplified product is hard to detect amid the dim background fluorescence of the large volume of intact probes. A low bright to dim ratio decreases the ability to successfully detect amplifiable product within a droplet by merely detecting the optical fluorescence of a droplet.

As an alternative to optical fluorescence detection, methods of the invention provide for detecting amplifiable product by analyzing the fluorescence polarization of fluorescent molecule (e.g., dyes and fluorophores) within a droplet. Using fluorescence polarization, a dye molecule attached to a probe can be distinguished from a dye molecule that is free floating in the droplet after cleavage. Prior to cleavage, a dye molecule intact with the quencher portion of a probe has a high degree of fluorescence polarization whereas a cleaved dye molecule has a low fluorescence polarization value.

Fluorescence polarization is based on the principle that, when temperature and viscosity of a solvent is held constant, the degree of fluorescence polarization detected when a fluorescent dye molecule is excited by polarized light depends on the molecular weight of the dye molecule. By monitoring the fluorescence polarization, one can detect significant changes in molecular weight of a dye molecule, which allows one to detect whether the dye molecule remains intact with the probe or free-floating due to cleavage. Detection of a free-floating dye molecule signifies amplifiable product, and enables detection of a corresponding target.

Light emitted from a fluorescent molecule becomes polarized when it is excited by polarized light at a certain wavelength. When the fluorescent molecule is in a medium, the molecule rotates. This rotation causes the observed fluorescence polarization to be proportional to the molecules rotational relaxation time. Rotational relaxation time correlates to the viscosity of the solvent, absolute temperature, molecular volume, and gas content, which is why a degree of fluorescence polarization is directly proportional to molecular weight when temperature and viscosity of a solvent is held constant. If a fluorescent molecule is large and has a high molecular weight (such as an intact probe having a fluorescent molecule coupled to a quencher probe), it slowly rotates within the solvent so that polarization of the fluorescent molecule is

preserved. However, if a fluorescent molecule is small and has a low molecular weight (such a fluorescent molecule cleaved from a probe during amplification), it rotates quickly and loses polarity or becomes depolarized. Based on the difference in the degree of fluorescent polarization between cleaved and intact probes, fluorescent polarization can be used to detect amplifiable products in any assay that releases a fluorescent molecule during amplification of a target nucleic acid, e.g. standard 5' nuclease assay for determining mutational status in a nucleic acid. It is understood that several factors may be adjusted to increase the sensitivity of this technique. For example, the excitation state of the dye, the size of the initial probe, the size of the cleaved fluorescent molecule, and the fluorescence/polarization lifetime of the fluorescent molecule after excitation. Detecting nucleic acids with fluorescence polarization is described in more detail in Latif et al., *Genome Re.* 2001 11:436-440, the contents of which is incorporated by reference in its entirety.

Another consideration effecting polarization of a fluorescent molecule, either quenched or free-floating, is the fluorescence lifetime of the molecule. The fluorescence polarization measurements are based on an inverse relationship between the fluorescence lifetime of a fluorescent molecule and the molecular rotation of the molecule by itself and/or other molecules bound thereto (e.g. probe) attached thereto. If the fluorescence lifetime of a fluorescent molecule is greater than the rotation correlation of time the molecule(s) bound thereto, the molecule(s) randomize during emission and quickly become unpolarized. If the fluorescence lifetime of a fluorescent molecule is lesser than the rotational correlation of time to molecules bound thereto, the molecules remain aligned during emission and the emission remains polarized. As such, certain embodiments provide for optimizing use of the fluorescence lifetime of a molecule in order to optimize polarization detection. This includes selecting a fluorescent molecule and designing probes (e.g. quencher probe length or adding a weight-bearing molecule) to maximize polarization differences between quenched and free-floating fluorescent molecules based on the fluorescence lifetime of the molecules.

In certain embodiments, any moiety that changes the rotational velocity can be added to the probe. For, example, a weight-bearing molecule is attached to a quencher portion of a probe to increase the combined molecular weight of an intact probe. The higher molecular weight of an intact probe (i.e. quenched probe), the greater disparity between the polarizations of a fluorescent molecule still attached to the probe and a cleaved fluorescent molecule. This is

because the added weight restricts motion of the intact probe, which better preserves the polarization of the fluorescent molecule. Examples of weight-bearing molecules suitable for increasing weight of an intact probe for motion reduction include, for example, attaching a random sequence to the '3 prime end of the intact probe or attaching a protein, bead, or other nanoparticle to the probe (via a random sequence, which allows access to template for PCR).

For detection of targets via fluorescent polarization in droplets, a laser is used to emit polarized light into a droplet containing a single template nucleic acid and subjected to an amplification process (such as PCR amplification). The laser will polarize fluorescent molecules of intact probes and free-floating fluorescent molecules that were cleaved during amplification. The wavelength of the polarized light used to excite the fluorescent molecules depends on the type of the fluorescent molecules utilized. Lasers suitable for polarizing the fluorescent molecules include, for example, lasers that emit any type of polarized light of certain wavelengths (such as a ring-laser). The emitted polarized light can include, for example, linear, elliptical, plane, circular polarized light. One or more detectors, such as a photomultiplier tubes, are used to detect the polarization of the intact and free-floating fluorescent molecules. Each detector is operably coupled to a polarizer for detection of the polarized light. A polarizer is an optical filter that passes light of a specific polarization and blocks waves of other polarizations. It can convert a beam of light of undefined or mixed polarization into a beam with well-defined polarization. The common types of polarizers are linear polarizers and circular polarizers. A polarizer may be chosen to correspond to the type of polarized light emitted from the laser.

Polarizers may be oriented parallel or perpendicular to the emitted polarized light (e.g. turning the polarizing filter so that its grid is perpendicular to a plane of emitted polarized light. In particular embodiments, a polarizing filter is placed cross (i.e. perpendicular) to the polarization excitation of the light source. The crossed polarizing filter blocks substantially more of the polarized light emitted from the quenched probes (having the fluorescent molecule intact) because the quenched probes rotate less and therefore maintain substantially the same polarization excitation of the light source. In contrast, the cross polarizing filter receives (i.e. allows light to pass through to the detector) emitted light from free-floating fluorescence molecules, which were cleaved from the probe) because those molecules are rotating more erratically and are more likely to emit polarized light that deviates from the polarization excitation of the light source. As a result, the crossed polarizing filter provides a better

brightness-to-dimness ratio between the free-floating cleaved fluorescent molecules and the intact quenched probes. Therefore, signals pertinent to the detection of the target (cleaved signals) are more readily apparent.

The configuration of the detectors also increases the specificity for detecting polarized light emitted from intact and free-floating fluorescent molecules. Figure 30A depicts detectors lined up adjacently along the channel. With the detectors spatially-arranged in an adjacent configuration, the detectors can sense changes in polarization emitted from an excited droplet over a period of time as the droplet moves through the channel past the detectors. Figure 30B illustrates a configuration where more than one detector is collecting polarized light emitted from a droplet at the same time. This increases the amount of signals being emitted by the fluorescent molecules that are detected at a certain time point, which leads to increased specificity of polarized light/signal emitted from the fluorescent molecules. In addition, the polarizers for the detectors may be the same or different. For example, one detector may have a polarizer crossed with the polarization excitation of the laser and another detector may have a parallel with the polarization excitation of the laser. This also increases the amount of data received.

Figure 31 depicts another configuration for detecting fluorescence polarization. As shown in Figure 31, a laser transmits polarized light to a splitter A. Splitter A directs polarized light from the laser to the channel for energizing the droplet. Polarized light emitted from one or more fluorescent molecules in the droplet is then transmitted through splitter A to splitter B. Splitter B splits light into detector 1 coupled to polarizer 1 and detector 2 coupled to polarizer 2. Polarizers 1 and 2 may be the same or different.

Figure 32 illustrates yet another configuration for detecting fluorescence polarization. As shown in Figure 32. As shown in Figure 31, a laser transmits polarized light to a splitter C. Splitter C directs polarized light from the laser to the channel for energizing the droplet. Polarized light (P1 + P2) emitted from one or more fluorescent molecules in the droplet is then transmitted through Splitter A to a polarizing beam splitter. P1 indicates emitted polarized light that is parallel to the polarization excitation of the laser and P2 indicates emitted polarized that is perpendicular to the polarization excitation of the laser. The polarizing beam splitter splits the emitted polarized light into two beams of differing linear polarization beams. The polarizing beam splitter is exemplified in Figure 32A. As shown the polarizing beam splitter splits polarized light (P1 and P2) emitted from the droplet into separate beams. Each beam P1 and P2

is transferred into a detector having two different detection channels. Although, it is understood that two separate detectors may also be utilized. With the polarized beam splitter, polarized light emitted from the droplet having polarization P1 and P2 can be detected.

One or more filters (not shown in Figures 30A-32), such as wavelength specific filters, can be utilized to prevent unwanted light from entering the detectors. It is also understood that the position of the laser can be adjusted in any one of the configurations depicted in Figures 30A-32.

The detected signals can be compared to reference polarization values of the fluorescence probes when intact (i.e. quenched) or free floating in order to determine whether the fluorescence molecules are intact (negative reaction/amplification) or free-floating (positive reaction/amplification).

In the case of two or more probes having fluorescence molecules for different targets, semi-automated genotype assignments may be made by taking a ratio between normalized fluorescence polarization values of two or more different fluorescent molecules (e.g. DYE1 and DYE2). For example, a sample with a low DYE1/DYE2 ratio represents a homozygous allele 1 sample (cleaved DYE1 probe, intact DYE2 probe); a high DYE1/DYE2 ratio represents a homozygous allele 2 sample (intact DYE1 probe, cleaved DYE2 probe); a DYE1/DYE2 ratio close to one represents a heterozygote if the absolute FP values of both dyes are low (both probes are cleaved); and indicating failed PCR when ratio is close to one but both dye values are high (both probes are intact).

The fluorescence polarization detection technique is not limited to TaqMan probes as discussed above, but may be adapted for use with any other fluorogenic DNA hybridization probes, such as molecular beacons, Solaris probes, scorpion probes, and any other probes that function by sequence specific recognition of target DNA by hybridization and result in increased fluorescence on amplification of the target sequence.

The following example illustrates a preferred method of detecting polarized light emitted. Two reaction mixes were prepared: Group #1. 0.4uM quenched (intact) SMN88G_FAM probe in 1x Genotyping mix and 0.5% tetronic; Group #2. 0.2uM free-floating 6FAM in 1x Genotyping mix and 0.5% tetronic. The reactions were mixed and diluted into eight separate samples. Intact Group #1 and free-floating Group #2 were excited with a laser beam emitting plane polarized light. A first Detector and a second Detector were used to detect light emitted from the excited

FAM molecules in both Group #1 and Group #2. The first Detector included a polarizer arranged perpendicular to the polarization excitation of the laser, and the second Detector included a polarizer arranged parallel to the polarization excitation of the laser. In addition, a third detector without a polarizer was also used to detect signals emitted from Group #1 and Group #2. Polarized light emitted from the excited FAM molecules in both Group #1 and Group #2 were detected. The below Table 4 shows the brightness-to-dimness ratio of Group #1 and Group #2 from the three detectors:

	Quenched	Free	Brightness-to-Dim Ratio
Detector 1: Parallel Polarizing Filter	0.17	0.97	5.7
Detector 2: Perpendicular Polarizing Filter	0.13	0.91	7.0
Detector 3: No Polarizing Filter	0.42	2.57	6.1

Table 4

As shown in Table 4, the brightness-to-dim ratio is improved when the polarizing filter is perpendicular (i.e. crossed) with the polarization excitement of the polarized laser source. A enhanced brightness-to-dim ratio with respect to quenched and free-floating fluorescent molecules allows one to focus on the signals relevant to the amplified product. That is, the detector having a polarizing filter crossed with the polarization excitement of the light source provides better signal return for the cleaved free-floating fluorescent molecules, and reduces signals received from the quenched probes that did not hybridize to the target.

Analysis

Analysis is then performed on the droplets. The analysis may be based on counting, i.e., determining a number of droplets that contain only wild-type target, and determining a number of droplets that contain only a variant of the target. Such methods are well known in the art.

See, e.g., Lapidus et al. (U.S. patent numbers 5,670,325 and 5,928,870) and Shuber et al. (U.S. patent number 6,203,993 and 6,214,558), the content of each of which is incorporated by reference herein in its entirety.

Generally, the presence of droplets containing only variant is indicative of a disease, such as cancer. In certain embodiments, the variant is an allelic variant, such as an insertion, deletion, substitution, translocation, or single nucleotide polymorphism (SNP).

Biomarkers that are associated with cancer are known in the art. Biomarkers associated with development of breast cancer are shown in Erlander et al. (US 7,504,214), Dai et al. (US 7,514,209 and 7,171,311), Baker et al. (US 7,056,674 and US 7,081,340), Erlander et al. (US 2009/0092973). The contents of the patent application and each of these patents are incorporated by reference herein in their entirety. Biomarkers associated with development of cervical cancer are shown in Patel (US 7,300,765), Pardee et al. (US 7,153,700), Kim (US 6,905,844), Roberts et al. (US 6,316,208), Schlegel (US 2008/0113340), Kwok et al. (US 2008/0044828), Fisher et al. (US 2005/0260566), Sastry et al. (US 2005/0048467), Lai (US 2008/0311570) and Van Der Zee et al. (US 2009/0023137). Biomarkers associated with development of vaginal cancer are shown in Giordano (US 5,840,506), Kruk (US 2008/0009005), Hellman et al. (Br J Cancer. 100(8):1303-1314, 2009). Biomarkers associated with development of brain cancers (e.g., glioma, cerebellum, medulloblastoma, astrocytoma, ependymoma, glioblastoma) are shown in D'Andrea (US 2009/0081237), Murphy et al. (US 2006/0269558), Gibson et al. (US 2006/0281089), and Zetter et al. (US 2006/0160762). Biomarkers associated with development of renal cancer are shown in Patel (US 7,300,765), Soyupak et al. (US 7,482,129), Sahin et al. (US 7,527,933), Price et al. (US 7,229,770), Raitano (US 7,507,541), and Becker et al. (US 2007/0292869). Biomarkers associated with development of hepatic cancers (e.g., hepatocellular carcinoma) are shown in Horne et al. (US 6,974,667), Yuan et al. (US 6,897,018), Hanausek-Walaszek et al. (US 5,310,653), and Liew et al. (US 2005/0152908). Biomarkers associated with development of gastric, gastrointestinal, and/or esophageal cancers are shown in Chang et al. (US 7,507,532), Bae et al. (US 7,368,255), Muramatsu et al. (US 7,090,983), Sahin et al. (US 7,527,933), Chow et al. (US 2008/0138806), Waldman et al. (US 2005/0100895), Goldenring (US 2008/0057514), An et al. (US 2007/0259368), Guilford et al. (US 2007/0184439), Wirtz et al. (US 2004/0018525), Filella et al. (Acta Oncol. 33(7):747-751, 1994), Waldman et al. (US 6,767,704), and Lipkin et al. (Cancer Research, 48:235-245, 1988). Biomarkers associated with

development of ovarian cancer are shown in Podust et al. (US 7,510,842), Wang (US 7,348,142), O'Brien et al. (US 7,291,462, 6,942,978, 6,316,213, 6,294,344, and 6,268,165), Ganetta (US 7,078,180), Malinowski et al. (US 2009/0087849), Beyer et al. (US 2009/0081685), Fischer et al. (US 2009/0075307), Mansfield et al. (US 2009/0004687), Livingston et al. (US 2008/0286199), Farias-Eisner et al. (US 2008/0038754), Ahmed et al. (US 2007/0053896), Giordano (US 5,840,506), and Tchagang et al. (Mol Cancer Ther, 7:27-37, 2008). Biomarkers associated with development of head-and-neck and thyroid cancers are shown in Sidransky et al. (US 7,378,233), Skolnick et al. (US 5,989,815), Budiman et al. (US 2009/0075265), Hasina et al. (Cancer Research, 63:555-559, 2003), Kebebew et al. (US 2008/0280302), and Ralhan (Mol Cell Proteomics, 7(6):1162-1173, 2008). The contents of each of the articles, patents, and patent applications are incorporated by reference herein in their entirety. Biomarkers associated with development of colorectal cancers are shown in Raitano et al. (US 7,507,541), Reinhard et al. (US 7,501,244), Waldman et al. (US 7,479,376); Schleyer et al. (US 7,198,899); Reed (US 7,163,801), Robbins et al. (US 7,022,472), Mack et al. (US 6,682,890), Tabiti et al. (US 5,888,746), Budiman et al. (US 2009/0098542), Karl (US 2009/0075311), Arjol et al. (US 2008/0286801), Lee et al. (US 2008/0206756), Mori et al. (US 2008/0081333), Wang et al. (US 2008/0058432), Belacel et al. (US 2008/0050723), Stedronsky et al. (US 2008/0020940), An et al. (US 2006/0234254), Eveleigh et al. (US 2004/0146921), and Yeatman et al. (US 2006/0195269). Biomarkers associated with development of prostate cancer are shown in Sidransky (US 7,524,633), Platica (US 7,510,707), Salceda et al. (US 7,432,064 and US 7,364,862), Siegler et al. (US 7,361,474), Wang (US 7,348,142), Ali et al. (US 7,326,529), Price et al. (US 7,229,770), O'Brien et al. (US 7,291,462), Golub et al. (US 6,949,342), Ogden et al. (US 6,841,350), An et al. (US 6,171,796), Bergan et al. (US 2009/0124569), Bhowmick (US 2009/0017463), Srivastava et al. (US 2008/0269157), Chinnaiyan et al. (US 2008/0222741), Thaxton et al. (US 2008/0181850), Dahary et al. (US 2008/0014590), Diamandis et al. (US 2006/0269971), Rubin et al. (US 2006/0234259), Einstein et al. (US 2006/0115821), Paris et al. (US 2006/0110759), Condon-Cardo (US 2004/0053247), and Ritchie et al. (US 2009/0127454). Biomarkers associated with development of pancreatic cancer are shown in Sahin et al. (US 7,527,933), Rataino et al. (US 7,507,541), Schleyer et al. (US 7,476,506), Domon et al. (US 7,473,531), McCaffey et al. (US 7,358,231), Price et al. (US 7,229,770), Chan et al. (US 2005/0095611), Mitchl et al. (US 2006/0258841), and Faca et al. (PLoS Med 5(6):e123, 2008).

Biomarkers associated with development of lung cancer are shown in Sahin et al. (US 7,527,933), Hutteman (US 7,473,530), Bae et al. (US 7,368,255), Wang (US 7,348,142), Nacht et al. (US 7,332,590), Gure et al. (US 7,314,721), Patel (US 7,300,765), Price et al. (US 7,229,770), O'Brien et al. (US 7,291,462 and US 6,316,213), Muramatsu et al. (US 7,090,983), Carson et al. (US 6,576,420), Giordano (US 5,840,506), Guo (US 2009/0062144), Tsao et al. (US 2008/0176236), Nakamura et al. (US 2008/0050378), Raponi et al. (US 2006/0252057), Yip et al. (US 2006/0223127), Pollock et al. (US 2006/0046257), Moon et al. (US 2003/0224509), and Budiman et al. (US 2009/0098543). Biomarkers associated with development of skin cancer (e.g., basal cell carcinoma, squamous cell carcinoma, and melanoma) are shown in Roberts et al. (US 6,316,208), Polsky (US 7,442,507), Price et al. (US 7,229,770), Genetta (US 7,078,180), Carson et al. (US 6,576,420), Moses et al. (US 2008/0286811), Moses et al. (US 2008/0268473), Dooley et al. (US 2003/0232356), Chang et al. (US 2008/0274908), Alani et al. (US 2008/0118462), Wang (US 2007/0154889), and Zetter et al. (US 2008/0064047). Biomarkers associated with development of multiple myeloma are shown in Coignet (US 7,449,303), Shaughnessy et al. (US 7,308,364), Seshi (US 7,049,072), and Shaughnessy et al. (US 2008/0293578, US 2008/0234139, and US 2008/0234138). Biomarkers associated with development of leukemia are shown in Ando et al. (US 7,479,371), Coignet (US 7,479,370 and US 7,449,303), Davi et al. (US 7,416,851), Chiorazzi (US 7,316,906), Seshi (US 7,049,072), Van Baren et al. (US 6,130,052), Taniguchi (US 5,643,729), Insel et al. (US 2009/0131353), and Van Bockstaele et al. (Blood Rev. 23(1):25-47, 2009). Biomarkers associated with development of lymphoma are shown in Ando et al. (US 7,479,371), Levy et al. (US 7,332,280), and Arnold (US 5,858,655). Biomarkers associated with development of bladder cancer are shown in Price et al. (US 7,229,770), Orntoft (US 6,936,417), Haak-Frendscho et al. (US 6,008,003), Feinstein et al. (US 6,998,232), Elting et al. (US 2008/0311604), and Wewer et al. (2009/0029372). The content of each of the above references is incorporated by reference herein in its entirety.

Devices and methods described herein may be used to assess the quality of a sample to be analyzed for methylation. DNA methylation is a chemical modification of DNA performed by enzymes called methyltransferases, in which a methyl group (m) is added to certain cytosines (C) of DNA, to yield 5-methylcytosine. This non-mutational (epigenetic) process (mC) is a critical factor in gene expression regulation. See, e.g., J. G. Herman, *Seminars in Cancer Biology*, 9: 359-67, 1999. Research suggests genes with high levels of 5-methylcytosine in a promoter

region are transcriptionally silent, which allows unchecked cell proliferation. Additionally, it is likely that there is a correlation between gene transcription and undermethylation. Methylation patterns of DNA from cancer cells are significantly different from those of normal cells. Therefore, detection of methylation patterns in appropriately selected genes of cancer cells can lead to discrimination of cancer cells from normal (i.e., non-cancerous) cells, thereby providing an approach to early detection of cancer.

A common method for assessing methylation status, e.g., the presence of CpG islands, is methylation specific PCR, also known as MSP. In MSP a nucleic acid sample is treated with a methylation reactant, typically bisulfite, and then amplified in the presence of two sets of primers. One primer set is complementary to sequences with converted Cs and the second primer set is complementary to non-converted Cs. Using these two separate primer sets, both the methylated and unmethylated DNA can be simultaneously amplified, and the amplification products compared (e.g., sequenced) to determine methylation sites in a given sequence. The MSP method, and variations on the MSP method, are described in greater detail in U.S. Patent Nos. 6,265,171, 6,331,393, 6,977,146, 7,186,512, and 7,229,759 all of which are incorporated by reference herein in their entireties.

In some instances, a bisulfite treatment to convert unmethylated Cs to Us degrades the sample, making the later amplification and sequencing steps of little value because the resulting sample is no longer contiguous in the region(s) of interest. Using the techniques described herein, it is possible to assess the post bisulfite treatment sample to determine the quality of the sample, for example the amount of contiguous DNA. Thus, a bisulfite treated sample can be partitioned into samples comprising nucleic acids of different lengths, primer pairs can be introduced along with appropriate probes, the nucleic acids amplified, and the make-up, e.g., the continuity of the nucleic acid sample can be determined. Because the sample may contain unmethylated Cs which are converted to Us as well as methylated Cs which are not converted to Us, it may be necessary to use additional primer sets which are complementary to targets having Cs as well as primer sets complementary to targets having Us.

In certain embodiments, methods of the invention may be used to monitor a patient for recurrence of a cancer. Since the patient has already been treated for the cancer, the genetic profile and particular mutation(s) associated with that patient's cancer are already known. Probes may be designed that specifically hybridize to the region of the nucleic acid that contains

the mutation(s) that is indicative of the cancer for which the patient was previously treated. A patient's sample (e.g., pus, sputum, semen, urine, blood, saliva, stool, or cerebrospinal fluid) may then be analyzed as described above to determine whether the mutant allele(s) is detected in the sample, the presence of which being indicative of recurrence of the cancer.

Droplet Sorting

Methods of the invention may further include sorting the droplets based upon whether the droplets contain a homogeneous population of molecules or a heterogeneous population of molecules. A sorting module may be a junction of a channel where the flow of droplets can change direction to enter one or more other channels, e.g., a branch channel, depending on a signal received in connection with a droplet interrogation in the detection module. Typically, a sorting module is monitored and/or under the control of the detection module, and therefore a sorting module may correspond to the detection module. The sorting region is in communication with and is influenced by one or more sorting apparatuses.

A sorting apparatus includes techniques or control systems, e.g., dielectric, electric, electro-osmotic, (micro-) valve, etc. A control system can employ a variety of sorting techniques to change or direct the flow of molecules, cells, small molecules or particles into a predetermined branch channel. A branch channel is a channel that is in communication with a sorting region and a main channel. The main channel can communicate with two or more branch channels at the sorting module or branch point, forming, for example, a T-shape or a Y-shape. Other shapes and channel geometries may be used as desired. Typically, a branch channel receives droplets of interest as detected by the detection module and sorted at the sorting module. A branch channel can have an outlet module and/or terminate with a well or reservoir to allow collection or disposal (collection module or waste module, respectively) of the molecules, cells, small molecules or particles. Alternatively, a branch channel may be in communication with other channels to permit additional sorting.

A characteristic of a fluidic droplet may be sensed and/or determined in some fashion, for example, as described herein (e.g., fluorescence of the fluidic droplet may be determined), and, in response, an electric field may be applied or removed from the fluidic droplet to direct the fluidic droplet to a particular region (e.g. a channel). In certain embodiments, a fluidic droplet is sorted or steered by inducing a dipole in the uncharged fluidic droplet (which may be initially

charged or uncharged), and sorting or steering the droplet using an applied electric field. The electric field may be an AC field, a DC field, etc. For example, a channel containing fluidic droplets and carrier fluid, divides into first and second channels at a branch point. Generally, the fluidic droplet is uncharged. After the branch point, a first electrode is positioned near the first channel, and a second electrode is positioned near the second channel. A third electrode is positioned near the branch point of the first and second channels. A dipole is then induced in the fluidic droplet using a combination of the electrodes. The combination of electrodes used determines which channel will receive the flowing droplet. Thus, by applying the proper electric field, the droplets can be directed to either the first or second channel as desired. Further description of droplet sorting is shown for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc.

Based upon the detected signal at the detection module, droplets containing a heterogeneous population of molecules are sorted away from droplets that contain a homogeneous population of molecules. Droplets may be further sorted to separate droplets that contain a homogeneous population of amplicons of the target from droplets that contain a homogeneous population of amplicons of the variant of the target.

Threshold Detection

Since the advent of digital detection, many methods have been developed that are based on diagnosing a disease state based on a ratio of normal (wild-type) nucleic acid to abnormal (mutant) nucleic acid in a sample. Digital detection involves partitioning a sample so that individual target molecules within the sample are localized and concentrated within many separate regions. Analysis is then based on analyzing each partitioned portion. In order to achieve only a single target molecule per partition, a sample must be significantly diluted. In fact, most partitioned portions are empty, i.e., include no target molecule. That means that during analysis, significant amounts of time and resources are used to analyze empty partitioned portions.

Embodiments of the invention recognize that due to the disperse nature of partitioned portions, efficiency can be increased by not analyzing empty partitioned portions. Aspects of the invention are accomplished by setting a threshold level on a detector that is above an optical

signal emitted by empty droplets (background), and recording an optical signal above the threshold level emitted by droplets that include amplified nucleic acid. Typically, millions of droplets are produced by systems of the invention, and at 1% loading, the probability that two signal producing droplets will be adjacent is less than 0.01%. Since not every single droplet is being analyzed in these methods of the invention, that means that droplets do not need to be mono-disperse within a channel. In fact, making a plurality of droplets past by the detector simultaneously decreases analysis time and saves on reagents.

This is further explained with reference to Figures 27 and 28. According to one technique for detecting a target nucleic acid, a plurality of droplets is directed through a detection channel (or a detection portion of a channel) in a single-file and spaced manner. This technique is shown in Figure 27A. As shown in Figure 27A, mono-dispersed droplets (e.g. droplets of uniform volume) are flowed through a channel in a single-filed fashion, in which each droplet is spaced apart from another droplet and passed by the detector one at a time. The detector detects the signal intensity of each droplet to determine a number of droplets having amplifiable product and also detects empty droplets. Although Figure 27A shows mono-dispersed droplets, the droplets may be non-mono-dispersed such that the droplets are not uniform in size/volume. As exemplified in Figure 28A, droplets having amplifiable product have a signal intensity (amplifiable intensity) that is higher than the intensity of empty droplets (dim or background intensity). If the detector is set to detect a threshold intensity below the background intensity level, each droplet flowing through the droplet is counted. The detected intensity of each droplet is recorded and can be used for subsequent analysis. Differences in amplifiable intensity can be used to determine which molecule (reference or mutant) was amplified in the droplet.

Droplets having amplifiable intensity (e.g., due to amplified wild-type target (or other reference sequence) and amplified target mutant molecule) may be counted to determine a number Y of reference molecules and a number X of target mutant molecules within the sample. The number Y of reference molecules is compared to the number X of target mutant molecules to ascertain a ratio which reflects the biological sample. In certain embodiments, the ratio is indicative of a condition. The total droplet count, which includes both droplets with amplifiable product and non-amplifiable product, can be used for Poisson correction of the amplifiable products to determine the statistical significance of the ratio of the reference molecules and the target mutant molecules in the sample. Methods of comparing a number of reference molecules

to a number of target mutant molecules in order to determine a clinical condition are described in more detail in U.S. Patent Number 6,440,706, the entirety of which is incorporated by reference.

Using the single-file and spaced-apart technique, the detector collects a large volume of data because every droplet, including amplifiable signals and background signals, is collected. For analysis of large nucleic acid samples, the number of droplets for analysis can be in the millions. Another drawback of the single-file technique is the time it takes to flow droplets in a single file line past the detector so that each droplet can be separately analyzed. As such, an alternative technique for counting droplets that obtains significantly less data in a short time span while still obtaining clinically significant data is desirable.

