

chosen for further downstream analysis. Demonstrating and validating total absence of expression (with the highest level of sensitivity, i.e., Digital PCR assays) in normal blood cells is important. In general, we found that only ~10% of candidate genes predicted based on computational models or RNA Seq data are truly negative in human blood samples.

In particular, candidate tumor-specific mRNA transcripts for the detection of CTCs were initially identified through the analysis of gene expression data sets (microarray and RNA-Seq) derived previously for human breast, prostate, lung, pancreas, hepatocellular, and melanoma cancers. Specific publically available data sets used for this analysis include The Cancer Genome Atlas (TCGA) (The Cancer Genome Atlas, available online at tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) and the Cancer Cell Line Encyclopedia (CCLE) (available online at broadinstitute.org/