Methods of the invention provide an alternative technique for assessing droplets that minimize the data obtained by the detector and increase the quantity of droplets flowing through a channel (or portion thereof) that simultaneously pass by the detector. According to the alternative technique of the invention, the detector is set to detect signals at a threshold above the background (i.e. dim) intensity level (Figure 28B). Figure 28B illustrates the threshold level of the detector being above the background intensity level. In this manner, the detector only collects signals from droplets containing amplifiable products (reference nucleic acid or template nucleic acid) and empty droplets are processed as background signal. Since only droplets having amplifiable products are being detected, this technique does not require that the droplets pass by the detector in a mono-dispersed configuration because the probability that adjacent droplets will simultaneously pass by the detector is very low, i.e., less than 0.01% at 1% loading. In other words, two or more droplets may pass by the detector at the same time without affecting the results because statistically only one of the droplets is likely to contain an amplifiable product (thereby producing an amplifiable signal above the threshold). In the event that two or more droplets are within the detection field at the same time, the signals may be separately identifiable because the peaks of the signals will be shown separated in time. An added benefit of flowing the plurality of droplets pass the detector in a non single-filed and spaced-apart configuration is that less carrier fluid is needed to flow the same amount of droplets passed the detector. Figure 27B illustrates a plurality of droplets passing the detector that are not in a single-filed and spaced apart configuration. Although the droplets shown in Figure 27B are monodispersed (uniform volume), this technique may also be used with non mono-dispersed droplets.

In another alternative technique for assessing droplets, droplets are passed by the detector in a non-spaced apart manner in order to decrease the time it takes for the each droplet to pass by the detector. In this technique (shown in Figure 27C), one or more droplets passing by the detector are in contact with each other without coalescing into a single droplet. This allows the detector to detect signals from more droplets over a period of time. In one embodiment, a neck formed from one or more indentations is placed in the channel above the detector to slow the pace of a droplet while it is within the detection field of the detector. As shown, the neck causes the droplet to elongate as it passes by the channel. In addition, the neck in Figure 27C includes two indentations; however the same effect may be accomplished with only one indentation. This increases the detector's ability to separate signals from each droplet. In the event that two or more droplets are within the detection field at the same time, the signals may be separately identifiable because the peaks of the signals will be shown separated in time. In this technique, the detector may also be set to detect signals at a threshold above the background (i.e. dim) intensity level (as shown in FIG 28B). While only one droplet is passed by the detector at a time using the necking technique, the high threshold significantly reduces the file size of the detected signals. With this technique, the droplets may be monodispersed or of non-uniform volume.

Because each droplet contains a single template molecule, each amplifiable intensity above background intensity that is recorded may be counted as one droplet (which represents, e.g., either a reference molecule or a target mutant molecule). This is true for detection methods shown in Figures 27B and 27C. Droplets having amplifiable intensity (e.g., due to amplified wild-type target (or other reference sequence) and amplified mutant target) may be counted to determine a number Y of reference molecules and a number X of target mutant molecules within the sample. Differences in amplifiable intensity can be used to determine which molecule (reference or mutant) was amplified in the droplet. The number Y of reference molecules is compared to the number of target mutant molecules to ascertain a ratio which reflects the biological sample. In certain embodiments, the ratio is indicative of a condition.

Release of Target from Droplet

Methods of the invention may further involve releasing amplified target molecules from the droplets for further analysis. Methods of releasing amplified target molecules from the droplets are shown in for example in Link et al. (U.S. patent application numbers 2008/0014589,

2008/0003142, and 2010/0137163) and European publication number EP2047910 to RainDance Technologies Inc.

In certain embodiments, sample droplets are allowed to cream to the top of the carrier fluid. By way of non-limiting example, the carrier fluid can include a perfluorocarbon oil that can have one or more stabilizing surfactants. The droplet rises to the top or separates from the carrier fluid by virtue of the density of the carrier fluid being greater than that of the aqueous phase that makes up the droplet. For example, the perfluorocarbon oil used in one embodiment of the methods of the invention is 1.8, compared to the density of the aqueous phase of the droplet, which is 1.0.

The creamed liquids are then placed onto a second carrier fluid which contains a de-stabilizing surfactant, such as a perfluorinated alcohol (e.g. 1H,1H,2H,2H-Perfluoro-1-octanol). The second carrier fluid can also be a perfluorocarbon oil. Upon mixing, the aqueous droplets begins to coalesce, and coalescence is completed by brief centrifugation at low speed (e.g., 1 minute at 2000 rpm in a microcentrifuge). The coalesced aqueous phase can now be removed and the further analyzed.

The released amplified material can also be subjected to further amplification by the use tailed primers and secondary PCR primers. In this embodiment the primers in the droplet contain an additional sequence or tail added onto the 5' end of the sequence specific portion of the primer. The sequences for the tailed regions are the same for each primer pair and are incorporated onto the 5' portion of the amplicons during PCR cycling. Once the amplicons are removed from the droplets, another set of PCR primers that can hybridize to the tail regions of the amplicons can be used to amplify the products through additional rounds of PCR. The secondary primers can exactly match the tailed region in length and sequence or can themselves contain additional sequence at the 5' ends of the tail portion of the primer. During the secondary PCR cycling these additional regions also become incorporated into the amplicons. These additional sequences can include, but are not limited to adaptor regions utilized by sequencing platforms for library preparation and sequencing, sequences used as a barcoding function for the identification of samples multiplexed into the same reaction. molecules for the separation of amplicons from the rest of the reaction materials such as biotin, digoxin, peptides, or antibodies and molecules such as fluorescent markers that can be used to identify the fragments.

In certain embodiments, the amplified target molecules are sequenced. In a particular embodiment, the sequencing is single-molecule sequencing-by-synthesis. Single-molecule sequencing is shown for example in Lapidus et al. (U.S. patent number 7,169,560), Quake et al. (U.S. patent number 6,818,395), Harris (U.S. patent number 7,282,337), Quake et al. (U.S. patent application number 2002/0164629), and Braslavsky, et al., PNAS (USA), 100: 3960-3964 (2003), the contents of each of these references is incorporated by reference herein in its entirety.

Briefly, a single-stranded nucleic acid (e.g., DNA or cDNA) is hybridized to oligonucleotides attached to a surface of a flow cell. The single-stranded nucleic acids may be captured by methods known in the art, such as those shown in Lapidus (U.S. patent number 7,666,593). The oligonucleotides may be covalently attached to the surface or various attachments other than covalent linking as known to those of ordinary skill in the art may be employed. Moreover, the attachment may be indirect, e.g., via the polymerases of the invention directly or indirectly attached to the surface. The surface may be planar or otherwise, and/or may be porous or non-porous, or any other type of surface known to those of ordinary skill to be suitable for attachment. The nucleic acid is then sequenced by imaging the polymerase-mediated addition of fluorescently-labeled nucleotides incorporated into the growing strand surface oligonucleotide, at single molecule resolution.

Determining the Nucleic Acid Make-Up of a Sample

Further aspects of the invention include methods for determining the nucleic acid make-up of a sample. Specifically, the method can determine the presence of a contiguous, intact nucleic acid, i.e., an unbroken chain of nucleotides, between two locations on the nucleic acid. Presence of a contiguous nucleic acid is determined via detection of both a first and second detectably labeled probe that hybridizes to a first and second location on the nucleic acid (e.g., a sequence, an oligomer, a polymer, a template, dsDNA). The detection of only one probe indicates a fragmented nucleic acid, in other words, a nucleic acid that is not contiguous through the entirety of the two aforementioned locations on the nucleic acid. In some embodiments, the method involves partitioning a sample comprising nucleic acid of different lengths into a plurality of partitioned portions, wherein each portion comprises, on average, a single nucleic acid molecule, introducing first and second primer pairs and first and second detectably labeled probes to the partitioned portions, wherein the first and second primer pairs are specific for first

and second locations on the nucleic acid, the first and second locations being spaced apart from each other, and wherein the first probe hybridizes to the first location and the second probe hybridizes to the second location, amplifying the nucleic acid in the partitioned portions, the presence of signal from both probes indicating the presence of a nucleic acid that is contiguous between the first and second locations, and determining a nucleic acid make-up of the sample based upon results of the detecting step.

Specific methods for the partitioning, introducing, amplifying, and detecting steps have been presented throughout the present disclosure. Further detail on determination step is now presented. As mentioned above, determination of contiguous or intact nucleic acid involves detecting a first and second detectably labeled probe that hybridizes to a first and second location on a nucleic acid. The detection of only one probe indicates the presence of a fragment of the longer nucleic acid. In some embodiments, the determining step may involve comparing relative amounts of contiguous nucleic acid to relative amounts of non-contiguous nucleic acid. In other embodiments, the determining step involves comparing amount of contiguous nucleic acid or non-contiguous nucleic acid to a total amount of amino acid.

For sequencing and other extensive molecular biology studies, a sample comprising mostly intact nucleic acid is desirable and is conducive to accurate results. The presence of a relatively large population of nucleic acid fragments, i.e., non-contiguous nucleic acids, may indicate that a sample is not suitable for sequencing. Because sequencing is relatively expensive, knowing the make-up of the sample prior to testing is advantageous. This is especially so when the sample is an FFPE sample or some other form of preserved sample in which the nucleic acid of interest has degraded into fragments. In some embodiments, if less than 90% of the nucleic acid sample, e.g., less than 80% of the nucleic acid sample, e.g., less than 70% of the nucleic acid sample, e.g., less than 50% of the nucleic acid sample, e.g., less than 40% of the nucleic acid sample, e.g., less than 30% of the nucleic acid sample, e.g., less than 20% of the nucleic acid sample, e.g., less than 10% of the nucleic acid sample is fragmented, the nucleic acid sample is suitable for further sequencing. Accordingly, once the nucleic acid make-up of a sample is determined, a further step may involve sequencing the sample, enriching the sample, or sequencing the sample after enrichment. Furthermore, because the method of determining the nucleic acid make-up incorporates the dPCR methods described throughout the present

disclosure, extremely small amounts of sample can be tested successfully. In some embodiments, the test sample may contain 50 ng or less of DNA or RNA.

A demonstration of concept in accordance with the invention is presented schematically in Figure 23. Two alleles are presented: Allele 1, comprising genes 88G and 815A, and Allele 2, comprising genes 88A and 815G. Alleles 1 and 2 were digested with a first enzyme to produce a first nucleic acid containing 88G and 815A (“88G_815A”) and a second nucleic acid containing 88A and 815G (“88A_815G”). A portion of these products were then subjected to a second digestion with a second enzyme to obtain four additional fragments, each containing only one gene of interest (“88G,” “815A”, “88A”, and “815G”). Methods of performing enzymatic digestion as well as selecting the proper enzymes are well-known in the art. While enzymatic digestion was used in this example, other methods of fragmenting or shearing the nucleic acid prior to partitioning the samples can be used. Such means may include, for example, sonication, the use of commercially available shearing devices, such as a Covaris® shearing device, or other means known in the art. In this demonstration of concept, the fragmentation is performed to simulate the degradation of a nucleic acid that may naturally occur in a sample, such as a FFPE sample. In certain embodiments of the invention, however, the nucleic acid is fragmented prior to partitioning the sample.

The six nucleic acids were then partitioned into a plurality of partitioned portions, each portion containing on average, a single nucleic acid molecule. Partitioning may involve any methods known in the art, such as serial or terminal dilution. Partitioning may also involve the methods described in the present disclosure, including forming droplets from a sample present in an aqueous fluid. Additional details on droplet formation have been described throughout the present disclosure.

As further shown in Figure 23, a first and second primer pair and a first and second detectably labeled probe specific for the regions of interest were introduced to the partitioned portions. In this example, the probes were labeled with either VIC™ or FAM. Further detail on probe embodiments has been described throughout the present disclosure. The nucleic acid was then amplified and the amplicons detected through the hybridized probes.

Sample data is presented in Figure 24, which shows the six populations of nucleic acids along with the empty droplets as peaks on a graph. 815A_88G and 815G_88A, oriented

approximately along the diagonal axis, are the longer nucleic acids. 815G, 88G, 815A, and 88A, arranged primarily along the y-axis and x-axis, respectively, are the shorter fragments.

Accordingly, the invention encompasses a method for determining the nucleic acid make-up of a sample. The method can determine the presence of intact or contiguous nucleic acids, represented in this experiment by peaks corresponding to 815A_88G and 815G_88A, as well as the presence of non-contiguous nucleic acids, represented by peaks corresponding to 815G, 88G, 815A, and 88A. In this example, the nucleic acid make-up can be assessed by looking at the number of counts in all the peaks. In a sample where the level of fragmentation is unknown, for example, in a FFPE sample, a prevalence of nucleic acid where only one probe has bound would indicate the presence of non-contiguous nucleic acid and sample degradation. In contrast, the prevalence of nucleic acid where both the first and second probe has bound would indicate the presence of a contiguous nucleic acid and a relatively more intact sample.

The methods described herein are not limited to the use of two probes, however. In some embodiments a plurality of probes are used to give additional information about the properties of nucleic acids in a sample. For example, three probes could be used wherein one probe was one color (e.g., VIC), and two probes were another color (e.g., FAM). Differences in intensity or polarization make it possible to distinguish between the probes of the same color, as discussed previously. Analysis using such a method may make it possible to determine the presence of multiple different contiguous lengths of nucleic acid molecules.

While methods described herein can encompass the use of several primer pairs, methods in accordance with the invention also encompass the use of a single primer pair. In some embodiments, the method includes providing a fluid comprising the sample nucleic acid and a plurality of one or more primer pairs, wherein each primer pair has at least one unique related probe and is selected to be complementary to one or more sequences of known length. The method also includes partitioning the fluid into a plurality of partitions, wherein at least a first portion of the partitions comprise one molecule of the nucleic acid sample having sequences complementary to one or more of the primer pairs, and at least one related probe, and a second portion of the partitions comprise no molecules of the sample nucleic acid having sequences complementary to one or more of the primer pairs. The method further includes conducting a PCR reaction in the partitions, thereby changing a fluorescent property of the first portion of the partitions, detecting the fluorescent property of each partition, and determining the number of

occurrences in the sample nucleic acid of one or more sequences of known length based on the detecting step. In some aspects of the invention, the method further includes comparing a first number of occurrences of a first sequence of known length to a second number of occurrences of a second sequence of a second known length.

Additional embodiments of the invention may also contemplate the use of a single primer pair as well as rely on something other than a probe for detecting the amplified sequence. In certain embodiments, the method comprises partitioning a sample comprising nucleic acid of different lengths into a plurality of partitioned portions, wherein each portion comprises, on average a single nucleic acid molecule. The method further includes introducing at least one primer pair, in which each primer of the pair is specific for a first and second location on the nucleic acid, the first and second locations being spaced apart from each other. The method further includes amplifying the nucleic acid in the partitioned portions, detecting the amplicons in the partitioned portions, and determining a nucleic acid make-up of the sample based on the results of the detecting step. In certain embodiments, the amplicons may be detected with a probe, for example, a fluorescently labeled probe. In other embodiments, the amplicon may be detected with a dye that intercalates within the nucleic acid. The invention also contemplates any other means of detecting nucleic acid sequences known in the art that do not interfere with the other steps described herein.

A demonstration of concept using a single primer pair is shown in Figure 25. Each primer within the primer pair is specific for a first and second location on the nucleic acid. The distance between the two primers for this sequence is known beforehand so that the resulting amplicons will be of a known length. The sequence between the primers is then amplified by PCR. In this particular demonstration of concept, the amplicons are detected via a fluorescent probe, specific for region X¹ on the nucleic acid sequence. As discussed above, other means of detection are encompassed by the invention.

Sample data is presented in Figure 26, which shows as peaks, the hybridization of the fluorescent probe to the amplified sequences along with the empty wells, i.e., partitions containing no molecules of sample nucleic acid. The amplicons, as detected by the X¹ specific probe, appear primarily along the Y-axis. Accordingly, the method encompasses a method of analyzing a sample nucleic acid or determining the nucleic acid make-up of a sample with just one pair of primers. As in the earlier experiment, the nucleic acid make-up of the sample can be

assessed by looking at the number of the counts in the peaks. In this case, by looking at the counts, one could determine the number of occurrences in the sample nucleic acid of a sequence of predetermined length. Accordingly, one could determine whether or not the sample was suitable for further testing based on the number of nucleic acids within the sample having the desired length.

Methods in accordance with the invention also encompass the analysis of cell-free nucleic acids. Collecting and assaying cell-free nucleic acids provides advantages over analysis of cellular nucleic acids in that anomalies, e.g., mutations, are easier to identify in the absence of massive quantities of normal nucleic acids. For example, circulating cell-free tumor DNA has been detected in the serum, plasma, and blood of cancer patients. Cell-free nucleic acids are versatile in that they can be analyzed to detect the presence of mutations, or epigenetic markers of a disease. Cell-free nucleic acids can also be used to identify the presence of foreign pathogens, e.g., a bacterial infection. In some embodiments, the biological sample can be blood, saliva, sputum, urine, semen, transvaginal fluid, cerebrospinal fluid, sweat, breast milk, breast fluid (e.g., breast nipple aspirate), stool, a cell or a tissue biopsy.

The methods of the invention can also be used to evaluate the quality of cell-free nucleic acids, for example cell-free DNA or RNA, which can be obtained from a biological sample. In some instances cell-free nucleic acid is greatly degraded, for example, because the nucleic acid was partially digested by normal metabolic processes in the body. The invention allows the cell-free nucleic acid to be evaluated for quality, e.g., continuity, prior to amplification and sequencing. Thus, a cell-free nucleic acid sample can be partitioned into samples comprising nucleic acids of different lengths, primer pairs can be introduced along with appropriate probes, the nucleic acids amplified, and the make-up, e.g., the continuity of the cell-free nucleic acid sample can be determined.

Primers for Maintaining the Representation of a Nucleic Acid Population for a Sequencing Reaction

This aspect of the invention relates to maintaining the representation of a nucleic acid population for a sequencing reaction. Methods of the invention remove the biases associated with bulk amplification reactions by compartmentalizing a sample into a plurality of compartmentalized portions that include on a subset of the nucleic acids from each sample,

preferably only a single nucleic acid from each sample. The amplification reaction is then conducted on the subset of nucleic acid in each compartmentalized portion, rather than conducting a bulk amplification reaction on all of the nucleic acid in a single vessel. In this manner, the over or under representation of any portion of the population of nucleic acids in the sample is avoided, and the resulting post-amplification nucleic acid population represents the true condition of the sample from which it was obtained. The amplicons from each compartmentalized portion are then pooled and are subjected to sequencing.

Compartmentalizing may involve diluting the sample such that it may be dispensed into different wells of a multi-well plate in a manner such that each well includes a subset of nucleic acids from the sample. In certain embodiments, each compartmentalized portion includes, on average, a single nucleic acid. Other exemplary compartmentalizing techniques are shown for example in, Griffiths et al. (U.S. patent number 7,968,287) and Link et al. (U.S. patent application number 2008/0014589), the content of each of which is incorporated by reference herein in its entirety.

In certain embodiments, the compartmentalizing involves forming droplets and the compartmentalized portions are the droplets. An exemplary method involves for forming droplets involves flowing a stream of sample fluid including the amplicons such that it intersects two opposing streams of flowing carrier fluid. The carrier fluid is immiscible with the sample fluid. Intersection of the sample fluid with the two opposing streams of flowing carrier fluid results in partitioning of the sample fluid into individual sample droplets. The carrier fluid may be any fluid that is immiscible with the sample fluid. An exemplary carrier fluid is oil, particularly, a fluorinated oil. In certain embodiments, the carrier fluid includes a surfactant, such as a fluorosurfactant. The droplets may be flowed through channels.

In certain embodiments, the compartmentalized portions include loci specific primers and secondary primers that include an adaptor sequence and a barcode sequence. During the amplification reaction, the secondary primers interact with the loci specific primers such that each produced amplicon includes an adaptor sequence and a barcode sequence.

Any amplification reaction known in the art may be used with methods of the invention, such as PCR, LCR, rolling circle amplification, transcription-mediated amplification (TMA), strand-displacement amplification (SDA), NASBA, the use of allele-specific oligonucleotides (ASO), allele-specific amplification. In certain embodiments, the amplification method is PCR.

Sequencing may be by any method known in the art. Sequencing-by-synthesis is a common technique used in next generation procedures and works well with the instant invention. However, other sequencing methods can be used, including sequence-by-ligation, sequencing-by-hybridization; gel-based techniques and others. In general, sequencing involves hybridizing a primer to a template to form a template/primer duplex, contacting the duplex with a polymerase in the presence of a detectably-labeled nucleotides under conditions that permit the polymerase to add nucleotides to the primer in a template-dependent manner. Signal from the detectable label is then used as to identify the incorporated base and the steps are sequentially repeated in order to determine the linear order of nucleotides in the template. Exemplary detectable labels include radiolabels, florescent labels, enzymatic labels, etc. In particular embodiments, the detectable label may be an optically detectable label, such as a fluorescent label. Exemplary fluorescent labels include cyanine, rhodamine, fluorescein, coumarin, BODIPY, alexa, or conjugated multi-dyes.

Another aspect of the invention generally relates to sample preparation for multiplex next-generation sequencing applications, but is also applicable across a broad range of detection assays. In sequencing applications, a barcode oligonucleotide is attached to nucleic acid from a sample. The barcode oligonucleotide is unique to nucleic acid from each sample such that no two samples have the same barcoded oligonucleotides. The barcodes serve to map from a given molecule to nucleic acid from a particular sample. Once barcoded, libraries are pooled, optionally amplified, and finally sequenced.

Methods of the invention provide for sample preparation in which sample and all reagents necessary for an amplification reaction are mixed prior to droplet formation, including reagents necessary for attaching adaptor sequences and barcode sequences to the nucleic acids in the sample. Such a method dramatically reduces reagent costs, and preparation time. In certain embodiments, nucleic acid is mixed with loci specific primers and secondary primers that include an adaptor sequence and a barcode sequence. This mixture is partitioned into droplets such that each droplet includes a subset of the nucleic acid, loci specific primers, and secondary primers. In certain embodiments, each droplet includes, on average, a single nucleic acid. The nucleic acid in the droplet is then amplified to thereby produce amplicons in which each amplicon includes an adaptor sequence and a barcode sequence. The droplets are then pooled,

the amplicons are released from the droplets, and the amplicons are then analyzed, for example, by sequencing.

In the methods described herein that include barcodes, the sequencing process consists of two reads, a read of the genomic region and a read of the barcode with the barcode read serving to allow the mapping of the genomic read to nucleic acid from a given sample.

In certain embodiments, methods of the invention utilize at least two sets of primers for maintain the representation of a nucleic acid population for a sequence reaction. In these embodiments, the at least two sets of primers include a loci specific set of primers and a secondary set of primers that interact with the loci specific primers. The loci specific primers include a tail, and the secondary primers include a portion that interacts with the tail of the loci specific primers and a second portion that includes an adaptor sequence and a barcode sequence. The two primer sets are shown in Figure 29.

Any method known in the art may be used to insert the tails onto the loci specific primers and the secondary primers, for example, a ligase, a polymerase, Topo cloning (e.g., Invitrogen's topoisomerase vector cloning system using a topoisomerase enzyme), or chemical ligation or conjugation. The ligase may be any enzyme capable of ligating an oligonucleotide (RNA or DNA) to the primers. Suitable ligases include T4 DNA ligase and T4 RNA ligase (such ligases are available commercially, from New England Biolabs). Methods for using ligases are well known in the art. The polymerase may be any enzyme capable of adding nucleotides to the 3' and the 5' terminus of template nucleic acid molecules.

Exemplary methods for designing sets of barcode sequences and other methods for attaching barcode sequences are shown in U.S. patent numbers 6,138,077; 6,352,828; 5,636,400; 6,172,214; 6,235,475; 7,393,665; 7,544,473; 5,846,719; 5,695,934; 5,604,097; 6,150,516; RE39,793; 7,537,897; 6,172,218; and 5,863,722, the content of each of which is incorporated by reference herein in its entirety.

The barcode sequence generally includes certain features that make the sequence useful in sequencing reactions. For example the barcode sequences can be designed to have minimal or no homopolymer regions, i.e., 2 or more of the same base in a row such as AA or CCC, within the barcode sequence. The barcode sequences can also be designed so that they do not overlap the target region to be sequence or contain a sequence that is identical to the target.

The barcode sequence is designed such that each sequence is correlated to a particular sample, allowing samples to be distinguished and validated. Methods of designing sets of barcode sequences is shown for example in Brenner et al. (U.S. patent number 6,235,475), the contents of which are incorporated by reference herein in their entirety. In certain embodiments, the barcode sequences range from about 2 nucleotides to about 50; and preferably from about 4 to about 20 nucleotides. Since the barcode sequence is sequenced along with the template nucleic acid or may be sequenced in a separate read, the oligonucleotide length should be of minimal length so as to permit the longest read from the template nucleic acid attached. Generally, the barcode sequences are spaced from the template nucleic acid molecule by at least one base.

The secondary primer also includes an adaptor sequence, which is generally a homopolymer region, e.g., a region of poly(A) or poly(T), that can hybridize to a universal primer for the sequence reaction. Adaptor sequences are further described in Sabot et al. (U.S. patent application number 2009/0226975), Adessi et al. (U.S. patent number 7,115,400), and Kawashima et al. (U.S. patent application number 2005/0100900), the content of each of which is incorporated by reference herein in its entirety. In certain embodiments, an “A” and a “B” adapter are introduced into each compartmentalized portion. The “A” adapter and “B” adapter sequences correspond to two surface-bound amplification primers on a flow cell used for amplification of the nucleic acids prior to sequencing, as is discussed in greater detail below.

The purpose of the two primer sets is such that during an amplification reaction, a sequence adaptor and a barcode sequence is added to each amplicon. As shown in Figure 29, A and A' are a primer pair (forward and reverse) of loci specific primers. Each of A and A' includes a target sequence specific portion that hybridizes to the target site on the nucleic acid. A and A' also include a tailed portion, B and B'. The secondary primers include a universal portion of B and B' also, such that the B and B' portions of the loci specific primers and the secondary primers can hybridize with each other. The secondary primers also include a second portion, C and C'. C and C' include an adaptor sequence and a barcode sequence. The result of the amplification reaction using these loci specific primers and secondary primers is amplicons in which each amplicon includes an adaptor sequence and a barcode sequence. Accordingly, the amplicons are ready for sequencing without further sample preparation.

The sample is diluted such that each compartmentalized portion includes only a subset of the nucleic acid in the sample. In certain embodiments, each compartmentalized portion includes, on average, a single nucleic acid. Poisson statistics dictate the dilution requirements needed to insure that each compartment contains only a subset of the nucleic acid in the sample. In particular embodiments, the sample concentration should be dilute enough that most of the compartments contain no more than a single nucleic acid with only a small statistical chance that a compartment will contain two or more molecules. The parameters which govern this relationship are the volume of the compartment and the concentration of nucleic acid in the sample solution. The probability that a compartment will contain two or more nucleic acid ($NAT_{\leq 2}$) can be expressed as:

$$NAT_{\leq 2} = 1 - \{1 + [NAT] \times V\} \times e^{-(NAT) \times V}$$

where "[NAT]" is the concentration of nucleic acid in units of number of molecules per cubic micron (μm^3), and V is the volume of the compartment in units of μm^3 . It will be appreciated that $NAT_{\leq 2}$ can be minimized by decreasing the concentration of nucleic acid in the sample solution.

In embodiments in which loci specific primers and secondary primers are pre-mixed with the sample prior to compartmentalizing the sample, the primers are provided in significantly excess the concentration of the nucleic acid in the sample, thereby ensuring that after sample dilution, every droplet will still receive a set of loci specific primers and secondary primers. One of skill in the art will readily be able to calculate the concentrations of primers needed to ensure that after dilution of the sample, each compartmentalized portion will receive a subset of nucleic acid from the sample and loci specific primers and secondary primers. This pre-mixture also includes reagents for the PCR reaction. Such reagents generally include Taq polymerase, deoxynucleotides of type A, C, G and T, magnesium chloride, all suspended within an aqueous buffer.

In embodiments in which the primers are pre-mixed with the sample, this step is not necessary. However, in certain embodiments, droplets containing nucleic acid are formed and primers are subsequently introduced to those droplets. In these embodiments, along with the primers, reagents for a PCR reaction are also introduced to the droplets. Such reagents generally

include Taq polymerase, deoxynucleotides of type A, C, G and T, magnesium chloride, all suspended within an aqueous buffer.

An exemplary method of introducing primers and PCR reagents to a sample droplet is as follows. After formation of the sample droplet from the first sample fluid, the droplet is contacted with a flow of a second sample fluid stream, which contains the loci specific primers and the secondary primers. Contact between the droplet and the fluid stream results in a portion of the fluid stream integrating with the droplet to form a mixed droplet containing a nucleic acid, primers and PCR reagents.

Droplets of the first sample fluid flow through a first channel separated from each other by immiscible carrier fluid and suspended in the immiscible carrier fluid. The droplets are delivered to the merge area, i.e., junction of the first channel with the second channel, by a pressure-driven flow generated by a positive displacement pump. While droplet arrives at the merge area, a bolus of a second sample fluid is protruding from an opening of the second channel into the first channel. The intersection of the channels may be perpendicular. However, any angle that results in an intersection of the channels may be used, and methods of the invention are not limited to the orientation of the channels.

The bolus of the second sample fluid stream continues to increase in size due to pumping action of a positive displacement pump connected to the second channel, which outputs a steady stream of the second sample fluid into the merge area. The flowing droplet containing the first sample fluid eventually contacts the bolus of the second sample fluid that is protruding into the first channel. Contact between the two sample fluids results in a portion of the second sample fluid being segmented from the second sample fluid stream and joining with the first sample fluid droplet 201 to form a mixed droplet.

In order to achieve the merge of the first and second sample fluids, the interface separating the fluids must be ruptured. In certain embodiments, this rupture can be achieved through the application of an electric charge. In certain embodiments, the rupture will result from application of an electric field. In certain embodiments, the rupture will be achieved through non-electrical means, e.g. by hydrophobic/hydrophilic patterning of the surface contacting the fluids.

Description of applying electric charge to sample fluids is provided in Link et al. (U.S. patent application number 2007/0003442) and European Patent Number EP2004316 to

Raindance Technologies Inc, the content of each of which is incorporated by reference herein in its entirety. Electric charge may be created in the first and second sample fluids within the carrier fluid using any suitable technique, for example, by placing the first and second sample fluids within an electric field (which may be AC, DC, etc.), and/or causing a reaction to occur that causes the first and second sample fluids to have an electric charge, for example, a chemical reaction, an ionic reaction, a photocatalyzed reaction, etc.

The electric field, in some embodiments, is generated from an electric field generator, i.e., a device or system able to create an electric field that can be applied to the fluid. The electric field generator may produce an AC field (i.e., one that varies periodically with respect to time, for example, sinusoidally, sawtooth, square, etc.), a DC field (i.e., one that is constant with respect to time), a pulsed field, etc. The electric field generator may be constructed and arranged to create an electric field within a fluid contained within a channel or a microfluidic channel. The electric field generator may be integral to or separate from the fluidic system containing the channel or microfluidic channel, according to some embodiments.

Techniques for producing a suitable electric field (which may be AC, DC, etc.) are known to those of ordinary skill in the art. For example, in one embodiment, an electric field is produced by applying voltage across a pair of electrodes, which may be positioned on or embedded within the fluidic system (for example, within a substrate defining the channel or microfluidic channel), and/or positioned proximate the fluid such that at least a portion of the electric field interacts with the fluid. The electrodes can be fashioned from any suitable electrode material or materials known to those of ordinary skill in the art, including, but not limited to, silver, gold, copper, carbon, platinum, tungsten, tin, cadmium, nickel, indium tin oxide ("ITO"), etc., as well as combinations thereof. In some cases, transparent or substantially transparent electrodes can be used.

The electric field facilitates rupture of the interface separating the second sample fluid and the droplet. Rupturing the interface facilitates merging of the bolus of the second sample fluid and the first sample fluid droplet. The forming mixed droplet continues to increase in size until it a portion of the second sample fluid breaks free or segments from the second sample fluid stream prior to arrival and merging of the next droplet containing the first sample fluid. The segmenting of the portion of the second sample fluid from the second sample fluid stream occurs as soon as the force due to the shear and/or elongational flow that is exerted on the forming

mixed droplet by the immiscible carrier fluid overcomes the surface tension whose action is to keep the segmenting portion of the second sample fluid connected with the second sample fluid stream. The now fully formed mixed droplet continues to flow through the first channel.

In another technique involves droplet merging. The merging of droplets can be accomplished using, for example, one or more droplet merging techniques described for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc.

In embodiments involving merging of droplets, two droplet formation modules are used. A first droplet formation module produces the sample droplets that on average contain a single target nucleic acid. A second droplet formation module produces droplets that contain reagents for a PCR reaction. Such droplets generally include Taq polymerase, deoxynucleotides of type A, C, G and T, magnesium chloride, loci specific primers and secondary primers, all suspended within an aqueous buffer.

The droplet formation modules are arranged and controlled to produce an interdigitation of sample droplets and PCR reagent droplets flowing through a channel. Such an arrangement is described for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc.

A sample droplet is then caused to merge with a PCR reagent droplet, producing a droplet that includes Taq polymerase, deoxynucleotides of type A, C, G and T, magnesium chloride, loci specific primers, secondary primers, and the target nucleic acid. Droplets may be merged for example by: producing dielectrophoretic forces on the droplets using electric field gradients and then controlling the forces to cause the droplets to merge; producing droplets of different sizes that thus travel at different velocities, which causes the droplets to merge; and producing droplets having different viscosities that thus travel at different velocities, which causes the droplets to merge with each other. Each of those techniques is further described in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc. Further description of producing and controlling dielectrophoretic forces on droplets to cause the droplets to merge is described in Link et al. (U.S. patent application number 2007/0003442) and European Patent Number EP2004316 to Raindance Technologies Inc.

Once compartmentalized, an amplification reaction is conducted on the nucleic acids in the compartmentalized portions. Amplification refers to production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction or other technologies well known in the art (e.g., Dieffenbach and Dveksler, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y. [1995]). The amplification reaction may be any amplification reaction known in the art that amplifies nucleic acid molecules, such as polymerase chain reaction, nested polymerase chain reaction, polymerase chain reaction-single strand conformation polymorphism, ligase chain reaction (Barany F. (1991) PNAS 88:189-193; Barany F. (1991) PCR Methods and Applications 1:5-16), ligase detection reaction (Barany F. (1991) PNAS 88:189-193), strand displacement amplification and restriction fragments length polymorphism, transcription based amplification system, nucleic acid sequence-based amplification, rolling circle amplification, and hyper-branched rolling circle amplification.

Polymerase chain reaction (PCR) refers to methods by K. B. Mullis (U.S. patent numbers 4,683,195 and 4,683,202, hereby incorporated by reference) for increasing concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. The process for amplifying the target sequence includes introducing an excess of primers (oligonucleotides) to a DNA mixture containing a desired target sequence, followed by a precise sequence of thermal cycling. The present invention includes, but is not limited to, various PCR strategies as are known in the art, for example QPCR, multiplex PCR, asymmetric PCR, nested PCR, hotstart PCR, touchdown PCR, assembly PCR, digital PCR, allele specific PCR, methylation specific PCR, reverse transcription PCR, helicase dependent PCR, inverse PCR, intersequence specific PCR, ligation mediated PCR, mini primer PCR, and solid phase PCR, emulsion PCR, and PCR as performed in a thermocycler, droplets, microfluidic reaction chambers, flow cells and other microfluidic devices.

Compartmentalized portions may be amplified using standard thermo cyclers and standard amplification protocols known in the art. See Sambrook et al., Molecular Cloning, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 2001, the content of which is incorporated by reference herein in its entirety. In embodiments in which the compartmentalized portions are droplets, the droplets may be amplified in PCR tubes according to well know techniques or may be amplified as they are flowing through channels.

Methods for performing PCR in droplets are shown for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163), Anderson et al. (U.S. patent number 7,041,481 and which reissued as RE41,780) and European publication number EP2047910 to Raindance Technologies Inc. The content of each of which is incorporated by reference herein in its entirety.

In certain embodiments, the droplets are flowed through a channel in a serpentine path between heating and cooling lines to amplify the nucleic acid in the droplet. The width and depth of the channel may be adjusted to set the residence time at each temperature, which can be controlled to anywhere between less than a second and minutes.

In certain embodiments, the three temperature zones are used for the amplification reaction. The three temperature zones are controlled to result in denaturation of double stranded nucleic acid (high temperature zone), annealing of primers (low temperature zones), and amplification of single stranded nucleic acid to produce double stranded nucleic acids (intermediate temperature zones). The temperatures within these zones fall within ranges well known in the art for conducting PCR reactions. See for example, Sambrook et al. (Molecular Cloning, A Laboratory Manual, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).

In certain embodiments, the three temperature zones are controlled to have temperatures as follows: 95°C (T_H), 55°C (T_L), 72°C (T_M). The prepared sample droplets flow through the channel at a controlled rate. The sample droplets first pass the initial denaturation zone (T_H) before thermal cycling. The initial preheat is an extended zone to ensure that nucleic acids within the sample droplet have denatured successfully before thermal cycling. The requirement for a preheat zone and the length of denaturation time required is dependent on the chemistry being used in the reaction. The samples pass into the high temperature zone, of approximately 95°C, where the sample is first separated into single stranded DNA in a process called denaturation. The sample then flows to the low temperature, of approximately 55°C, where the hybridization process takes place, during which the primers anneal to the complementary sequences of the sample. Finally, as the sample flows through the third medium temperature, of approximately 72°C, the polymerase process occurs when the primers are extended along the single strand of DNA with a thermostable enzyme.

The nucleic acids undergo the same thermal cycling and chemical reaction as the droplets passes through each thermal cycle as they flow through the channel. The total number of cycles in the device is easily altered by an extension of thermal zones. The sample undergoes the same thermal cycling and chemical reaction as it passes through N amplification cycles of the complete thermal device.

In other embodiments, the temperature zones are controlled to achieve two individual temperature zones for a PCR reaction. In certain embodiments, the two temperature zones are controlled to have temperatures as follows: 95°C (T_H) and 60°C (T_L). The sample droplet optionally flows through an initial preheat zone before entering thermal cycling. The preheat zone may be important for some chemistry for activation and also to ensure that double stranded nucleic acid in the droplets are fully denatured before the thermal cycling reaction begins. In an exemplary embodiment, the preheat dwell length results in approximately 10 minutes preheat of the droplets at the higher temperature.

The sample droplet continues into the high temperature zone, of approximately 95°C, where the sample is first separated into single stranded DNA in a process called denaturation. The sample then flows through the device to the low temperature zone, of approximately 60°C, where the hybridization process takes place, during which the primers anneal to the complementary sequences of the sample. Finally the polymerase process occurs when the primers are extended along the single strand of DNA with a thermostable enzyme. The sample undergoes the same thermal cycling and chemical reaction as it passes through each thermal cycle of the complete device. The total number of cycles in the device is easily altered by an extension of block length and tubing.

The amplification reaction produces numerous amplicons within each compartmentalized portion, and each amplicon now includes an adaptor sequence and a barcode sequence.

Methods of the invention may further involve pooling the compartmentalized portions and releasing the nucleic acid from the compartmentalized portions for further analysis. For well plate based methods, samples may simply be pooled into a single vessel. For droplet based embodiments, methods of releasing amplified target molecules from droplets are shown in for example in Link et al. (U.S. patent application numbers 2008/0014589, 2008/0003142, and 2010/0137163) and European publication number EP2047910 to Raindance Technologies Inc.

In certain embodiments, sample droplets are allowed to cream to the top of the carrier fluid. By way of non-limiting example, the carrier fluid can include a perfluorocarbon oil that can have one or more stabilizing surfactants. The droplet rises to the top or separates from the carrier fluid by virtue of the density of the carrier fluid being greater than that of the aqueous phase that makes up the droplet. For example, the perfluorocarbon oil used in one embodiment of the methods of the invention is 1.8, compared to the density of the aqueous phase of the droplet, which is 1.0.

The creamed liquids are then placed onto a second carrier fluid which contains a de-stabilizing surfactant, such as a perfluorinated alcohol (e.g. 1H,1H,2H,2H-Perfluoro-1-octanol). The second carrier fluid can also be a perfluorocarbon oil. Upon mixing, the aqueous droplets begins to coalesce, and coalescence is completed by brief centrifugation at low speed (e.g., 1 minute at 2000 rpm in a microcentrifuge). The coalesced aqueous phase can now be removed and the further analyzed.

In certain embodiments, the barcoded amplicons are then sequenced. Sequencing may be by any method known in the art.

Experimental Detail

What follows is experimental detail for the various experiments details above.

Primers and probes

All TaqMan[®] primers and probes used here are listed in Table 2. Unless otherwise noted by reference in the table, the primers and probes were designed with the “Custom TaqMan[®] Assay Design Tool” from Applied Biosystems Inc. (ABI) and procured through ABI (Carlsbad, CA). Probes were labeled with 6-carboxyfluorescein (FAM, λ_{ex} 494 nm \ λ_{em} 494 nm) or VIC[™] (from ABI, λ_{ex} 538 nm \ λ_{em} 554 nm).

<i>Target</i>	<i>Assay</i>	<i>Primers (5' to 3')</i>	<i>Probe (5' to 3')</i>	<i>5-plex assay conditions</i>	<i>Ref.</i>
SMN1	Copy number	(f) AATGCTTTTAA- CATCCATATAAAG CT (SEQ ID NO.: 1) (r) CCTTAATTTAAG-	FAM- CAGGGTTTC*AGACAAA -MGBNFQ (SEQ ID NO.: 3)	0.37×	Anhuf <i>et al.</i> , 2003

SMN2	Copy number	GAATGTGAGCACC (SEQ ID NO.: 2)			
		(f) AATGCTTTTAA- CATCCATATAAAG CT (SEQ ID NO.: 4)	FAM- TGATTTTGTCTA*AAA- CCC-MGBNFQ (SEQ ID NO.: 6)	0.76×	Anhuf <i>et al.</i> , 2003
BCKDH A	Copy number	(r) CCTTAATTTAAG- GAATGTGAGCACC (SEQ ID NO.: 5)			
		(f) CAACCTACTCTT- CTCAGACGTGTA (SEQ ID NO.: 7)	(FAM/VIC)- CAGGAGATGCCCG- CCCAGCTC-TAMRA (SEQ ID NO.: 9)	FAM: 0.18× VIC: 0.56×	DiMatteo <i>et al.</i> , 2008
c.815A> G	SNP	(r) TCGAAGTGATCC- AGTGGGTAGTG (SEQ ID NO.: 8)			
		(f) TGCTGATGCTTT- GGGAAGTATGTTA (SEQ ID NO.: 10)	(A) (FAM/VIC)- CATGAGTGG- CTA*TCATAC-MGBNFQ (SEQ ID NO.: 11)	0.9×	
RNaseP	Copy number	(r) TGTCAGGAAAAG- ATGCTGAGTGATT (SEQ ID NO.: 12)	(G) FAM- ATGAGTGGCTG*TC- ATAC-MGBNFQ (SEQ ID NO.: 13); VIC-CATGA- GTGGCTG*TCATAC - MGBNFQ (SEQ ID NO.: 14)	FAM: 0.9× VIC: 0.45×	
		Unknown	unknown	n/a	Standard product, 4403326 , ABI

Table 2 5'-exonuclease genotyping assay design. Assay conditions in column 5 are specific to the multiplexed SMA assay. References: D. Anhuf, T. Eggermann, S. Rudnik-Schöneborn and K. Zerres, *Hum Mutat.*, 2003, **22**, 74-78; D. DiMatteo, S. Callahan and E. B. Kmiec, *Exp Cell Res.*, 2008, **15**, 878-886.

Target DNA

For some genetic targets, BCKDHA and SMN2, plasmid DNA was synthesized (GeneArt, Regensburg, Germany) containing the sequence spanning between the primers (see Table 2) and cloned into the GeneArt standard vector (2.5 kb). The target fragment was released from the cloning vector by restriction digestion with SfiI to avoid any DNA supercoiling that might affect the assay. For simplicity, these gene fragments are called “plasmid DNA”

throughout the text. A string of different gene fragments was also synthesized (GeneArt) and cloned into the GeneArt standard vector for demonstration of multiplexed reactions, called an “artificial chromosome” in the text. In this case, the fragments were separated from each other by restriction digestion at flanking EcoRV sites. Human DNA was obtained in already purified form from cell lines (See Table 3; Coriell, Camden, NJ) and fragmented before use with a K7025-05 nebulizer following manufacturer’s instructions (Invitrogen, Carlsbad, CA). DNA concentration was quantified by measuring absorbance at 260 nm on a Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE).

<i>Patient number</i>	<i>Coriell cell line</i>	<i>Patient number, cont’d.</i>	<i>Coriell cell line, cont’d.</i>
1	NA14638	11	NA13714
2	NA14637	12	NA13712
3	NA14097	13	NA13709
4	NA14096	14	NA13707
5	NA14094	15	NA13705
6	NA14093	SMA carrier	NA03814
7	NA14092	SMA 1	NA03813
8	NA14091	SMA 2	NA00232
9	NA14090	SMA 3	NA09677
10	NA13715	SMA 4	NA10684

Table 3 Map of patient numbers used in the text to Coriell cell lines.

Microfluidics

Microfluidic chips were manufactured by conventional soft lithography. Molding masters were fabricated by spin coating SU-8 negative photoresist (MicroChem Corp., Newton, MA) onto 6 inch silicon wafers and transferring the fluidic features from photomasks (CAD/Art Services, Bandon, OR) by contact lithography with an OAI Hybralign Series 200 aligner (OAI, San Jose, CA). Chips contained channels with two depths: deep channels with low hydrodynamic resistance (100 ± 10 μm) for transporting fluid from external ports to the functional regions of the chip, and shallow channels (20 ± 1 μm) for droplet manipulation and detection. SU-8 photoresists 2100 and 2025 were used for deep and shallow channels respectively. Polydimethylsiloxane (PDMS) (Sylgard 184, Dow Corning, Midland, MI) chips were molded from the negative masters within mold housings of custom design. Glass cover

slides were permanently bonded to the fluidic side of the chips by surface activation in an AutoGlo oxygen plasma system (Glow Research, Phoenix, AZ) followed by immediate contact bonding. To create hydrophobic surfaces, the microfluidic channels were exposed for ~2 min to 1H,1H,2H,2H-perfluorodecyltrichlorosilane (Alfa Aesar, Ward Hill, MA) dissolved in FC-3283 (3M Specialty Materials, St. Paul, MN) prepared as a mixture of 18 g silane in 100 μ L solvent.

Two different microfluidic devices were used, one for droplet generation and the other for fluorescence readout after thermal cycling. The droplet generation chip created an emulsion of uniformly sized aqueous droplets of template DNA and PCR master mix that were suspended in an inert fluorinated oil with an emulsion stabilizing surfactant, called “carrier oil” from this point forward (REB carrier oil; RainDance Technologies, Lexington, MA). Droplets were generated in a cross-shaped microfluidic intersection, or “nozzle”. As shown in Fig 3a, under typical operation the aqueous phase flowed into the nozzle from the right (160 μ L/hr), joining flows of the carrier oil from the top and bottom (750 μ L/hr of total oil), and producing 4 μ L droplets at a rate of 11 kHz. The channel widths at the intersection measured 15 μ m for the aqueous inlet, 12.5 for the oil inlets, and 15 μ m widening to 40 μ m at the outlet. Flow was driven by custom OEM pumps (IDEX Corporation, Northbrook, IL).

Approximately 25 μ L of the PCR reaction mixture was collected as an emulsion from the droplet generation chip and thermally cycled in a DNA Engine (Bio-Rad, Hercules, CA). The reaction mixture contained 1 $\times$ TaqMan universal PCR master mix (Applied Biosystems, Carlsbad, CA), 0.2 mM dNTP (Takara Bio, Madison, WI), and various amounts of primer pairs and probes as described in the results. 1 $\times$ assay concentration is defined as 0.2 μ M probes with 0.9 μ M primers. In all cases, when varied from the 1 $\times$ concentration, the primers and probes were varied by the same amount. The cycler program included a 10 min hot start at 95 $^{\circ}$ C, and 45 cycles of 15 s at 95 $^{\circ}$ C and 60 s at 60 $^{\circ}$ C.

The droplets became concentrated during off-chip handling because the carrier oil is more dense than the aqueous phase and drained down from the emulsion. Hence the droplets were reinjected into the readout chip as a tightly packed emulsion that required dilution prior to readout to properly distinguish one droplet from another. A “spacer” nozzle similar to the droplet generation nozzle above was used to inject uniform plugs of extra carrier oil between droplets immediately before readout. As shown in Fig. 3B, the droplet entrance into the nozzle tapered down into a constriction about the size of an individual droplet forcing the droplets to

enter the nozzle in single file and consequently at a stable rate. Opposed flow of the carrier oil from the top and bottom channels separated the droplets uniformly. The channel leaving the spacer nozzle increased in width along the direction of flow, and the droplets were interrogated by laser induced fluorescence at the location along the channel where the width was smaller than or equal to the droplet diameter (marked with an arrow in Fig. 3B). The nozzle dimensions were 15 μm for the droplet entrance and exit, and 20 μm for the oil lines.

Instrumentation

Fluorescence readout was performed by conventional epifluorescence microscopy with a custom microscope. A 20 mW, 488 nm laser source (Cyan; Picarro, Sunnyvale, CA) was expanded 2x and focused by the objective lens (20x / 0.45 NA; Nikon, Japan) onto the microfluidic channel. Two band pass filters discriminated the fluorescence collected through the objective lens: 512/25 nm and 529/28 nm for FAM and VIC fluorophores respectively (Semrock, Rochester, NY). Fluorescence was detected by two H5784-20 photomultipliers (Hamamatsu, Japan) and was typically recorded at a 200 kHz sampling rate with a USB-6259 data acquisition card (National Instruments, Austin, TX). The data traces were smoothed by a seven-point, second-order Savitzky-Golay algorithm before subsequent analysis. Concurrent with the fluorescence read out, the droplets were imaged through the same objective lens with backside illumination from an 850 nm LED (TSHG6200; Vishay Semiconductors, Shelton, CT), a short pass filter to separate the optical paths for fluorescence detection and imaging, and a Guppy CCD camera (Allied Vision Technologies, Newburyport, MA). Droplets were imaged with short illumination pulses (5-20 μs) to avoid image streaking.

Data Analysis

Data was analyzed with custom LabView software (National Instruments, Austin, TX) that interpreted droplet events as contiguous bursts of fluorescence intensity above a threshold value. The signal-to-noise ratio was generally quite high and the signal levels were consistent from day to day, hence a fixed threshold value of 50 mV was used predominantly, otherwise the threshold was set by eye. The peak fluorescence intensity was recorded for each droplet event for both VIC and FAM fluorophores. Some coalescence of droplets did occur during thermal cycling, typically as isolated events between two intact droplets forming “doublets.” Doublets and the rare larger coalesced events were easily filtered from the data set on based on the duration of the fluorescence burst.

Incorporation by Reference

References and citations to other documents, such as patents, patent applications, patent publications, journals, books, papers, web contents, have been made throughout this disclosure. All such documents are hereby incorporated herein by reference in their entirety for all purposes.

Equivalents

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the invention described herein.

What is claimed is:

1. A method of detecting a target molecule, the method comprising:
 - providing a plurality of droplets, wherein one or more droplets comprise a target molecule;
 - flowing the plurality of droplets through a microfluidic channel, wherein the plurality of droplets are not single-file within a portion of the channel; and
 - flowing a plurality of droplets simultaneously past a detector, thereby detecting the target molecule in one or more of the droplets.
2. The method of claim 1, wherein the target molecule is selected from the group consisting of: a nucleic acid, a protein, an antibody, an aptamer, a cell, and a microorganism.
3. The method of claim 1, wherein prior to the providing step, the method comprises forming the plurality of droplets.
4. The method of claim 1, wherein the plurality of droplets are aqueous droplets in an immiscible fluid.
5. The method of claim 4, wherein the immiscible fluid is oil.
6. The method of claim 4, wherein the oil comprises a surfactant.
7. The method of claim 6, wherein the surfactant is a fluorosurfactant.
8. The method of claim 5, wherein the oil is a fluorinated oil.
9. The method of claim 1, wherein a plurality of the droplets comprise no more than a single target molecule.
10. The method of claim 1, wherein one or more droplets are empty.

11. A method for diagnosing a condition, the method comprising:

providing a plurality of droplets, wherein a subset of droplets comprise nucleic acid from a sample;

amplifying nucleic acid within the droplets;

detecting a signal above background for only droplets that comprise amplified nucleic acid; and

determining a ratio of a first type of nucleic acid to a second type of nucleic acid from the detected droplets, wherein the ratio is indicative of a condition.

12. The method of claim 11, wherein one or more droplets are empty.

13. The method of claim 12, wherein detecting comprising:

setting a threshold level on a detector that is above an optical signal emitted by the empty droplets; and

recording an optical signal above the threshold level emitted by droplets that comprise amplified nucleic acid.

14. The method of claim 11, wherein a plurality of the droplets comprise no more than a single nucleic acid.

15. The method of claim 11, wherein prior to the providing step, the method comprises forming the plurality of droplets.

16. The method of claim 11, wherein the plurality of droplets are aqueous droplets in an immiscible fluid.

17. The method of claim 16, wherein the immiscible fluid is oil.

18. The method of claim 17, wherein the oil comprises a surfactant.

19. The method of claim 18, wherein the surfactant is a fluorosurfactant.

20. The method of claim 17, wherein the oil is a fluorinated oil.
21. The method of claim 11, wherein the first type of nucleic acid is different from the second type of nucleic acid.
22. A method of detecting a target molecule, the method comprising:
 providing a plurality of droplets, wherein one or more droplets comprise a target molecule;
 flowing the plurality of droplets through a microfluidic channel; and
 flowing a plurality of droplets past a detector, wherein the plurality of droplets contact each other without coalescing as they flow past the detector, thereby detecting the target molecule in one or more of the droplets.
23. The method of claim 22, wherein the target molecule is selected from the group consisting of: a nucleic acid, a protein, an antibody, an aptamer, a cell, and a microorganism.
24. The method of claim 22, wherein prior to the providing step, the method comprises forming the plurality of droplets.
25. The method of claim 22, wherein the plurality of droplets are aqueous droplets in an immiscible fluid.
26. The method of claim 25, wherein the immiscible fluid is oil.
27. The method of claim 25, wherein the oil comprises a surfactant.
28. The method of claim 27, wherein the surfactant is a fluorosurfactant.
29. The method of claim 26, wherein the oil is a fluorinated oil.

30. The method of claim 21, wherein a plurality of the droplets comprise no more than a single target molecule.

31. The method of claim 21, wherein one or more droplets are empty.

32. The method of claim 21, wherein the microchannel comprises a neck-down region that corresponds with where the detector interrogates the droplets.

33. The method of claim 21, wherein the droplets are uniform in size.

34. The method of claim 21, wherein the droplets are non-uniform in size.

1/51

FIG. 1

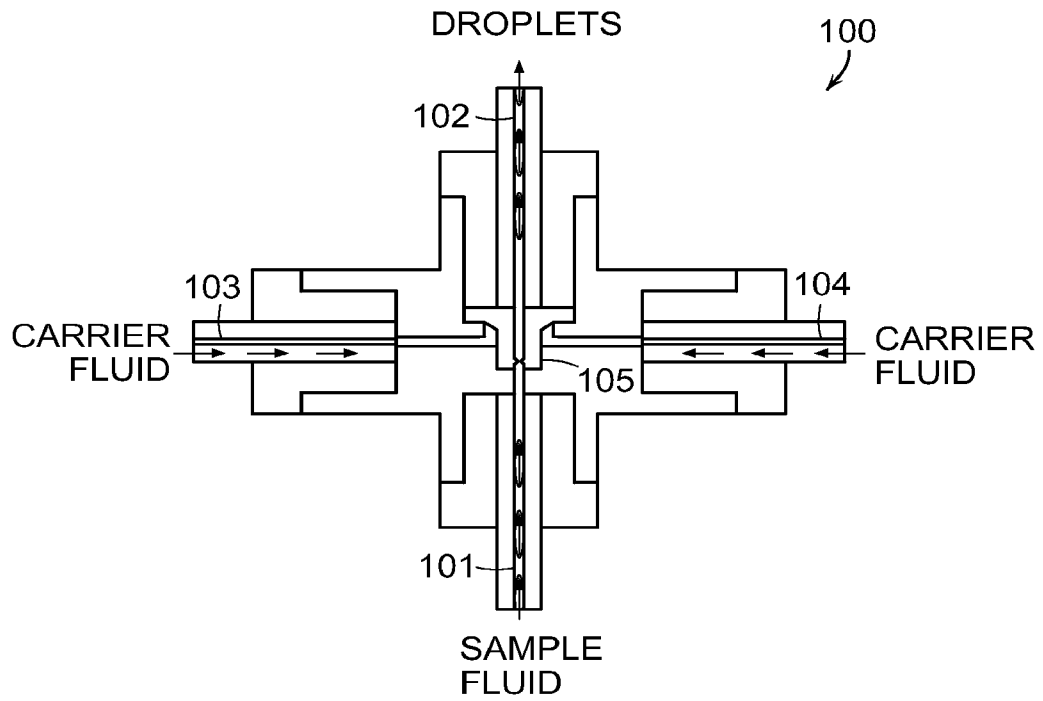
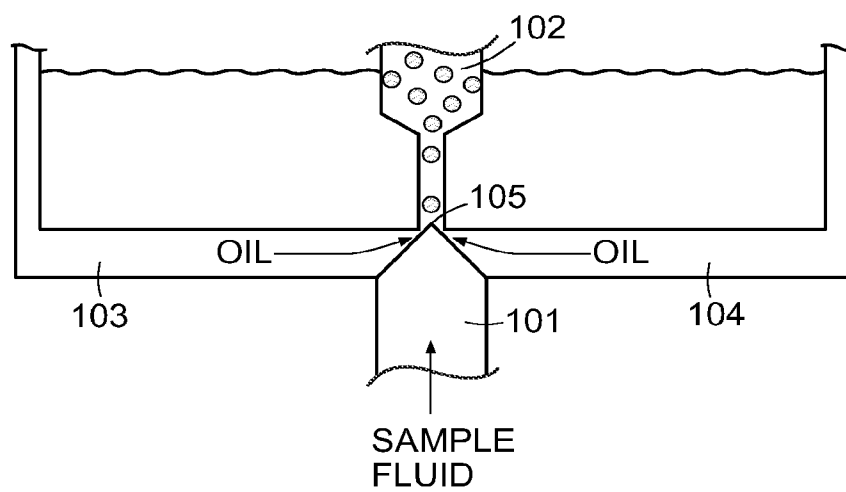


FIG. 2



2/51

FIG. 3A

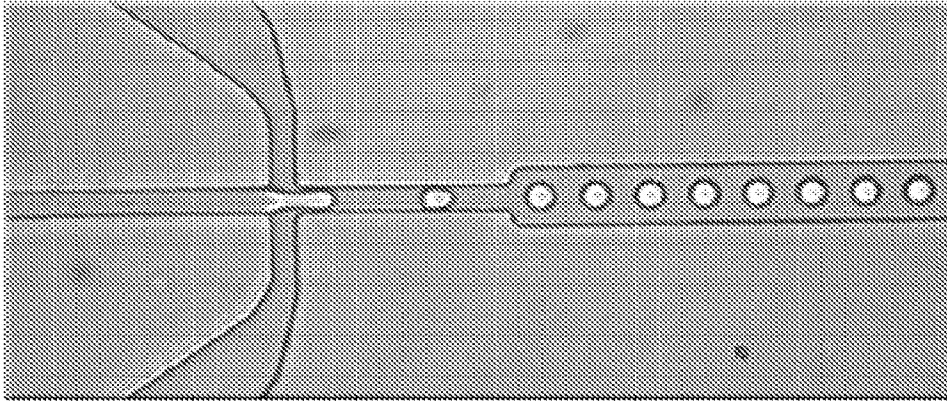


FIG. 3B

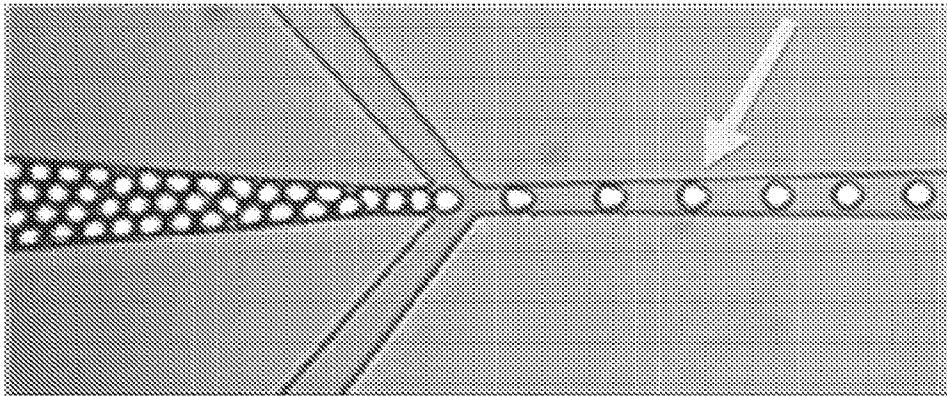
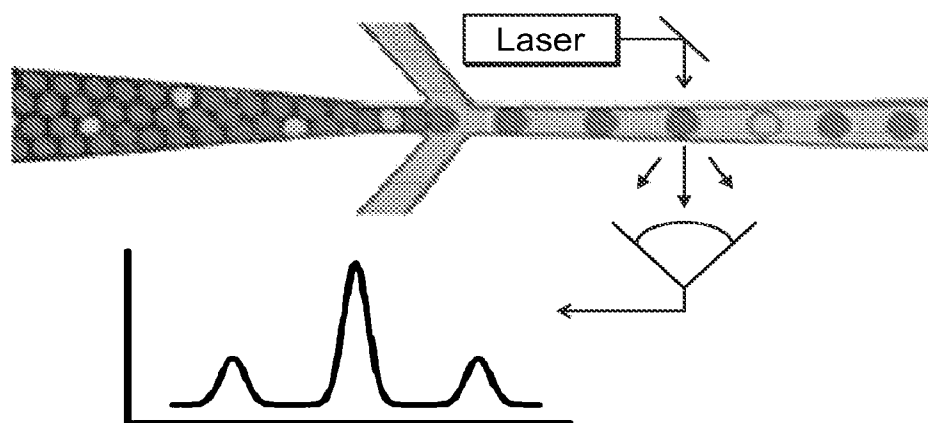


FIG. 3C



3/51

FIG. 4A

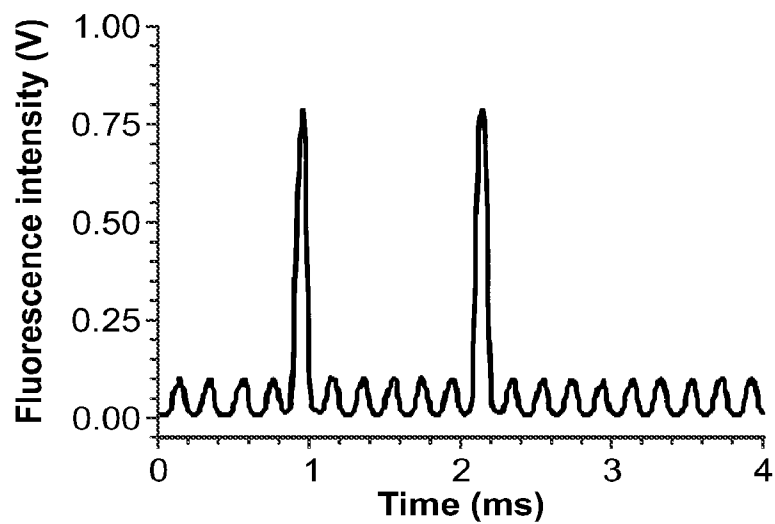


FIG. 4B

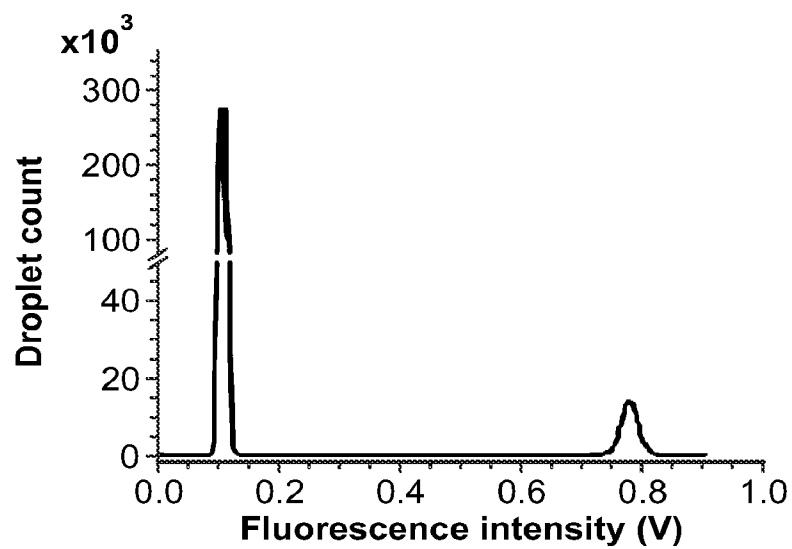
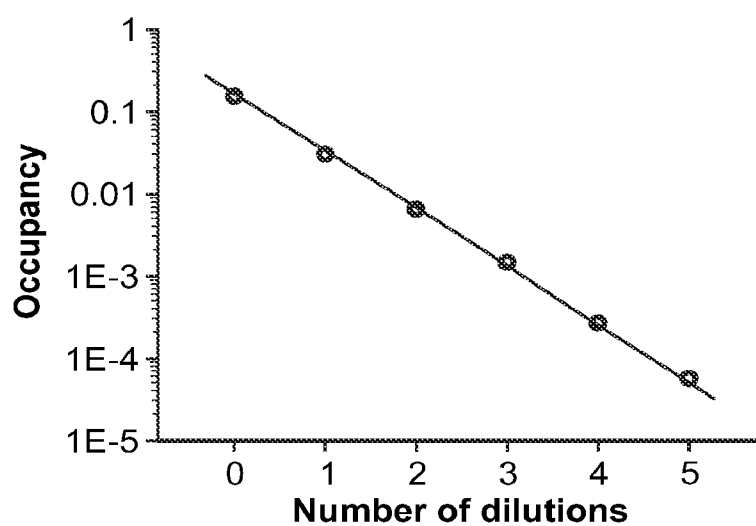


FIG. 4C



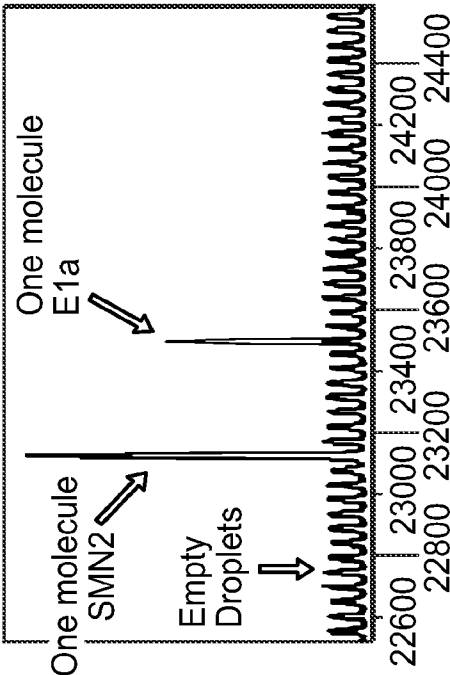


FIG. 5B

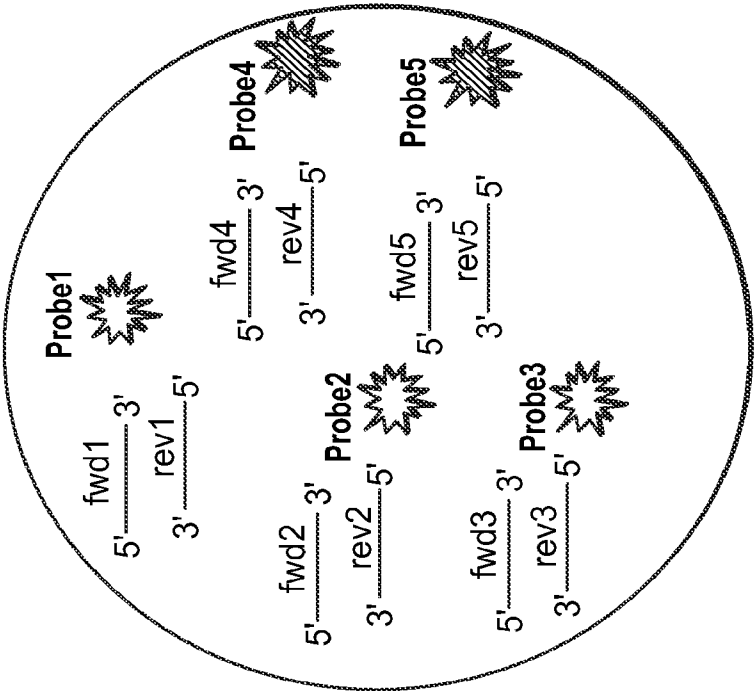


FIG. 5A

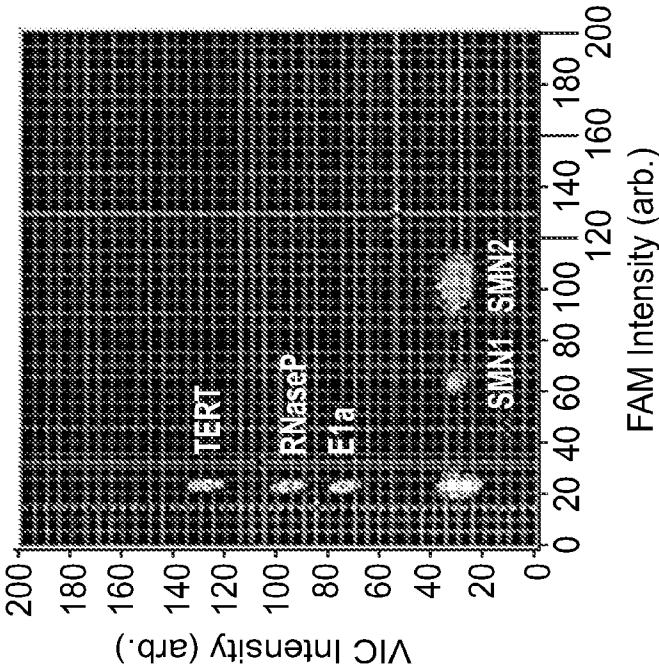


FIG. 5C

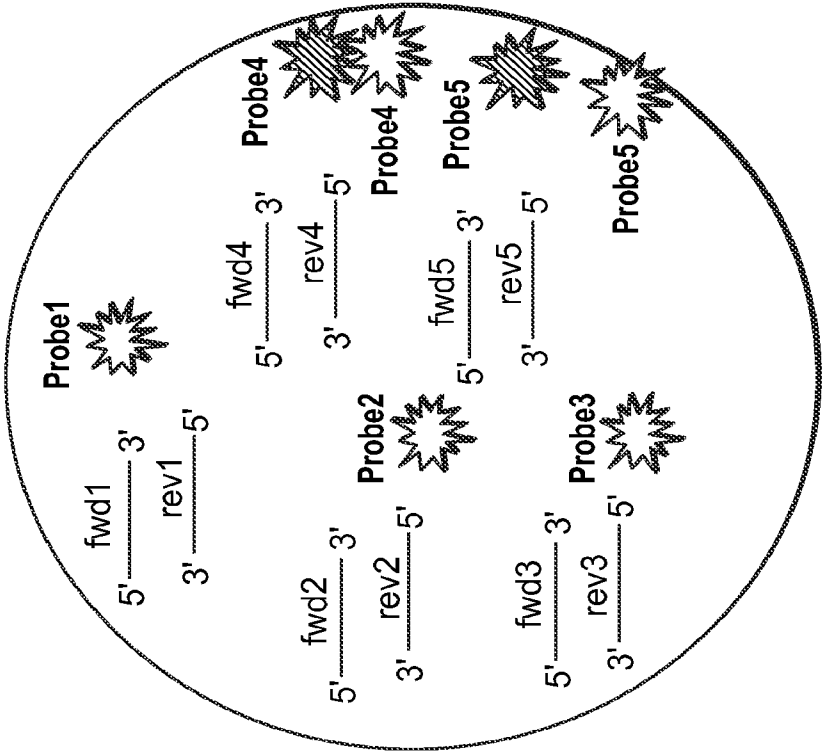
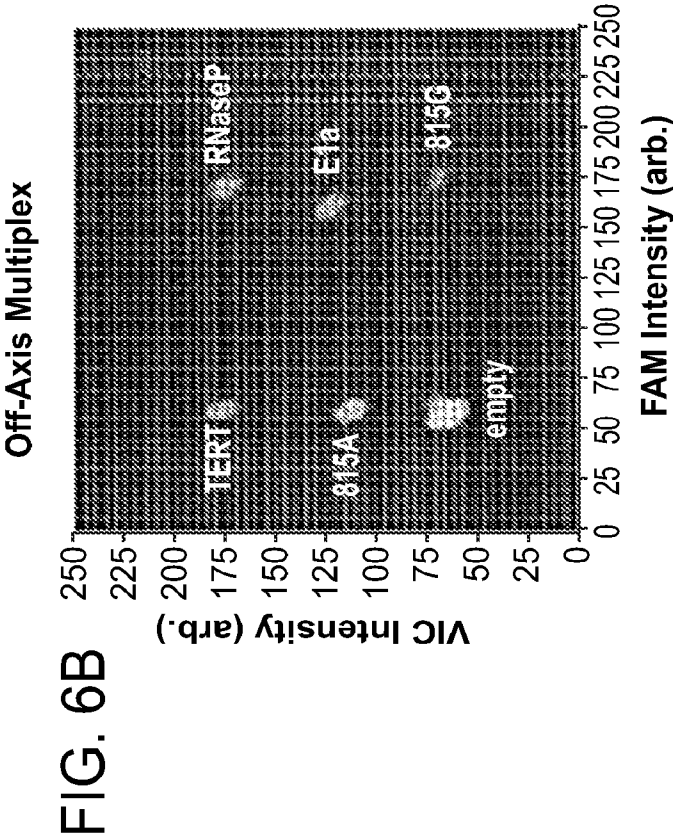


FIG. 6C

NA03814+815G 215,897 drops	Copy Number	Measured Ratio/TERT
RNaseP	2	1.01
E1a	2	0.997
815A	6	3.02
815G	1	0.497
TERT	2	1

6/51

FIG. 7A

One Color One Target

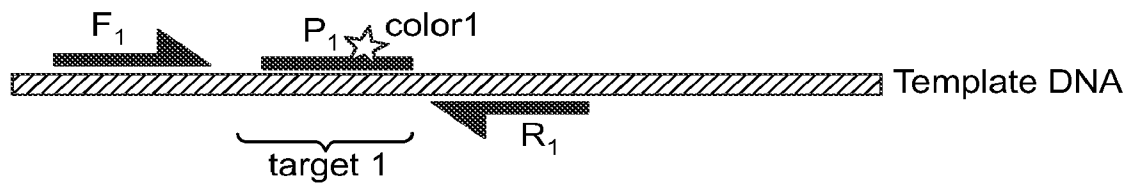


FIG. 7B

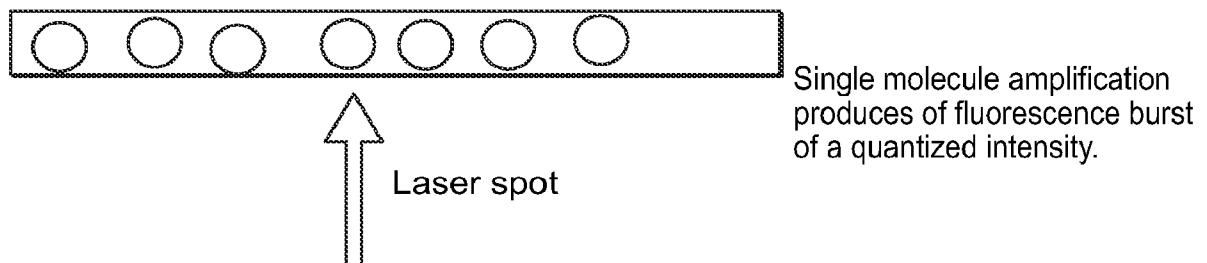
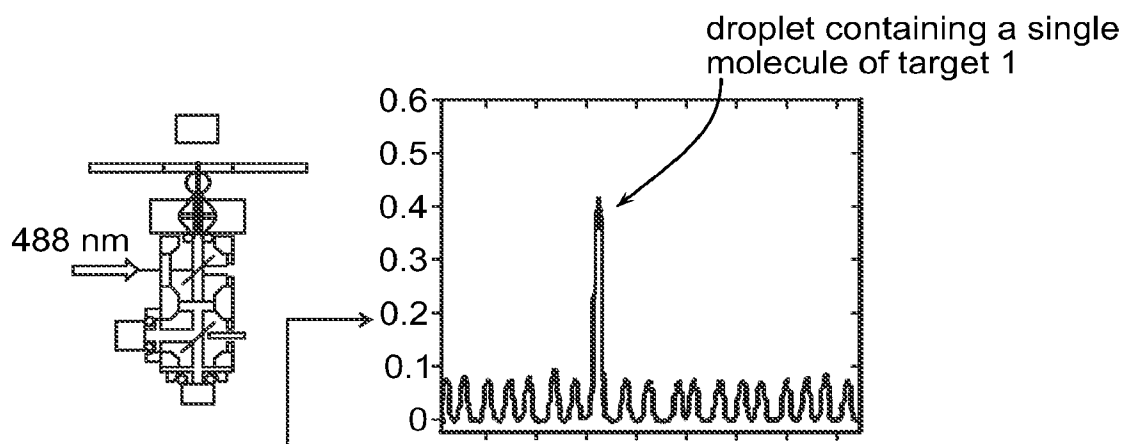


FIG. 7C



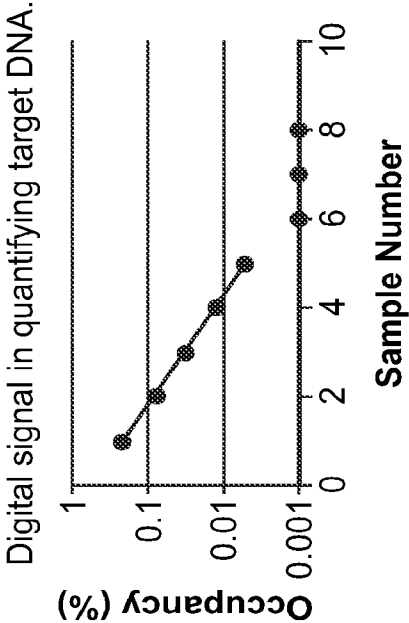


FIG. 7E

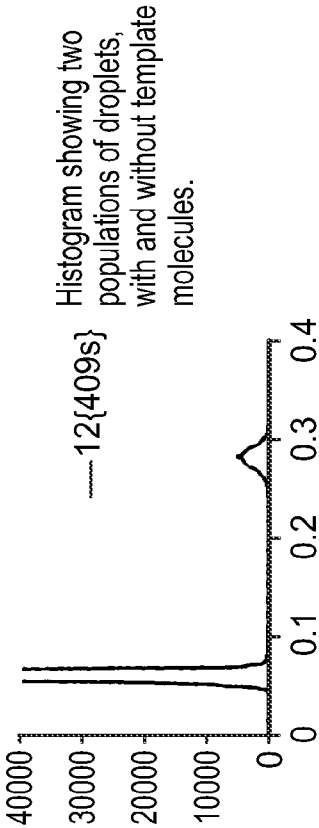


FIG. 7D

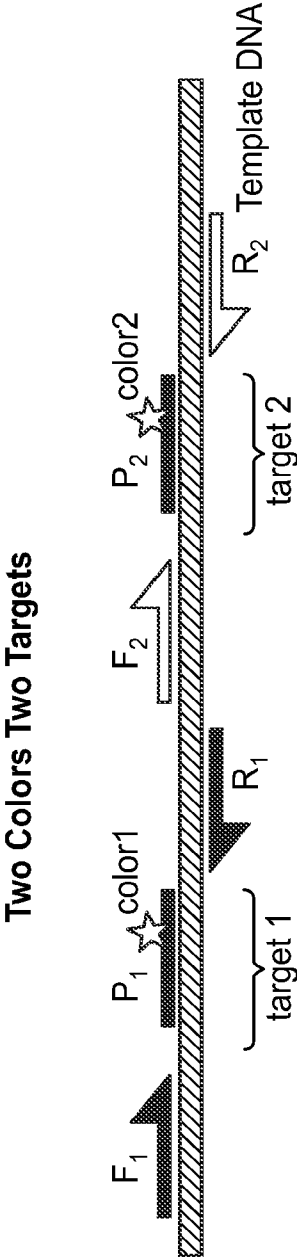
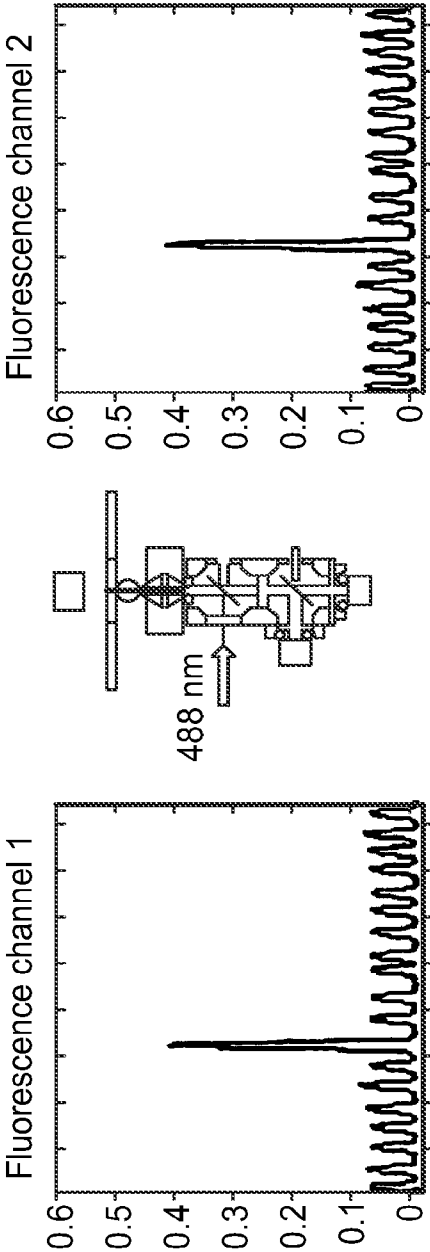
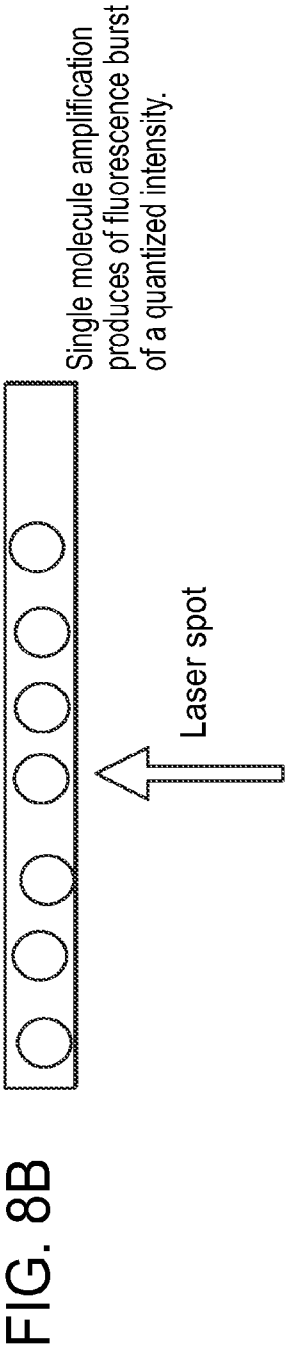


FIG. 8A



9/51

Histogram showing three populations of droplets; with target1, with target2, without target1 or target2.

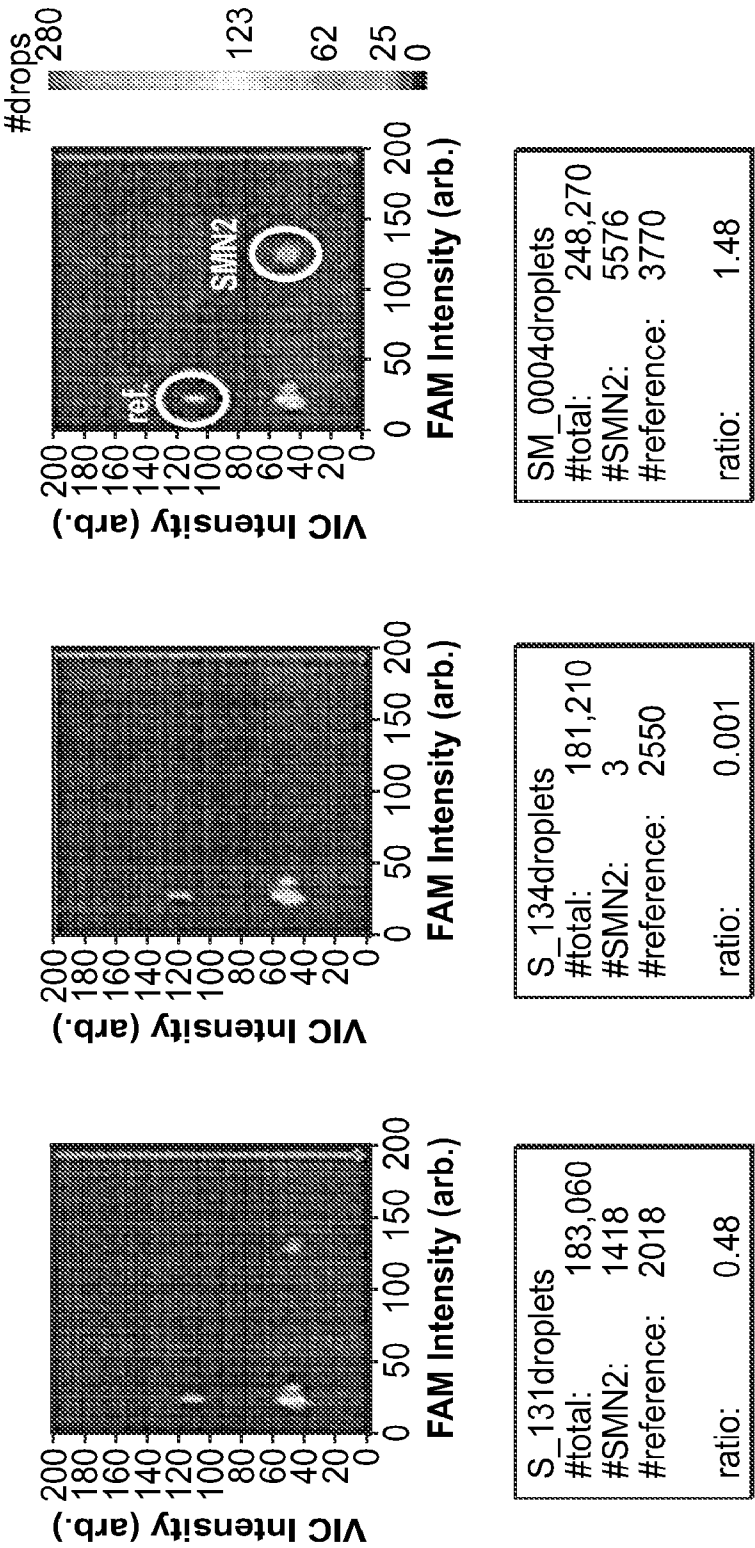


FIG. 8D

Two Colors THREE Targets

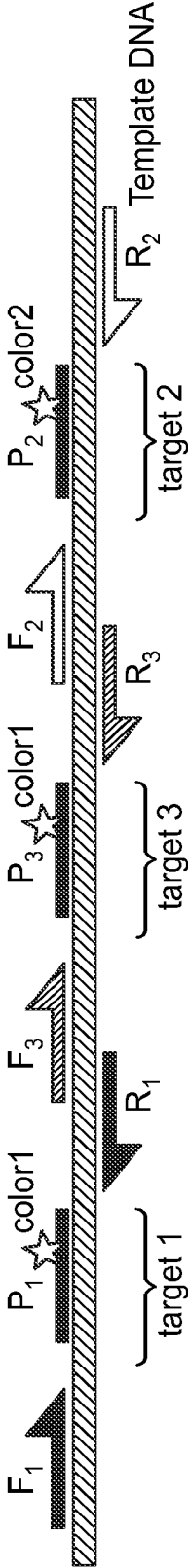


FIG. 9A

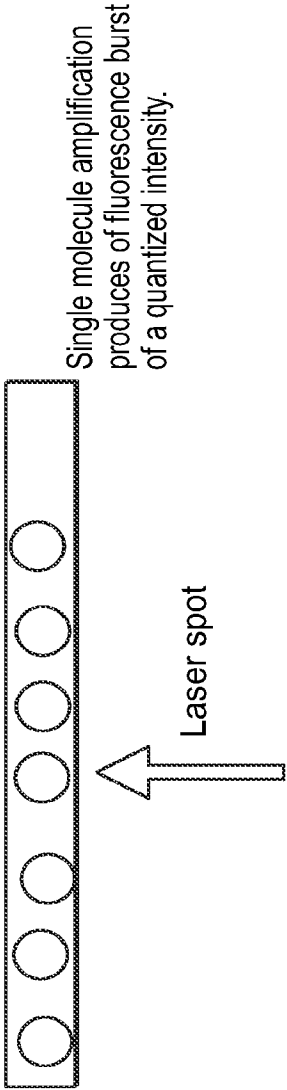


FIG. 9B

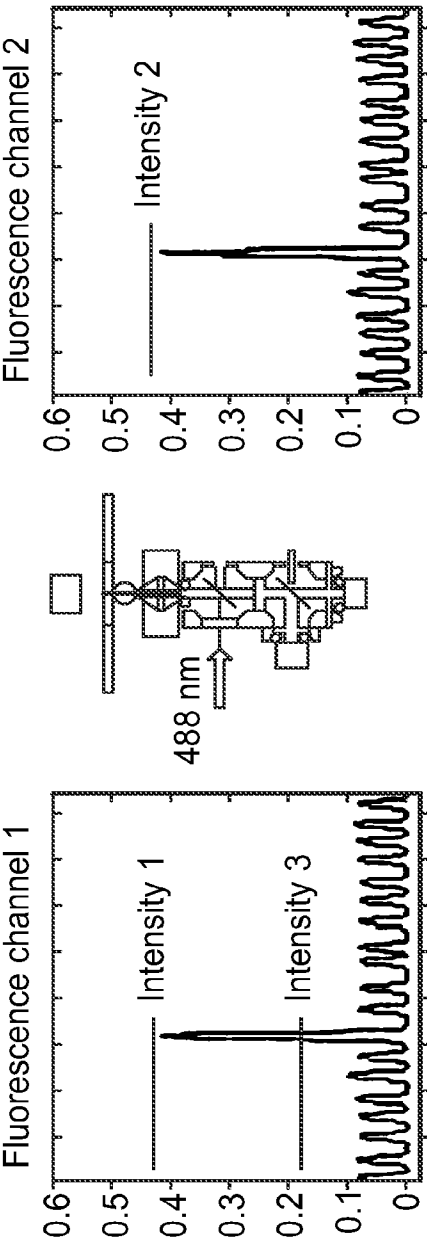


FIG. 9C

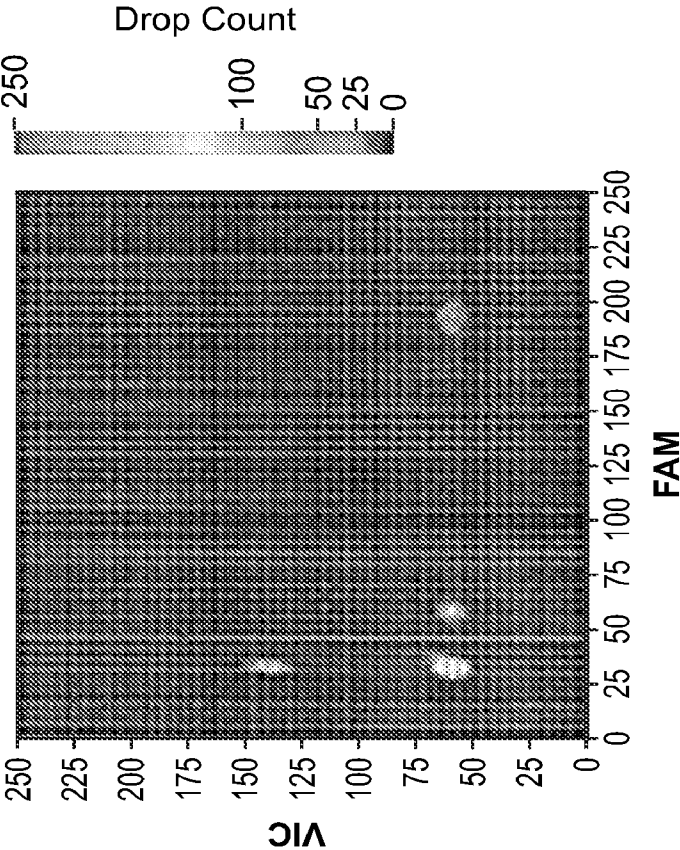


FIG. 9D

Histogram showing four populations of droplets; with target1, with target2, with target3 and without any of the three targets.

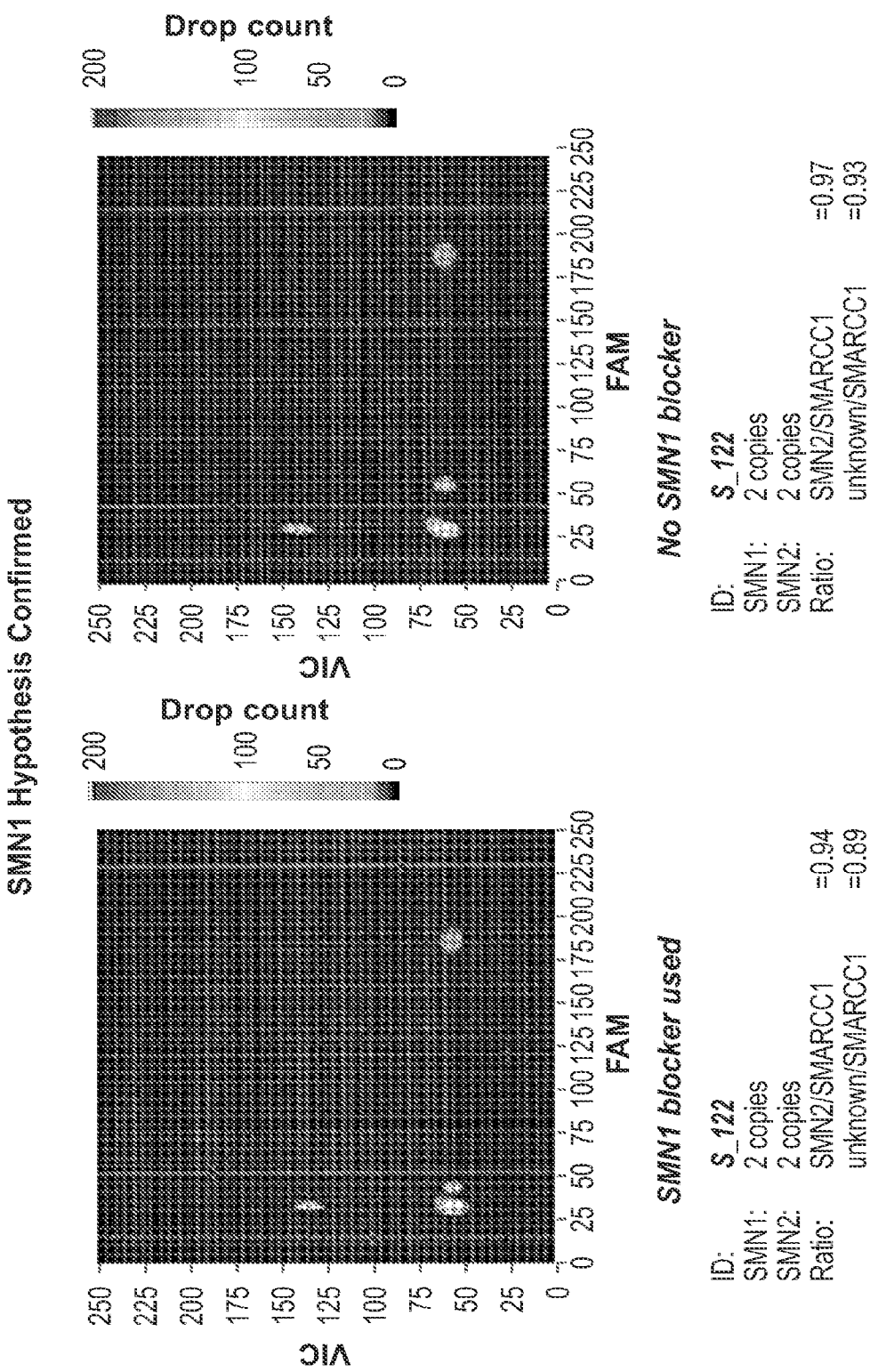


FIG. 10

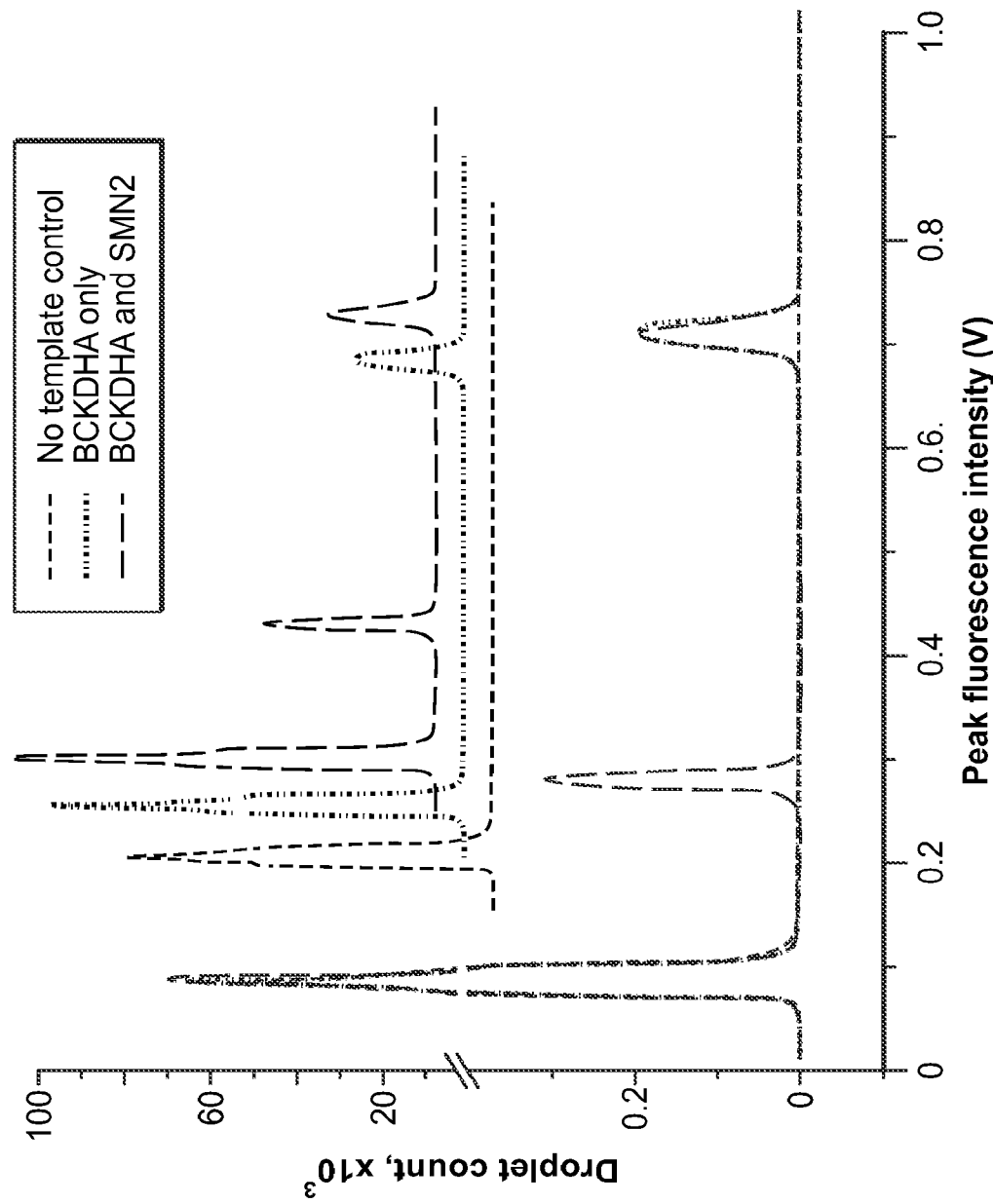


FIG. 11A

14/51

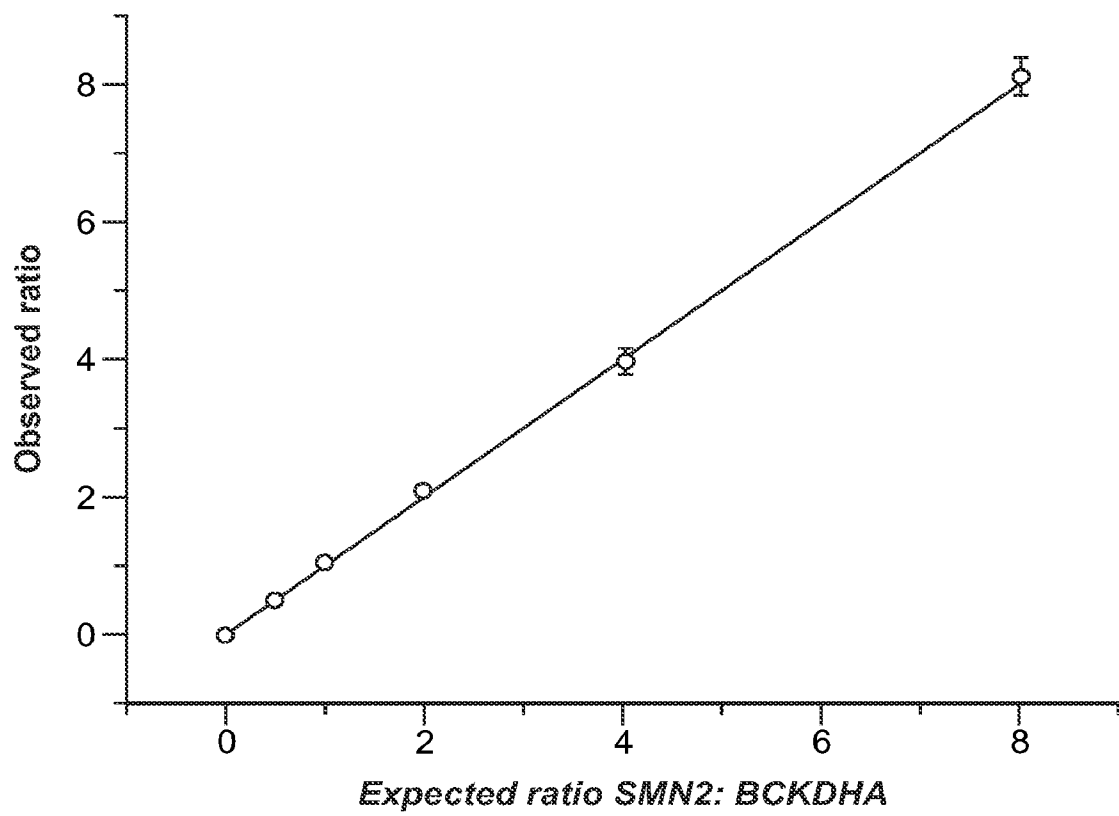


FIG. 11B

15/51

FIG. 12A

Tuning the Intensity of Target 2

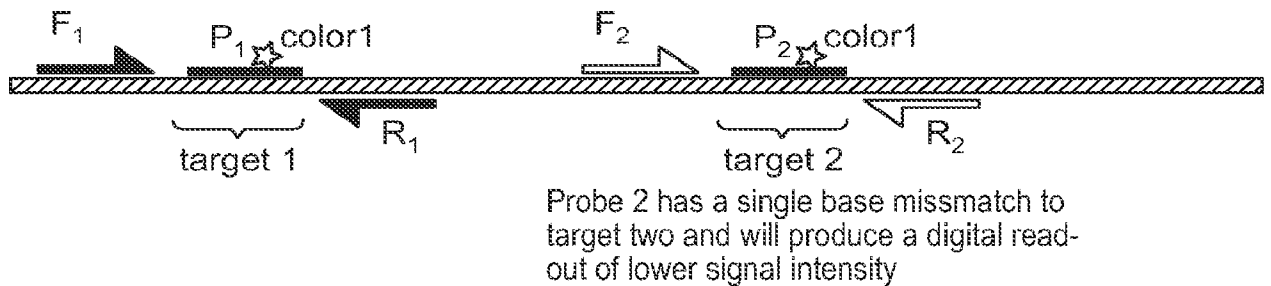


FIG. 12B

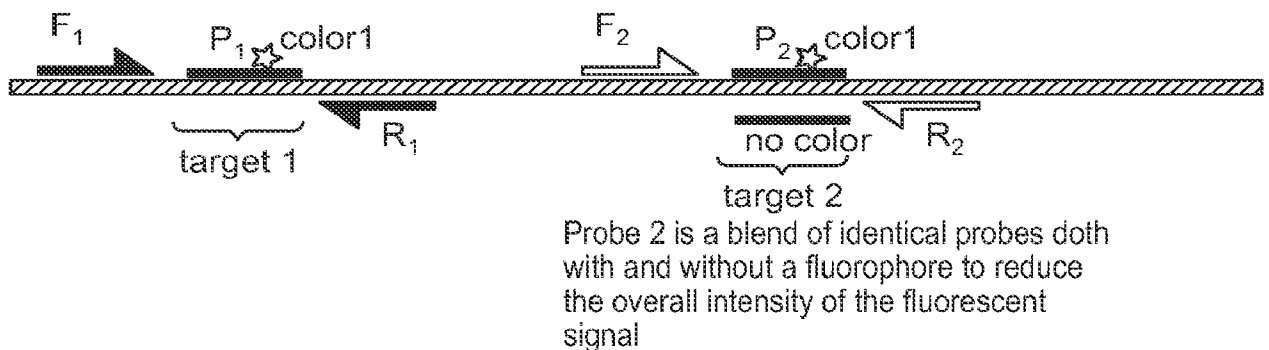
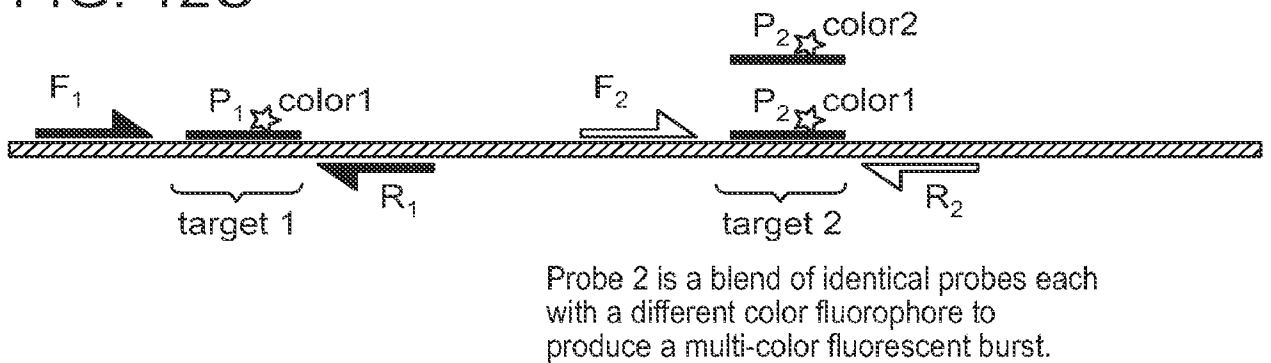


FIG. 12C



16/51

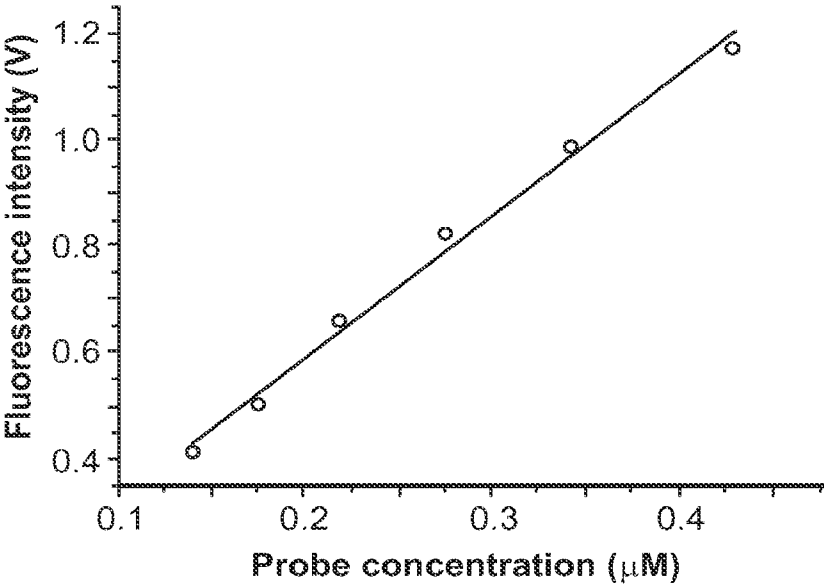


FIG. 13

17/51

FIG. 14A

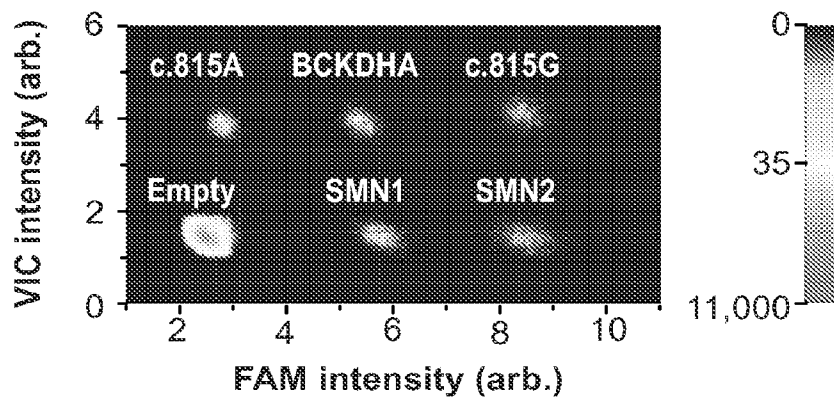
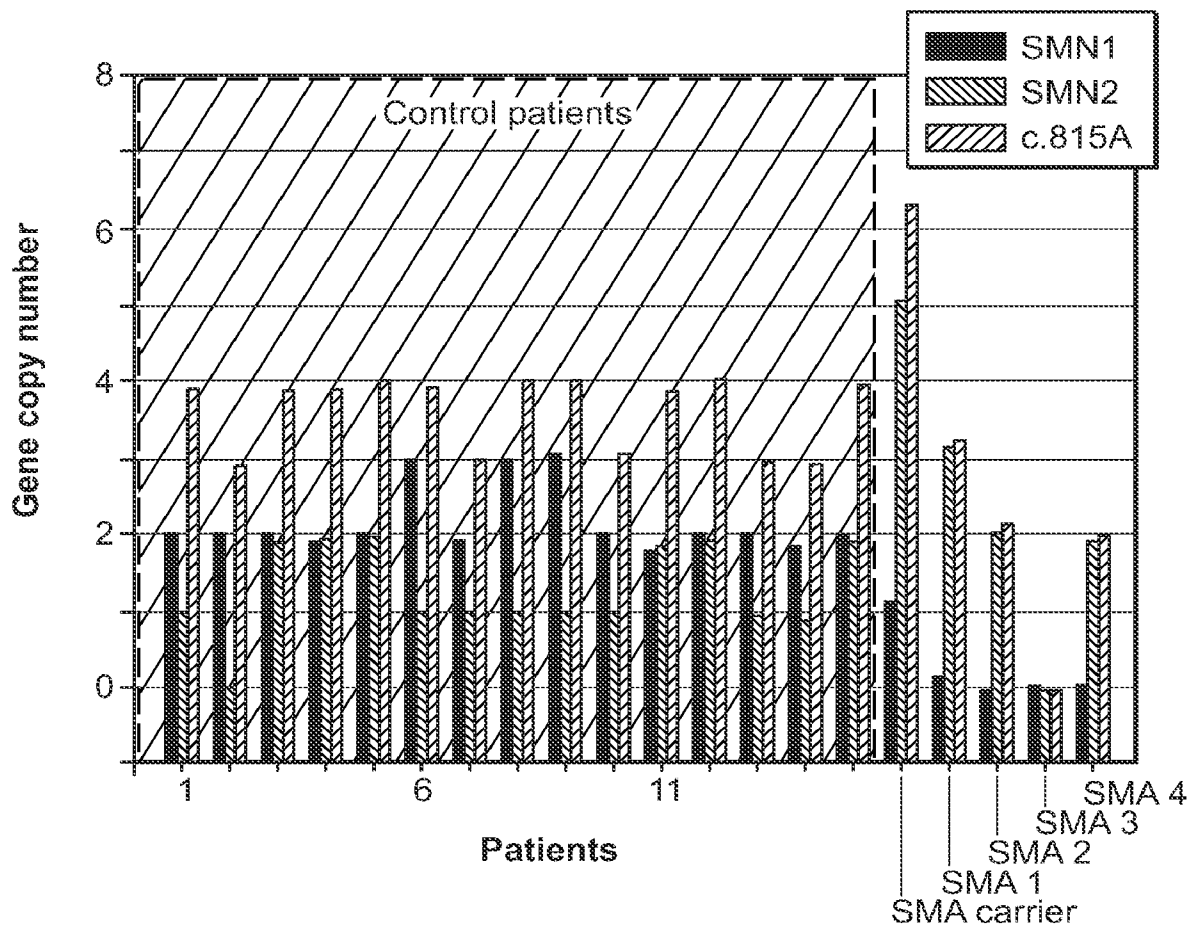


FIG. 14B



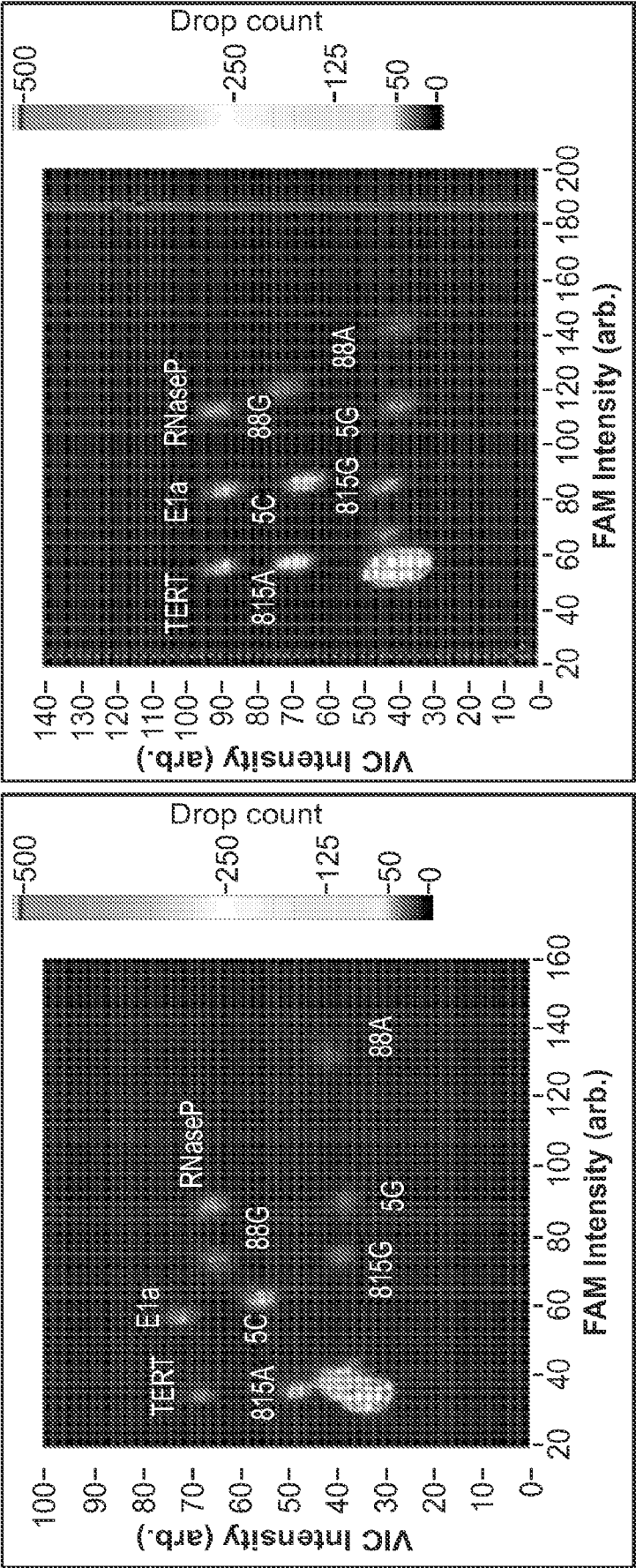


FIG. 15A

3 iterations

FIG. 15B

19/51

FIG. 16

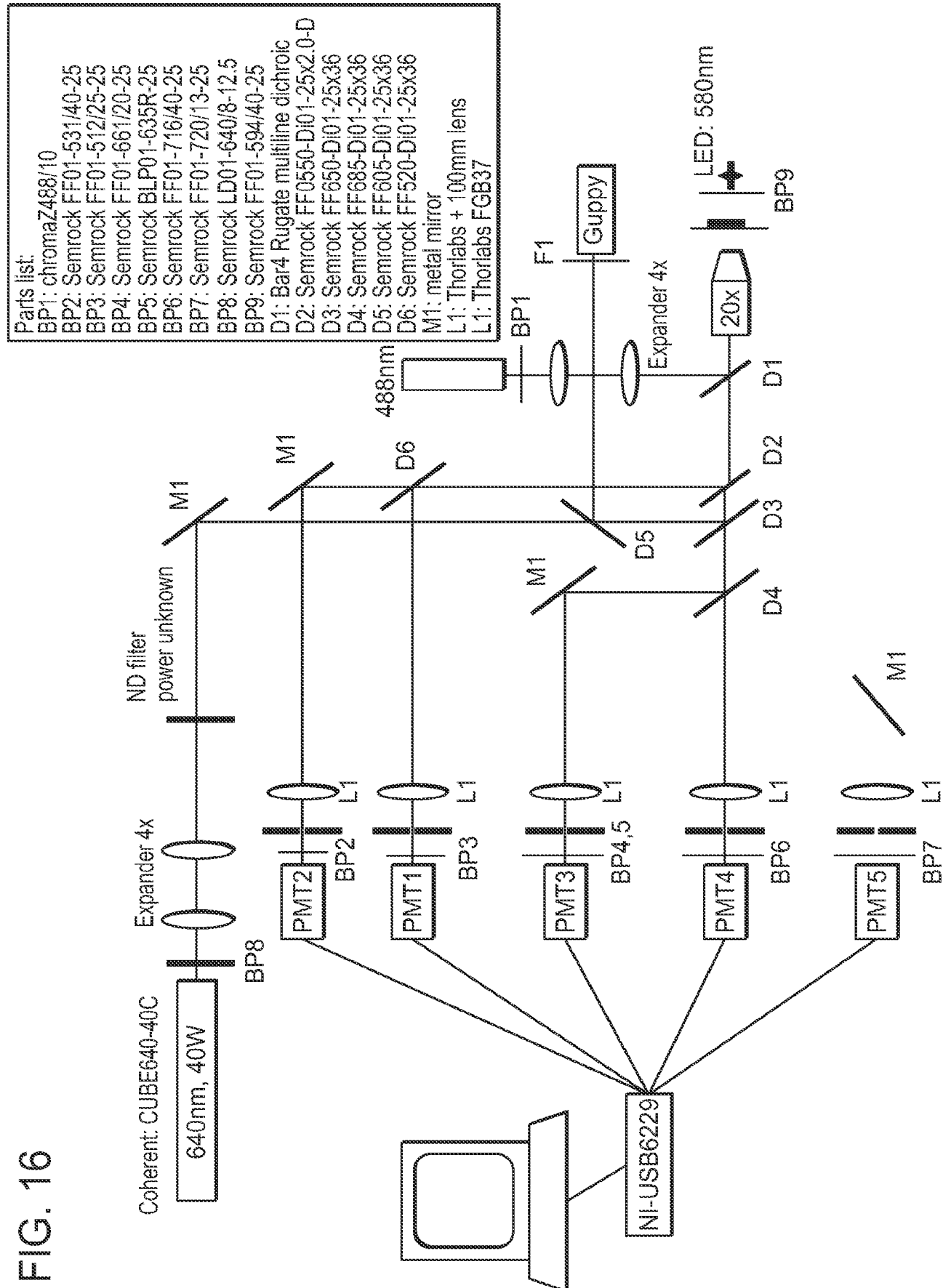


FIG. 17

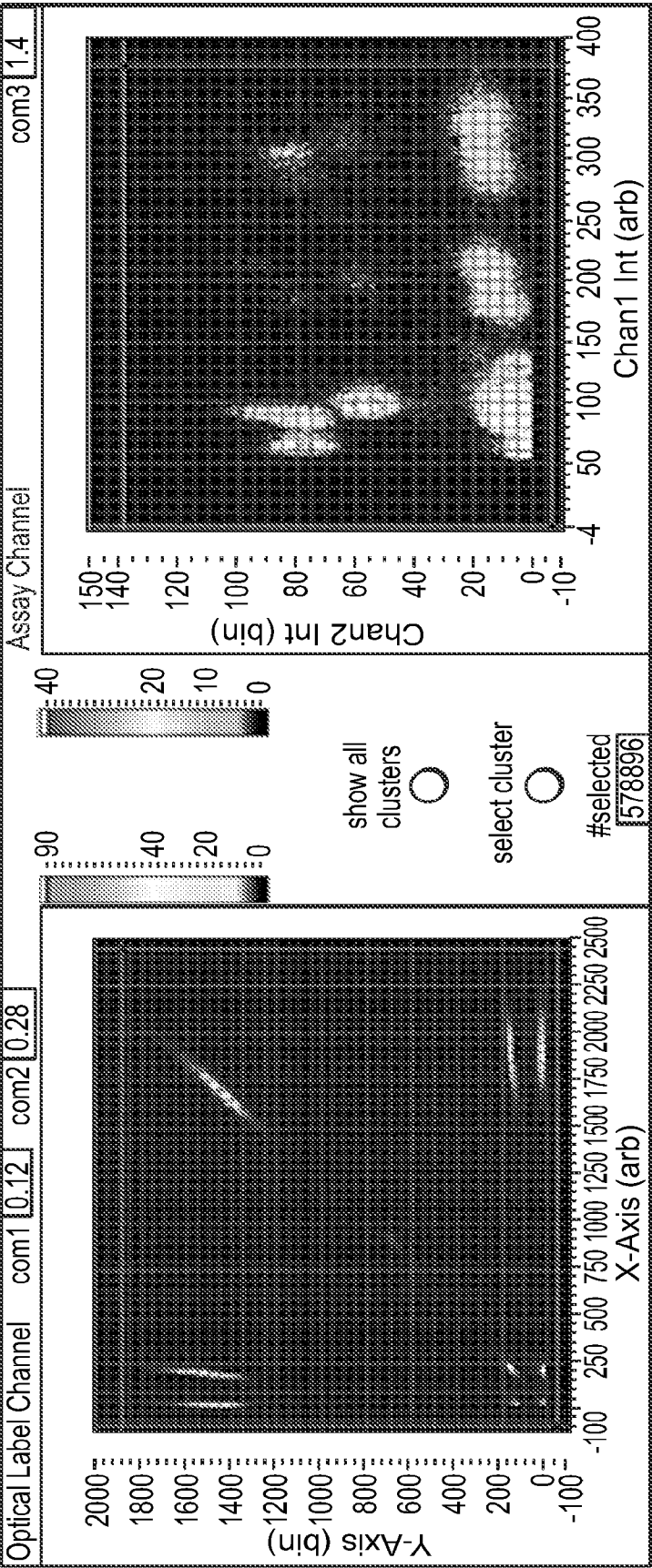


FIG. 18A

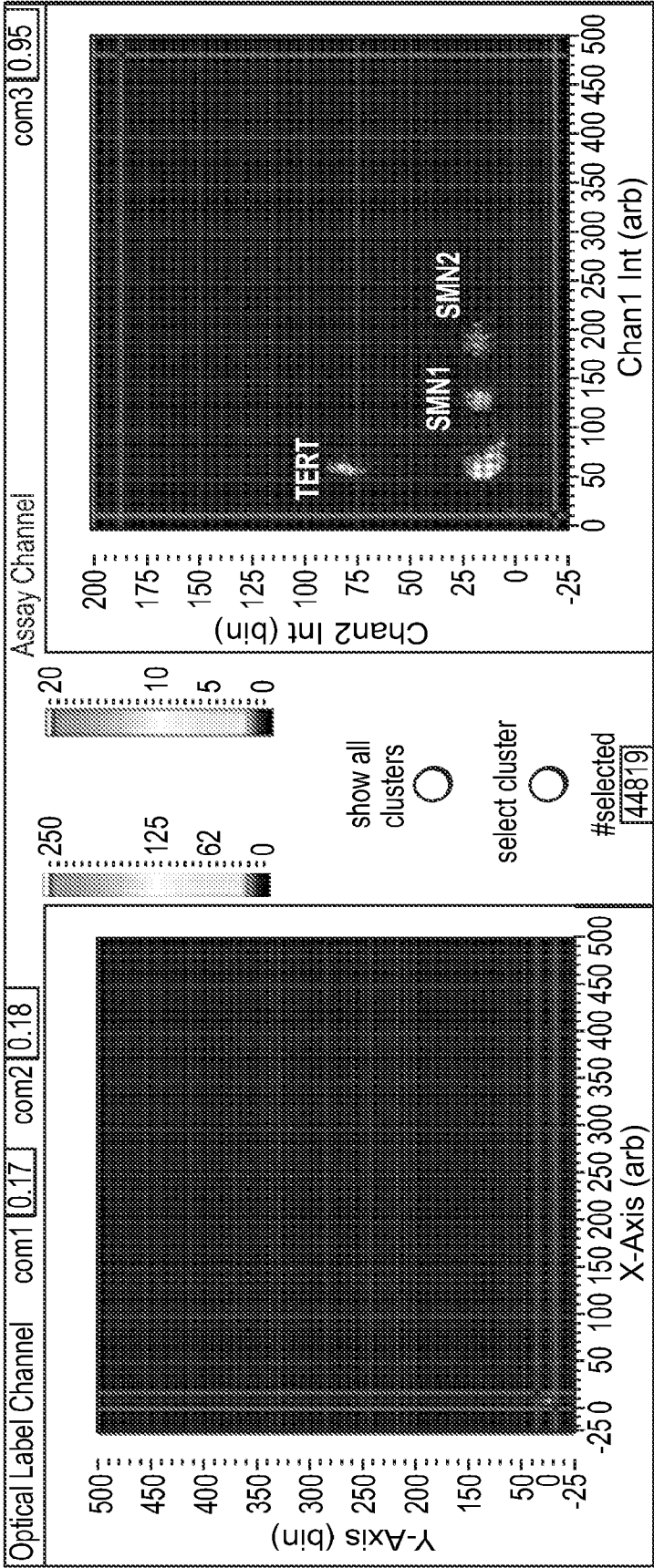


FIG. 18B

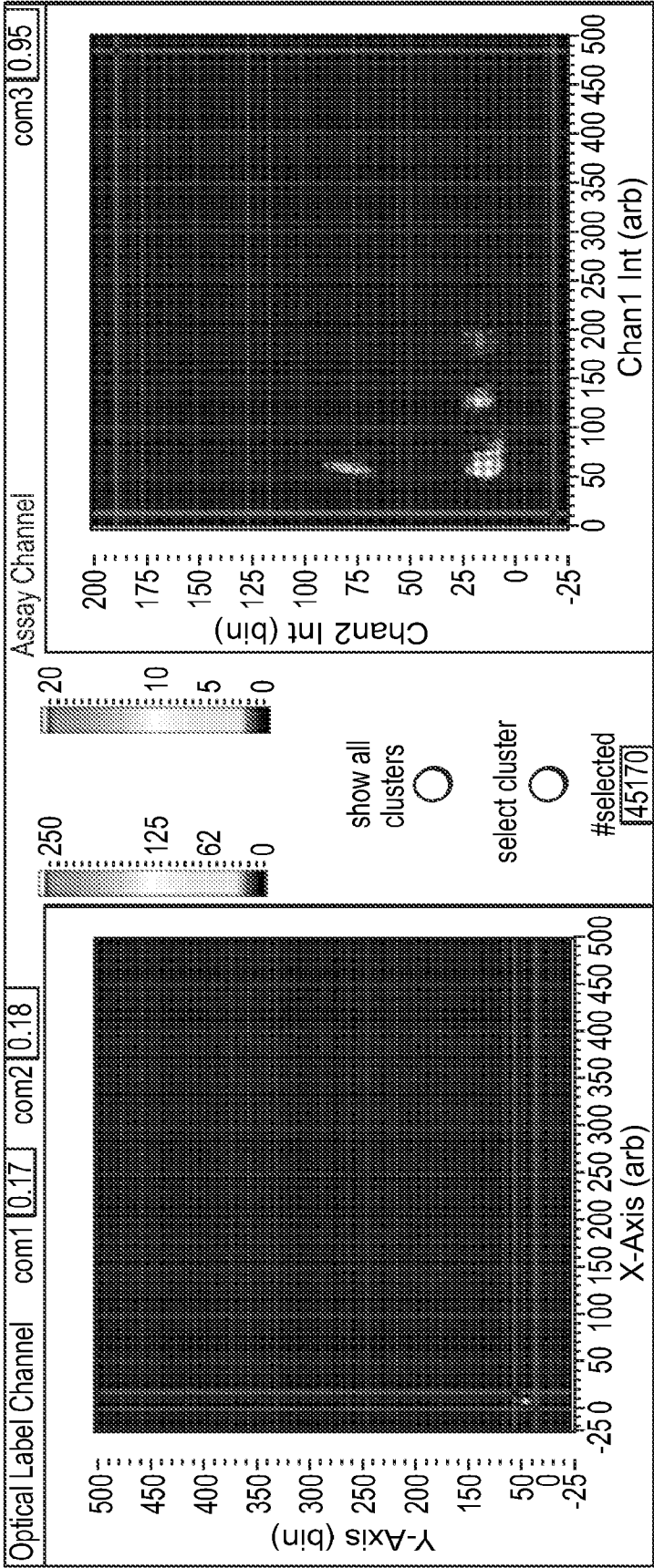


FIG. 18C

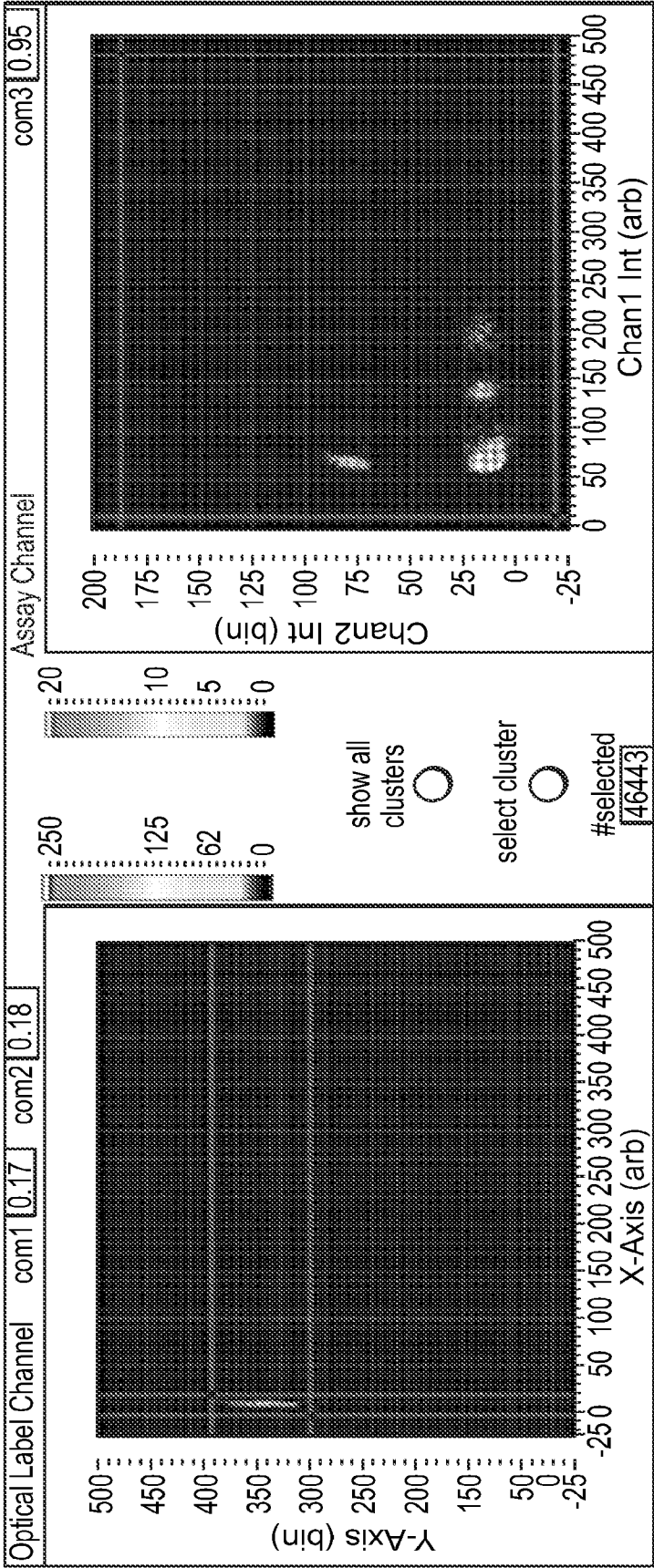


FIG. 19A

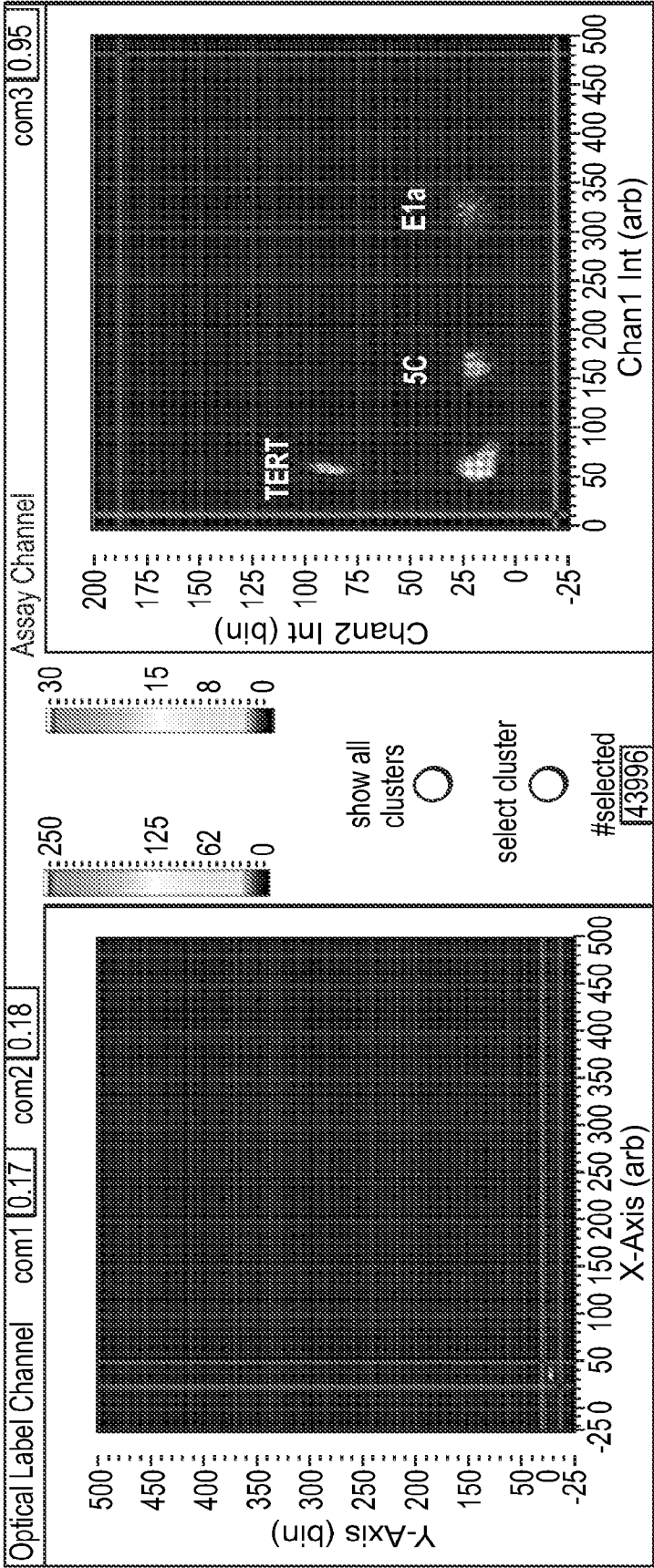


FIG. 19B

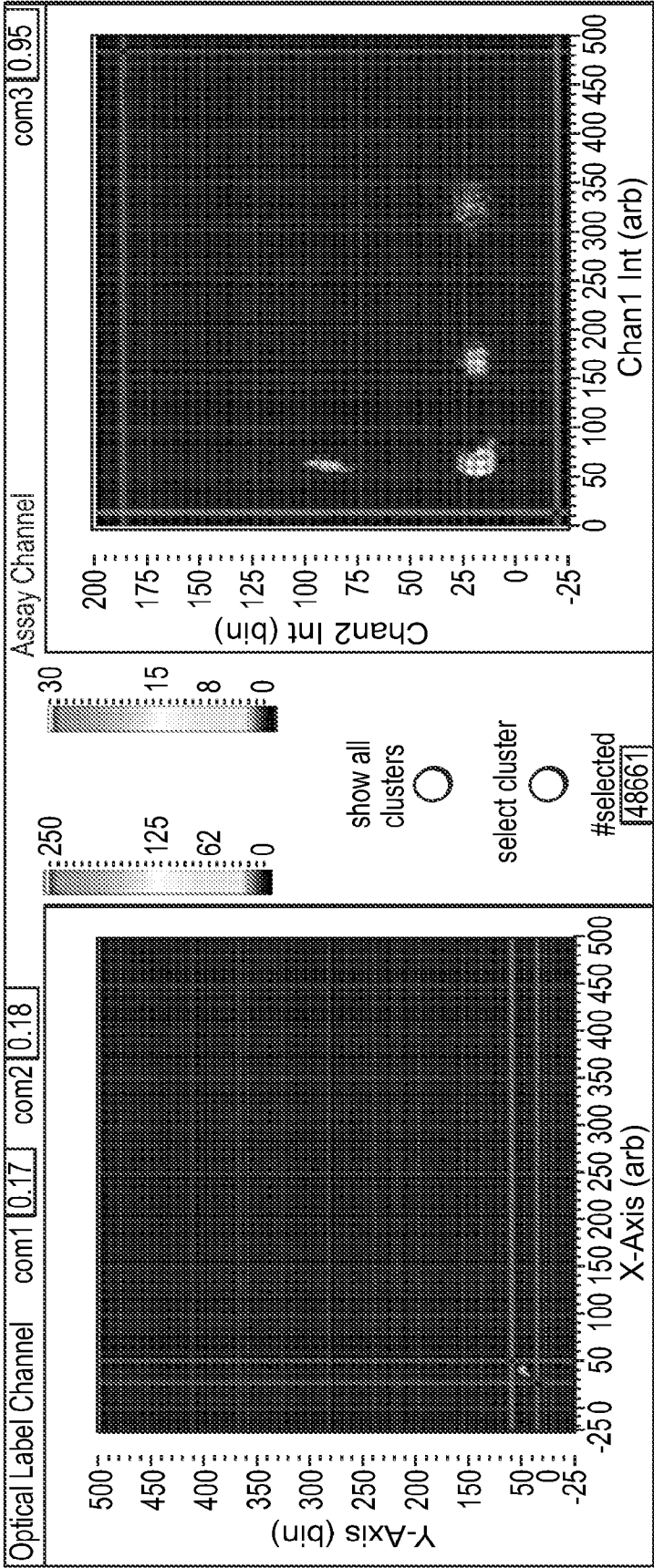


FIG. 19C

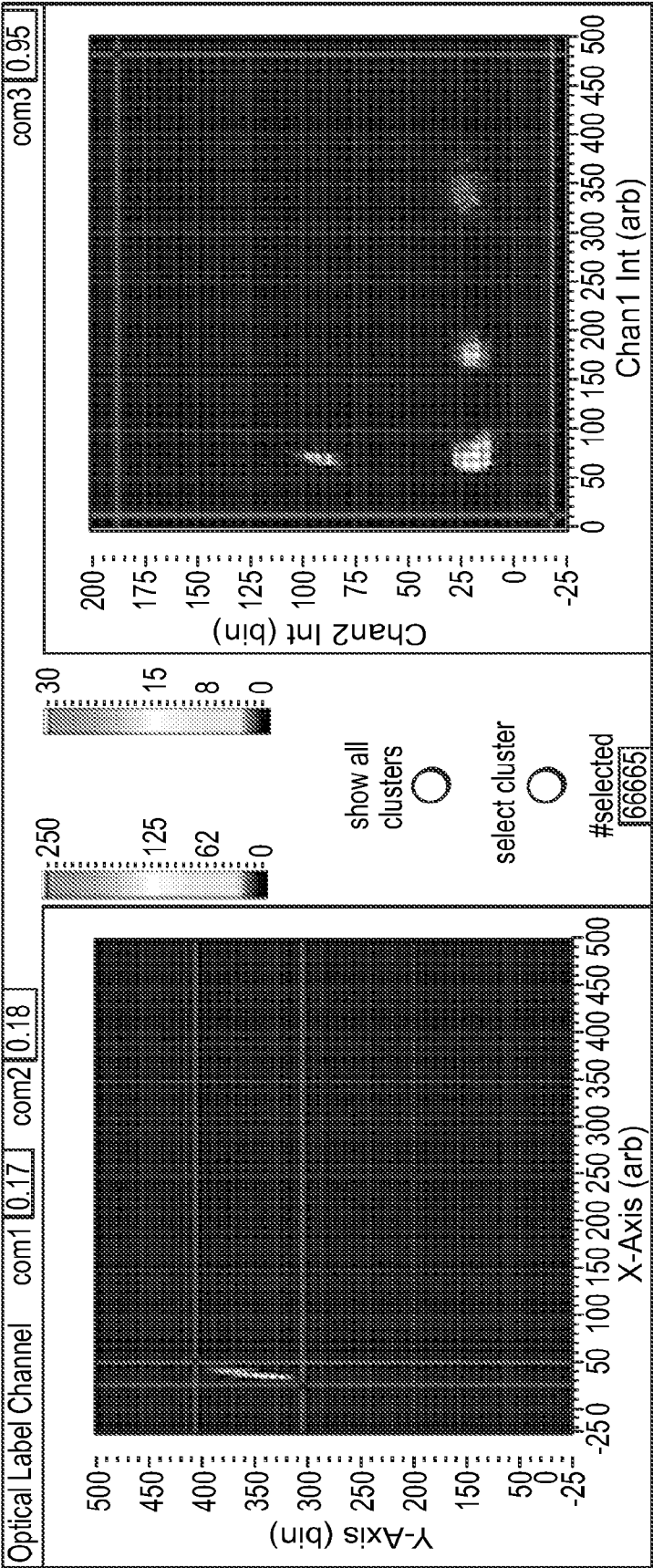


FIG. 20A

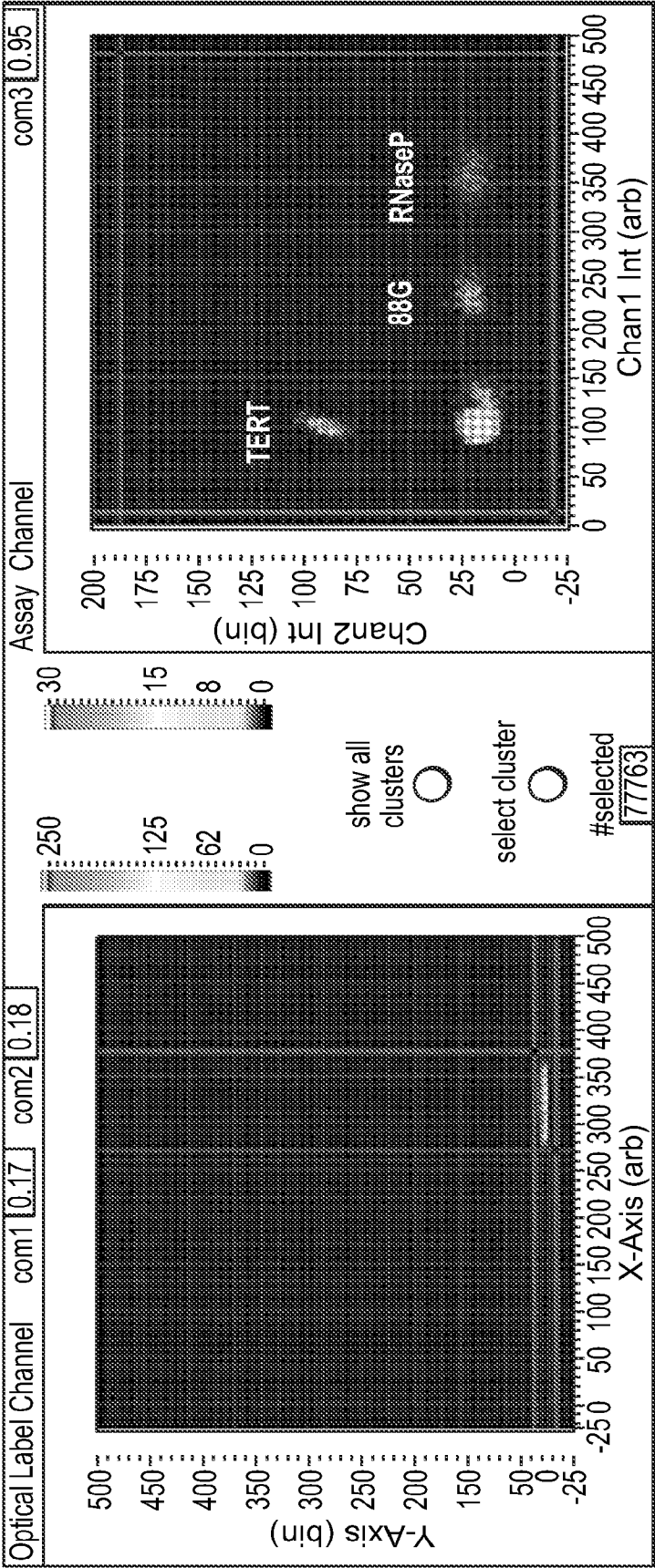


FIG. 20B

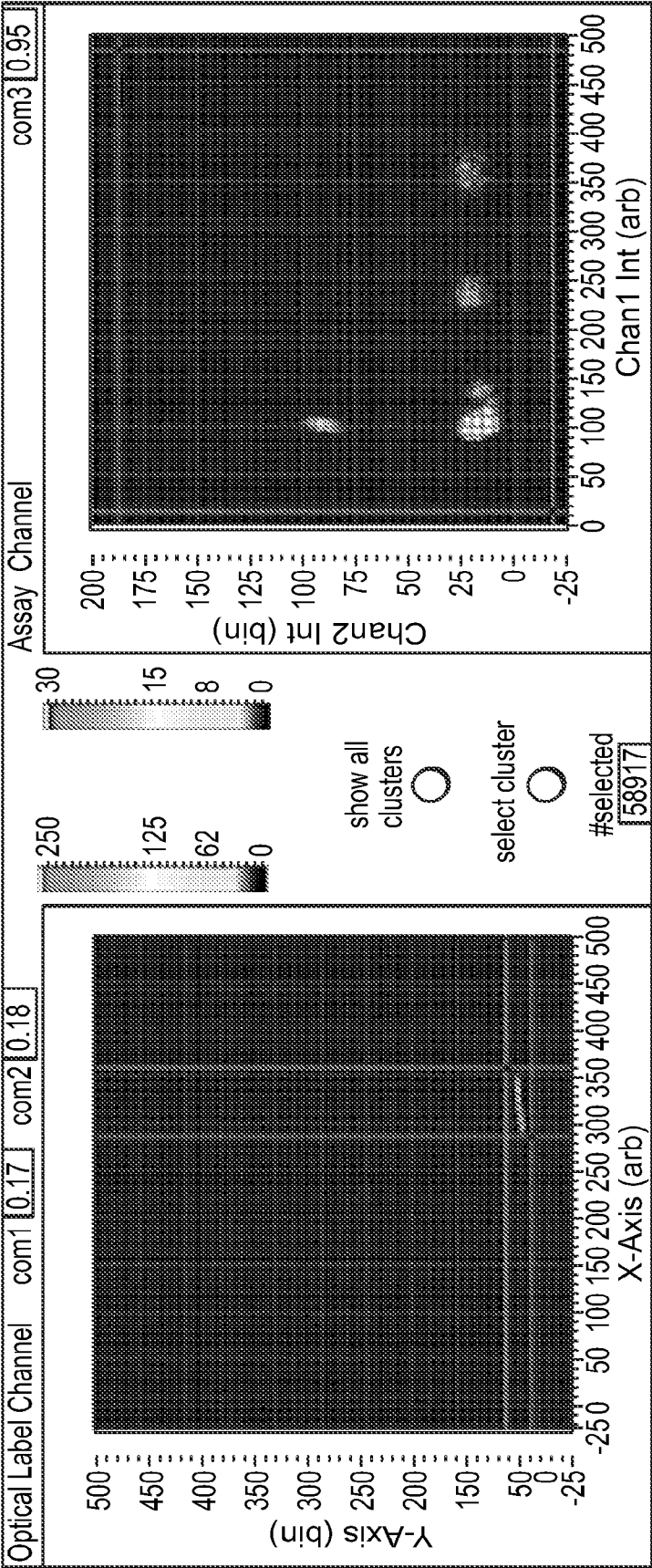


FIG. 20C

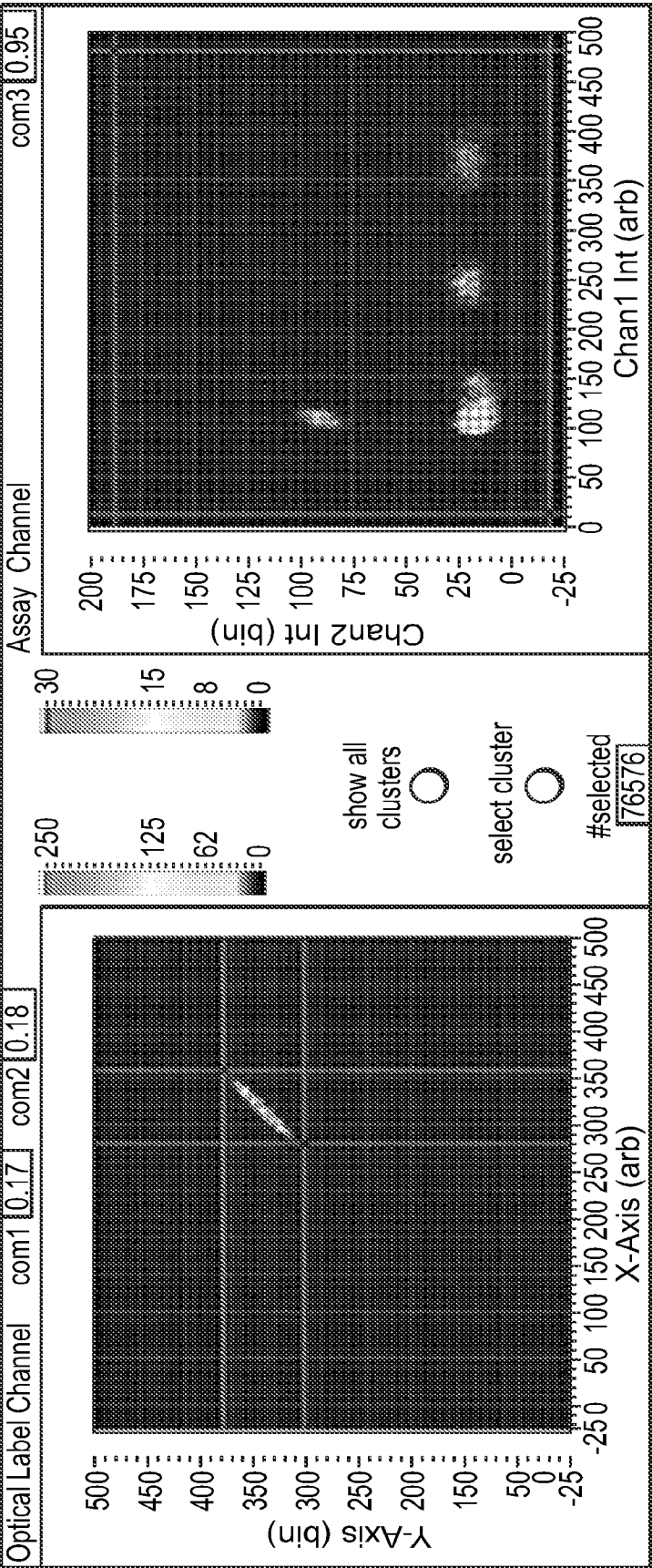


FIG. 21A

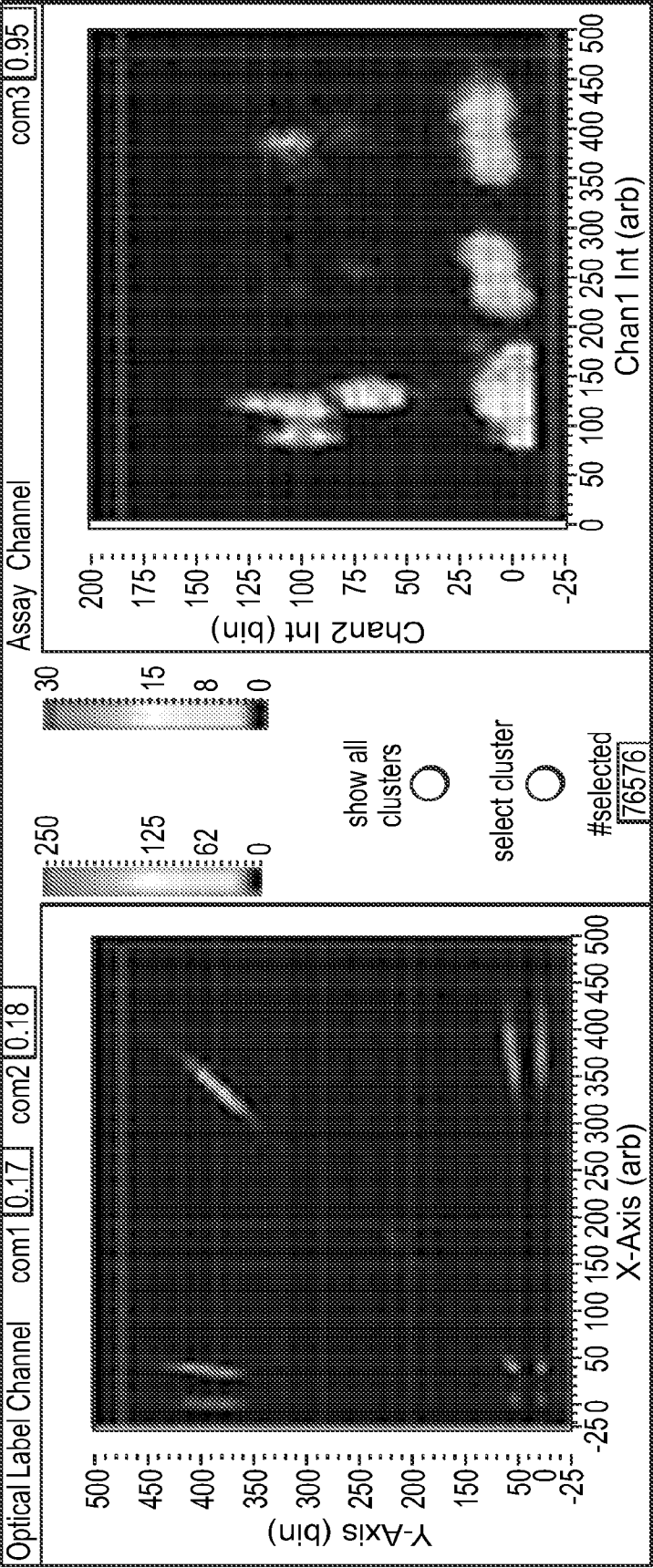


FIG. 21B

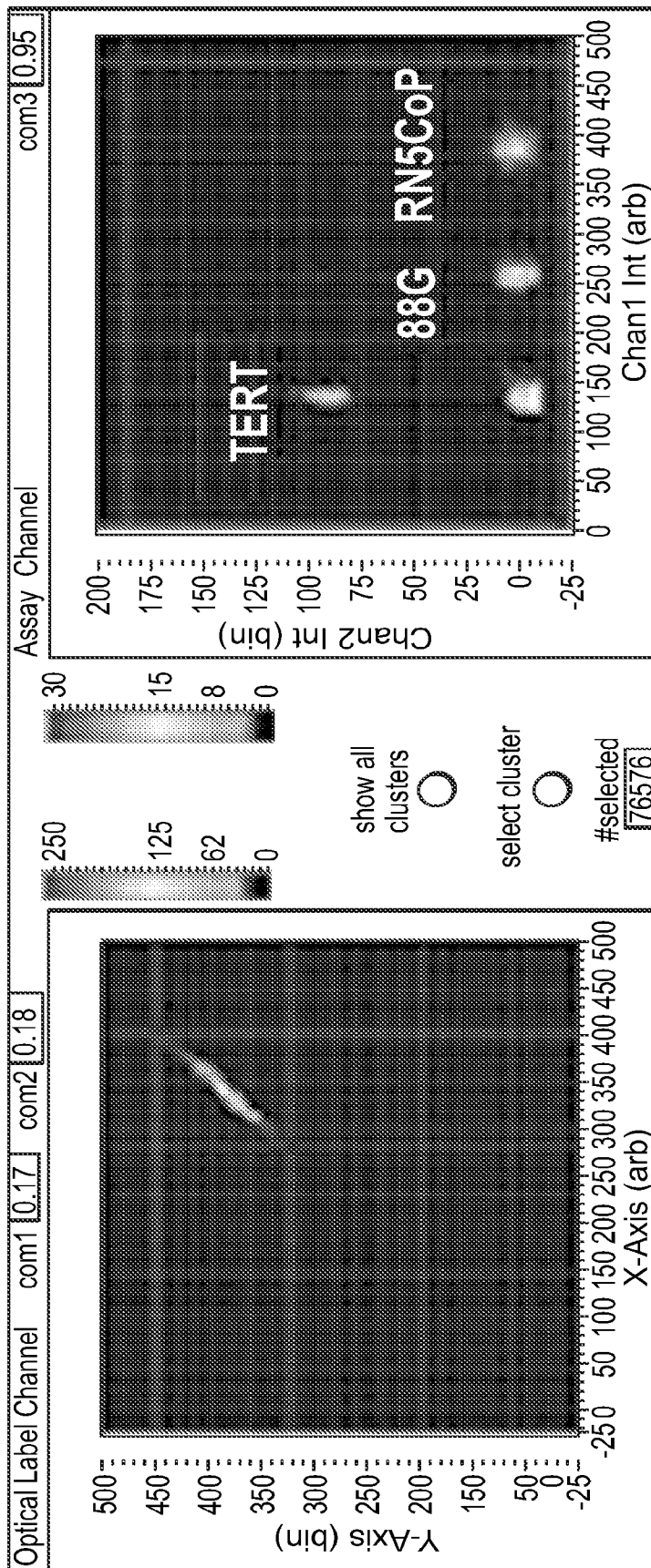


FIG. 21C

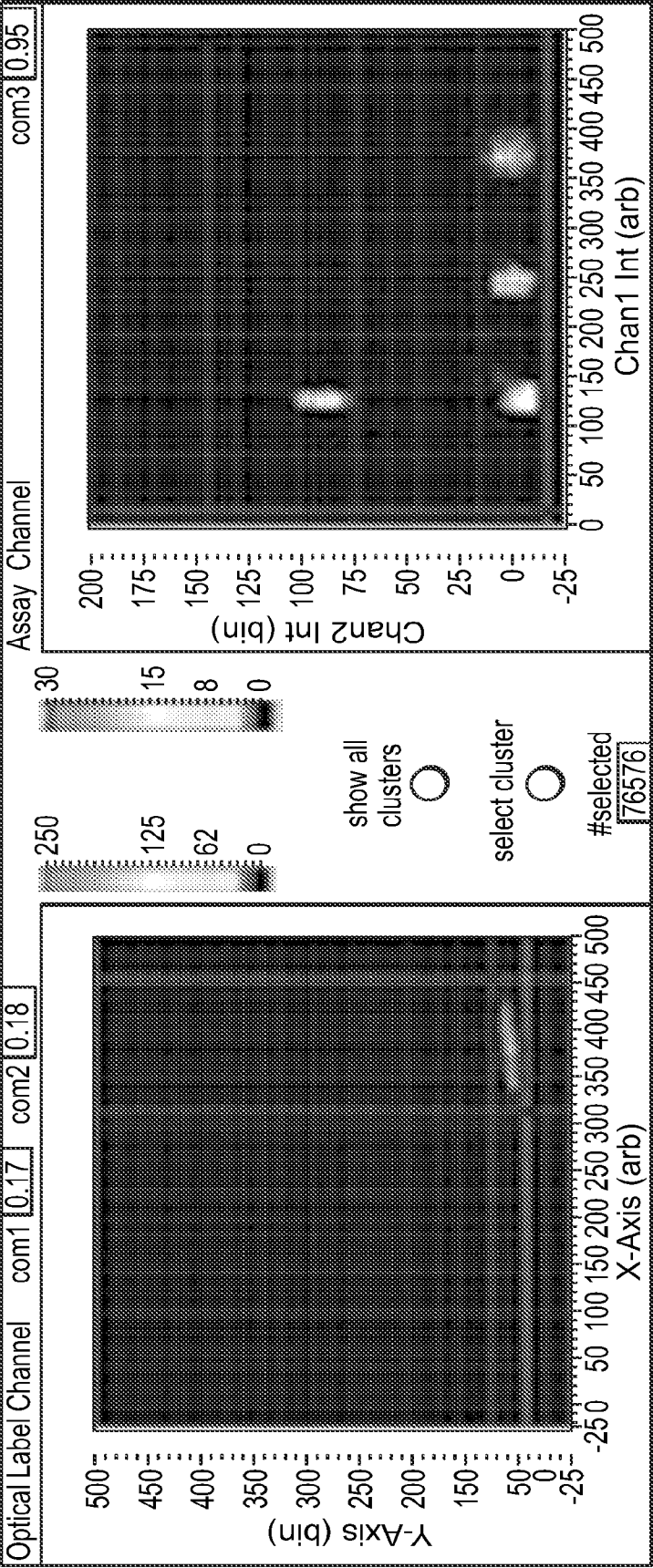


FIG. 21D

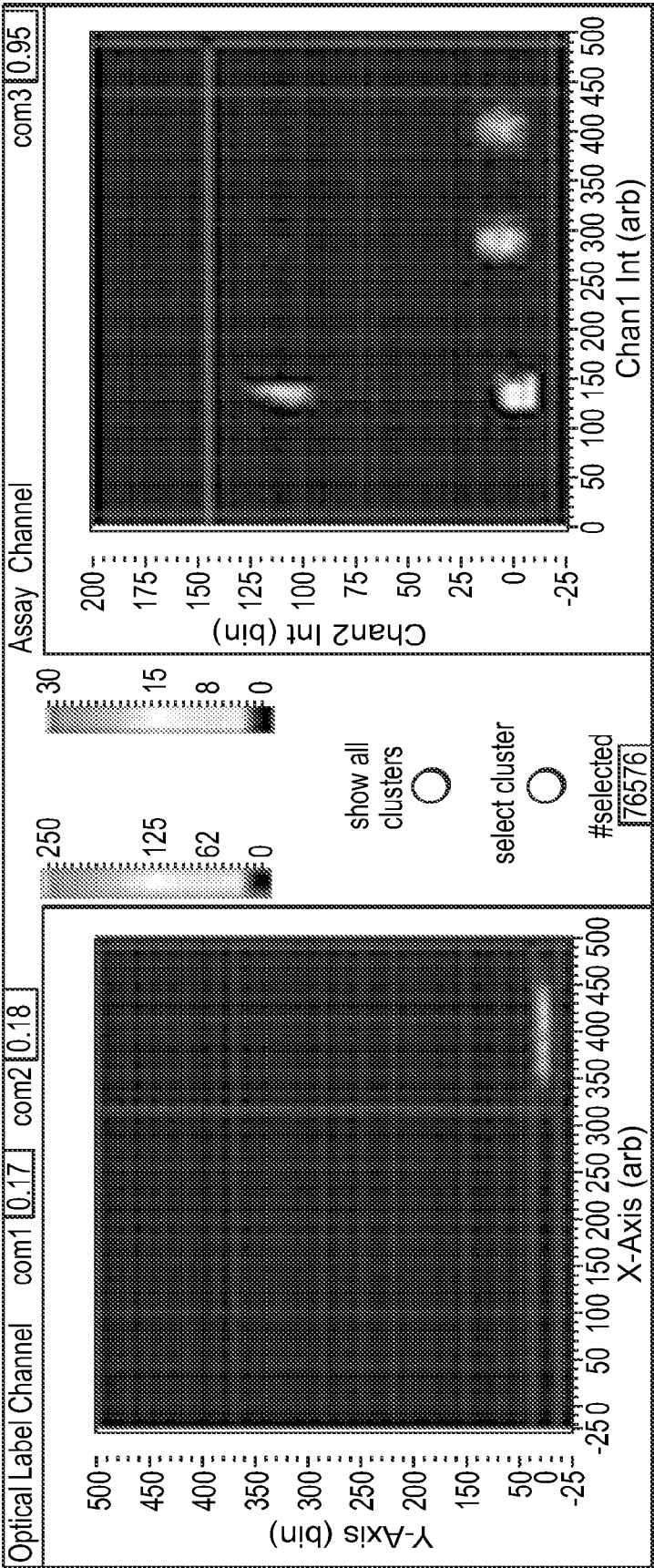


FIG. 21E

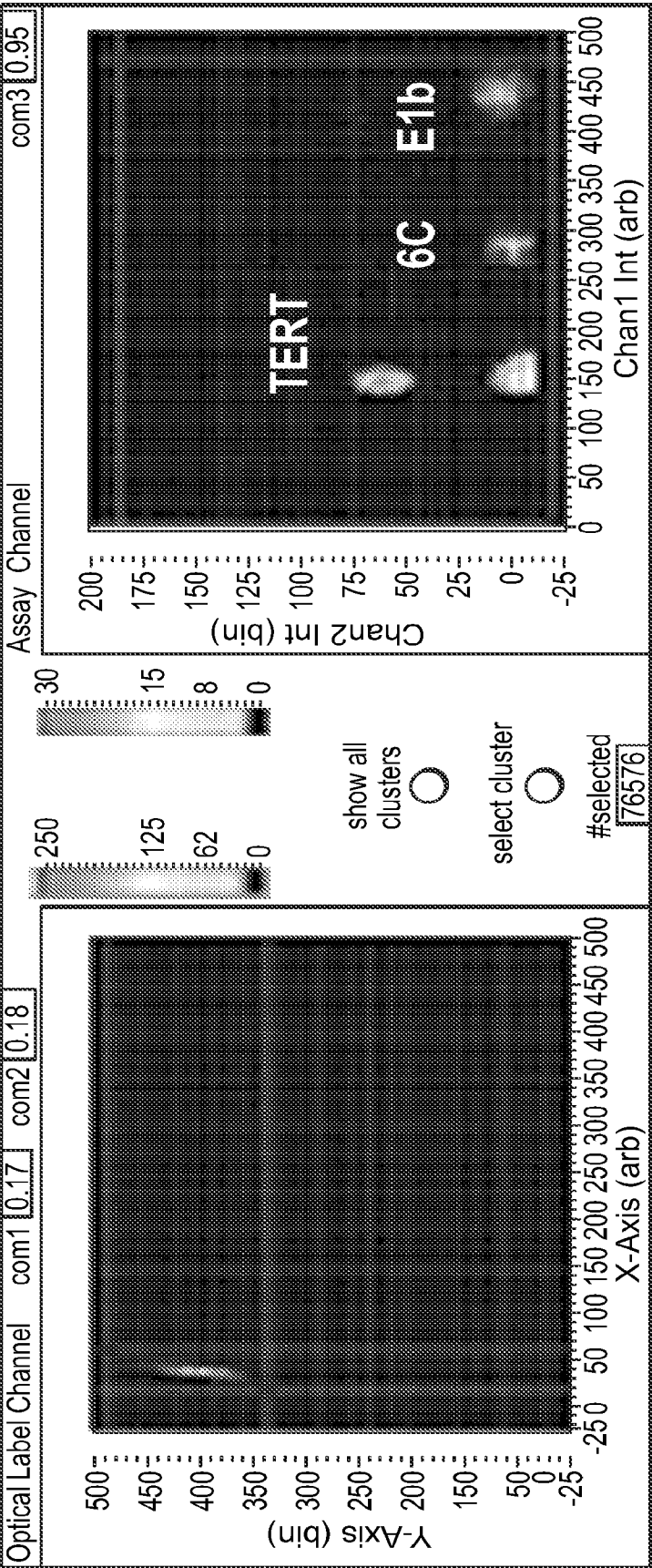


FIG. 21F

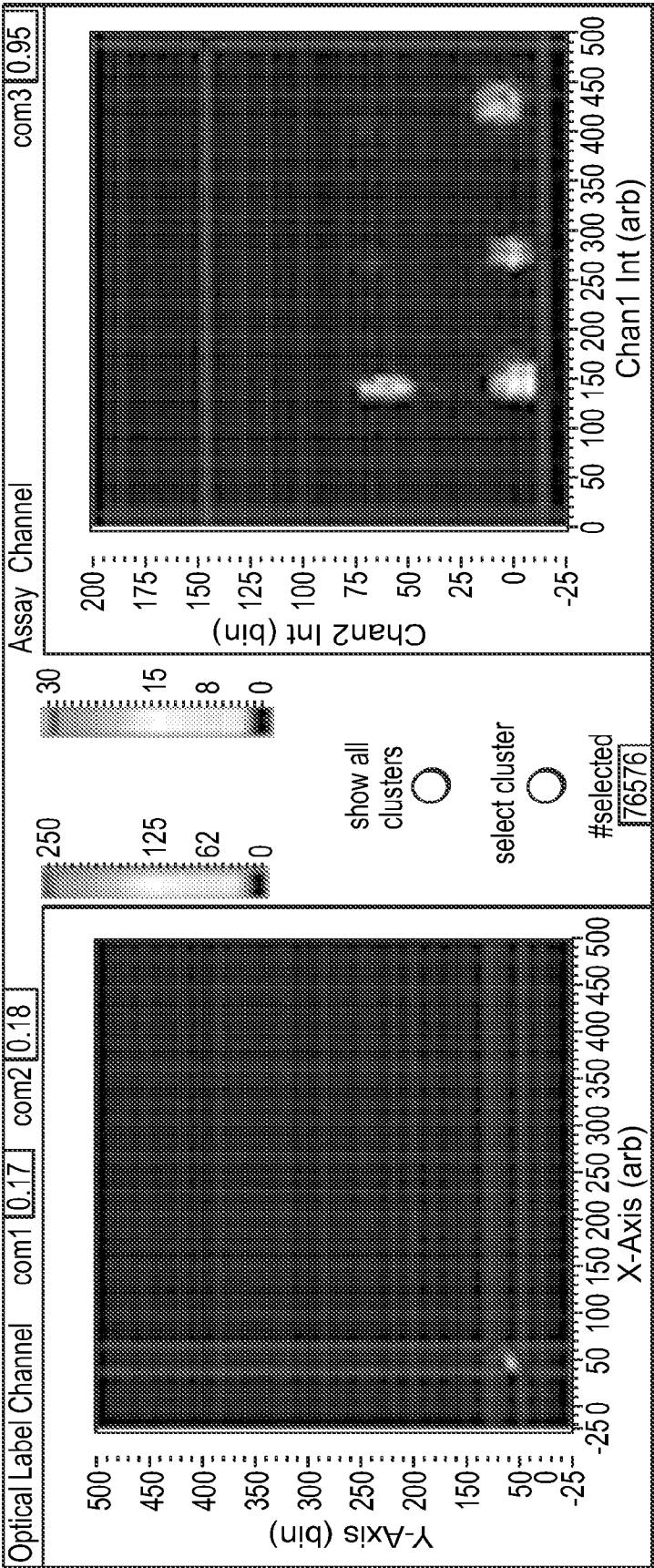


FIG. 21G

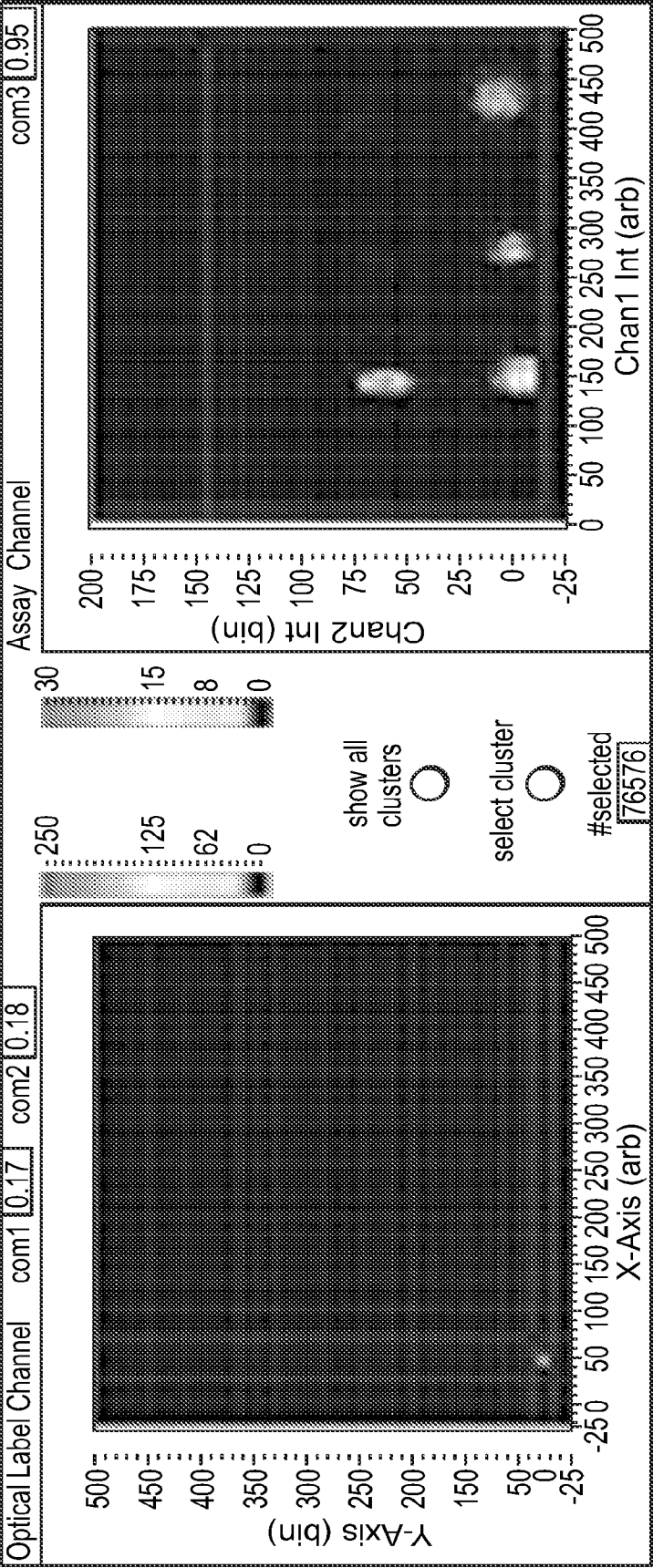


FIG. 21H

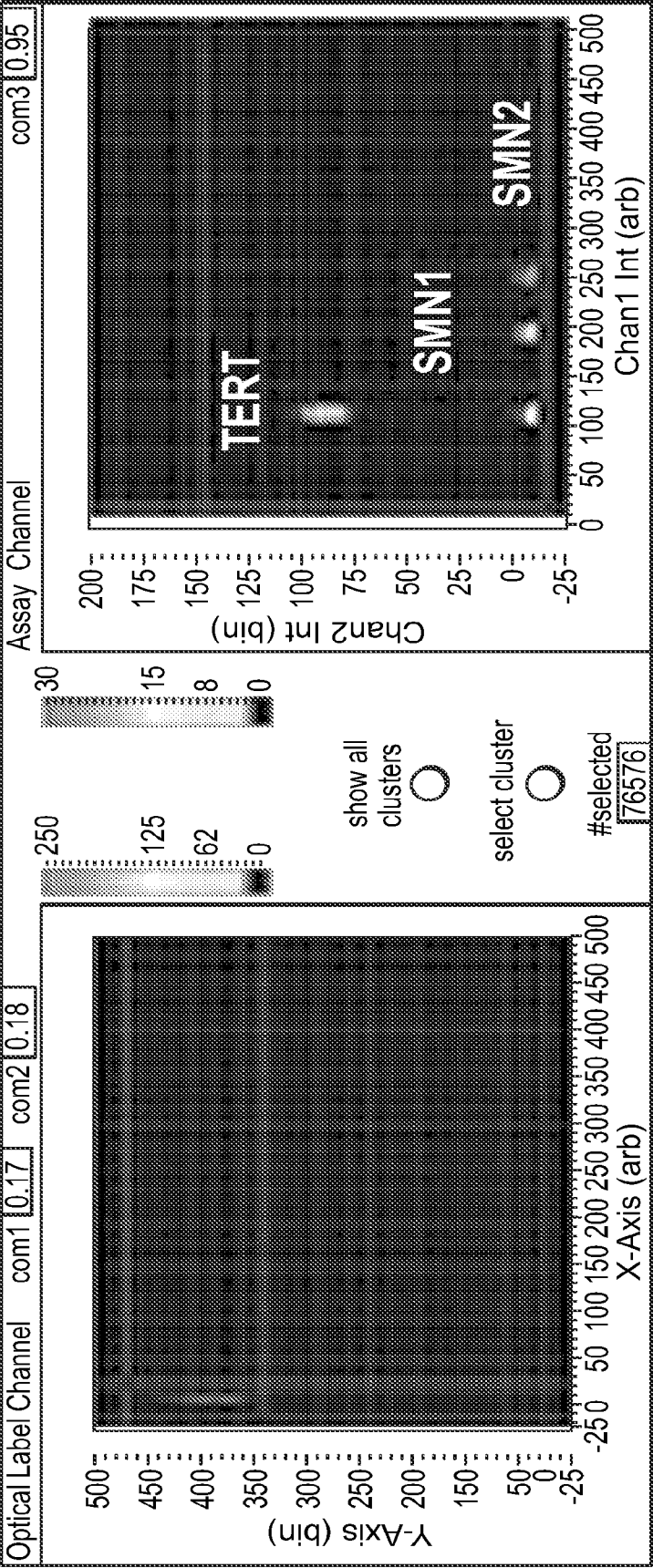


FIG. 21I

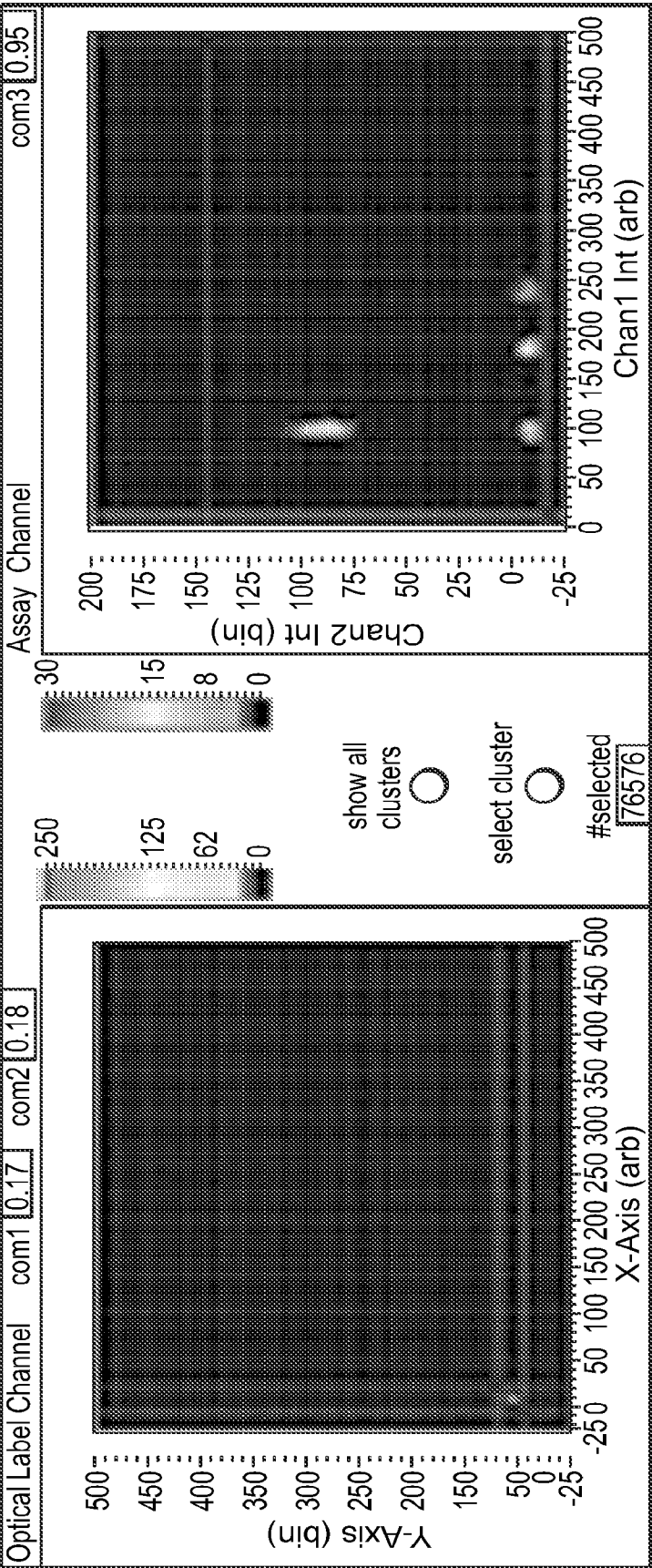
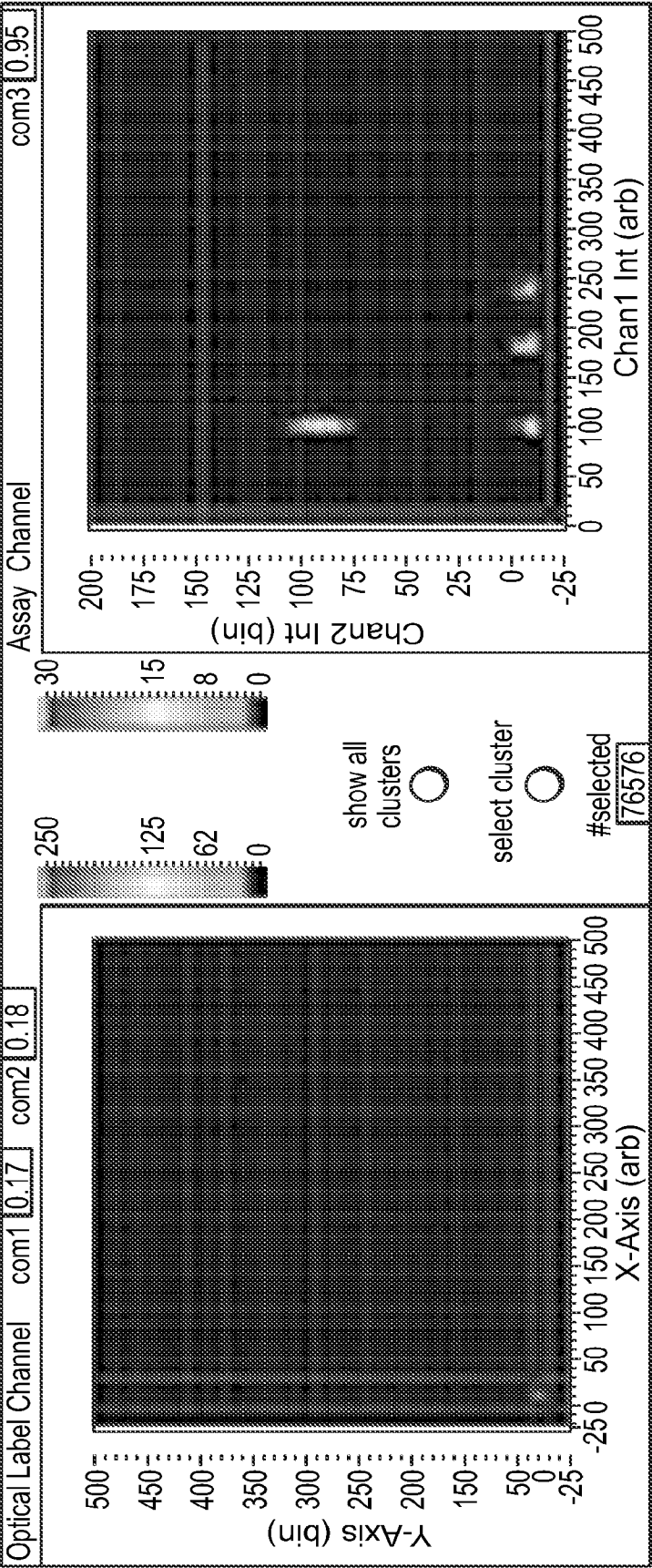


FIG. 21J



40/51

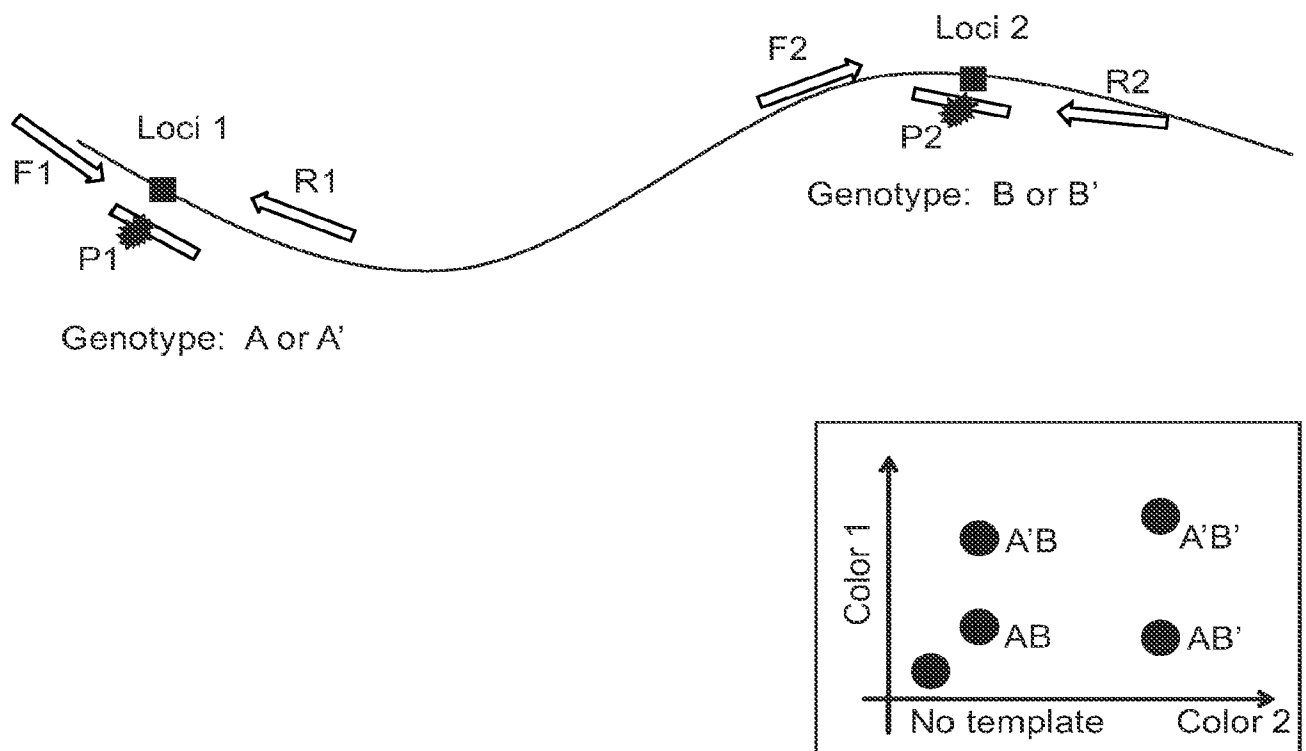
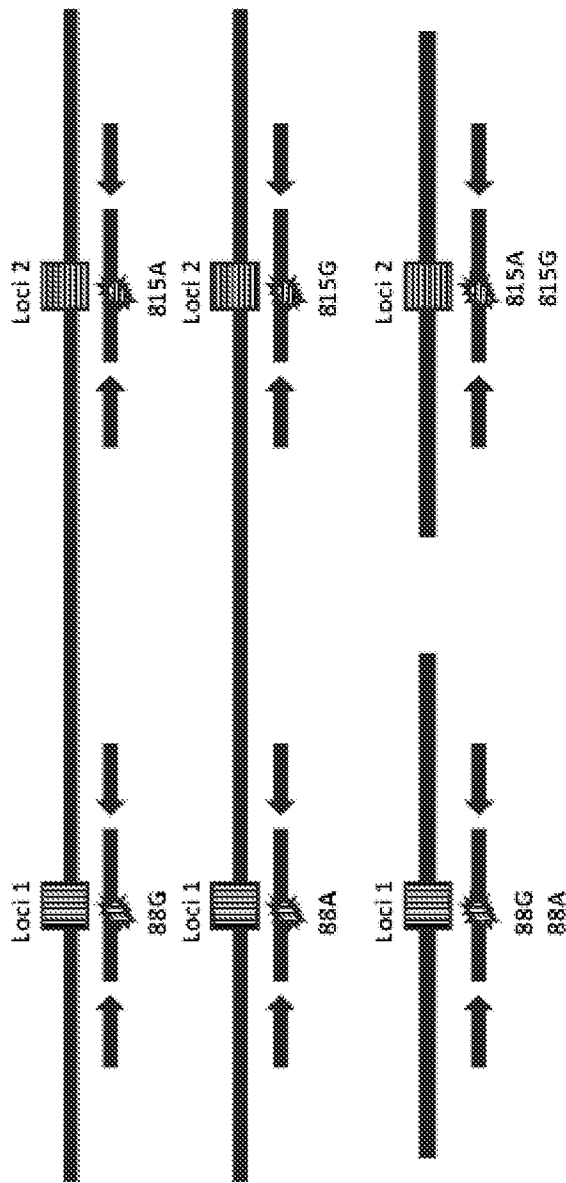


FIG. 22



First Enzyme Digest, Allele 1: 88G, 815A
First Enzyme Digest, Allele 2: 88A, 815G
Second Enzyme Digest, Allele 1: 88G, 815A
Second Enzyme Digest, Allele 2: 88A, 815G

FIG. 23

42/51

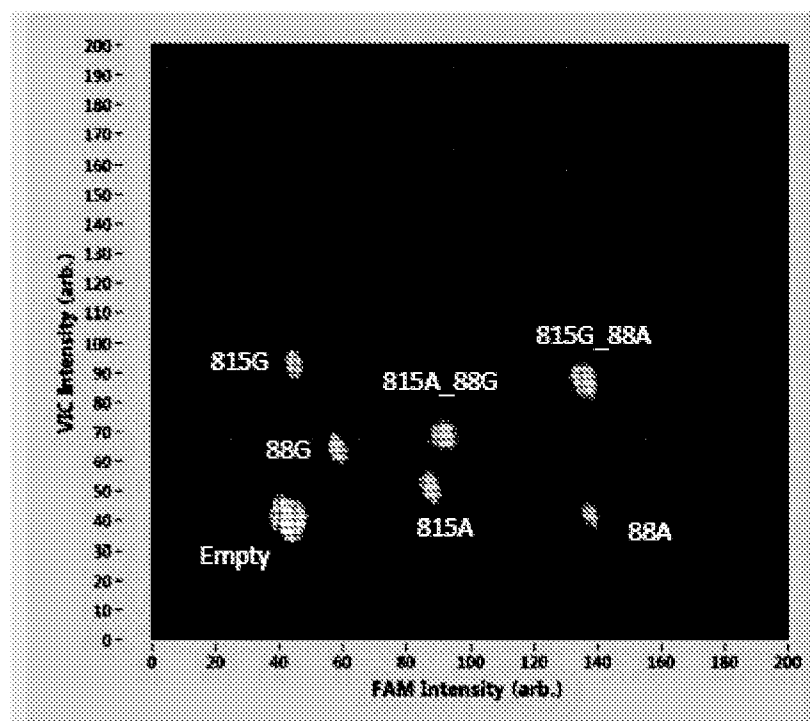


FIG. 24

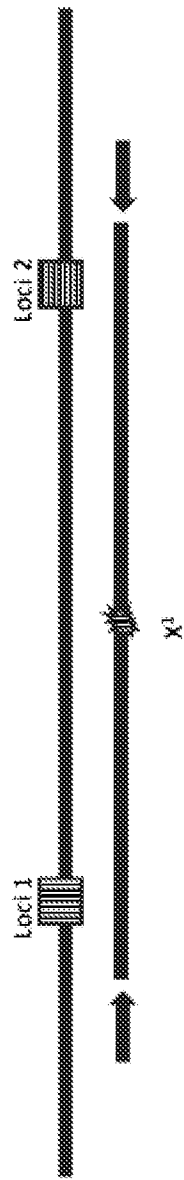


FIG. 25

44/51

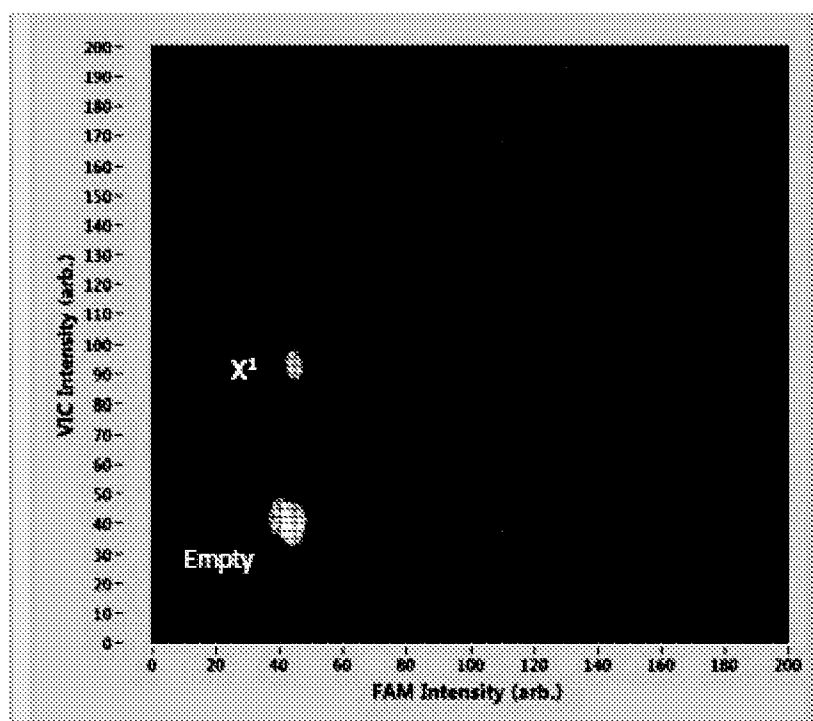


FIG. 26

45/51

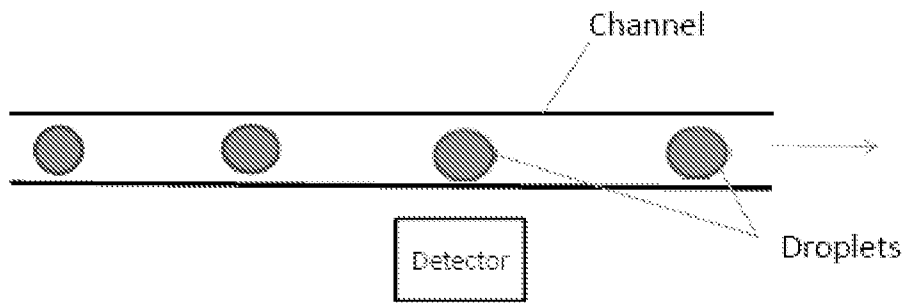


Figure 27A

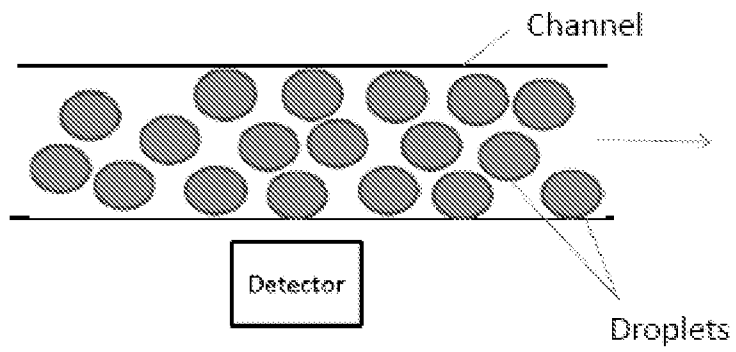


Figure 27B

46/51

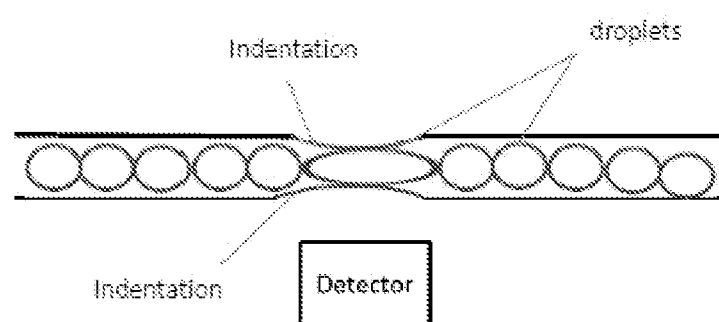


Figure 27C

47/51

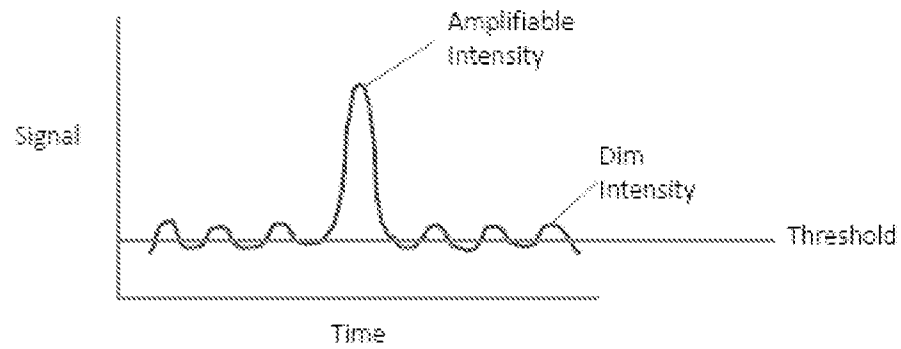


Figure 28A

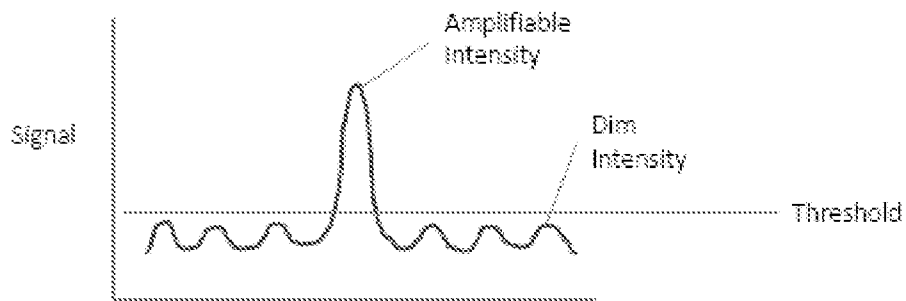


Figure 28B

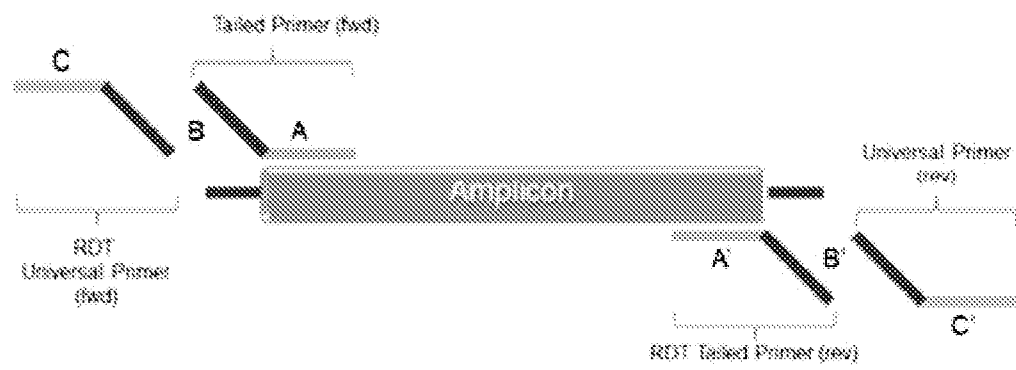


Figure 29

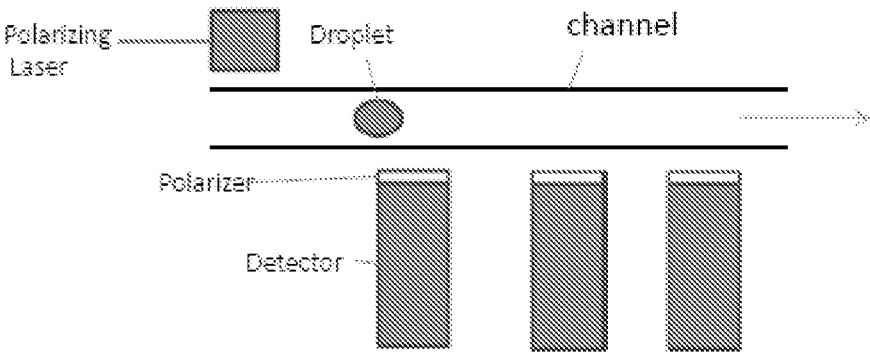


Figure 30A

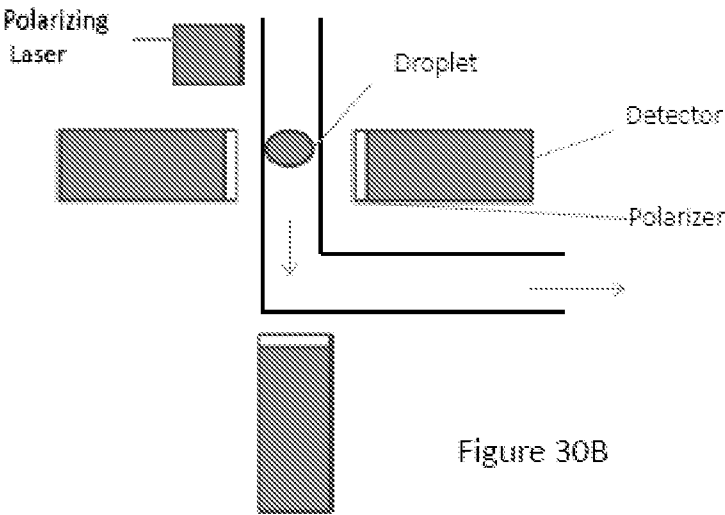


Figure 30B

50/51

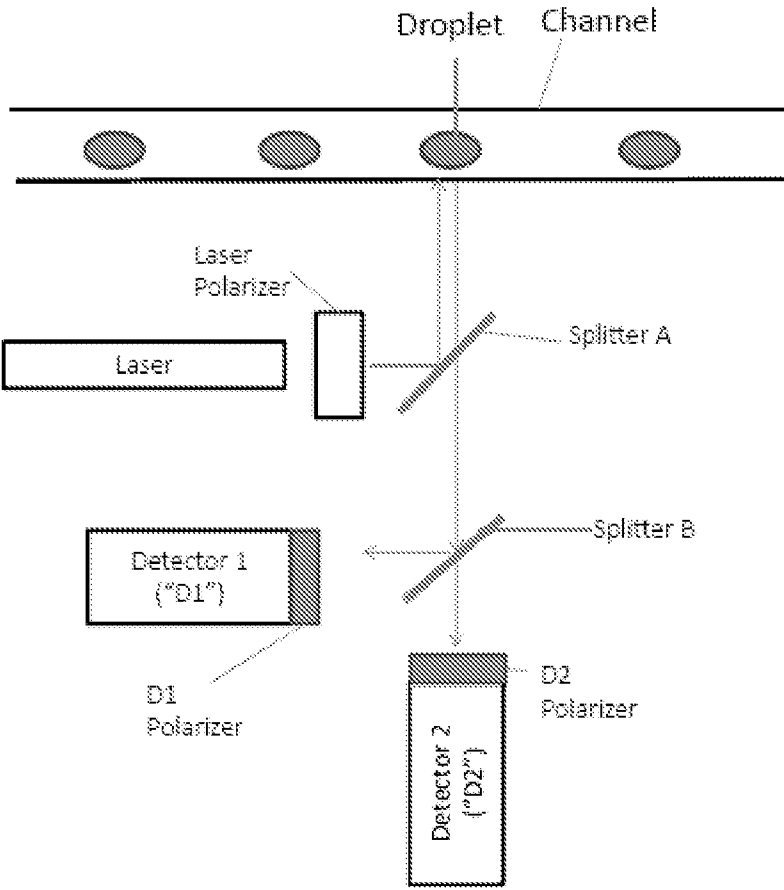


Figure 31

Figure 32

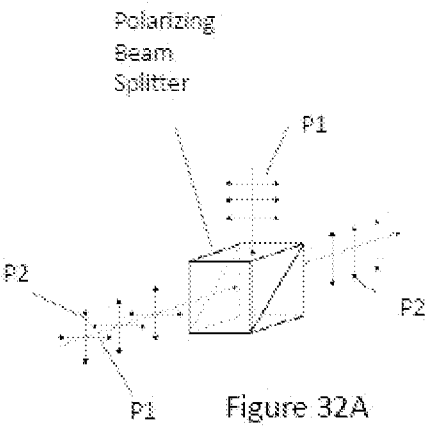
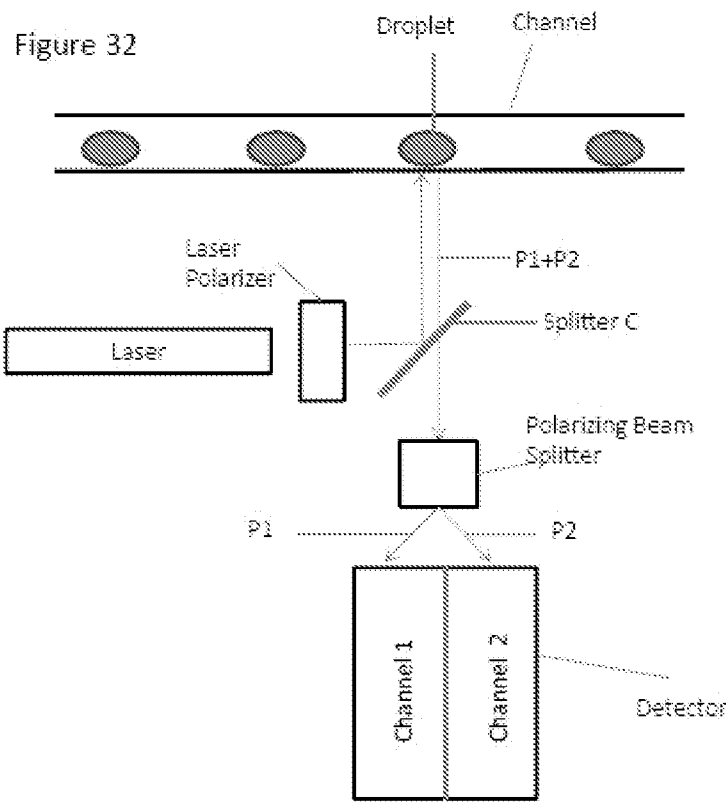


Figure 32A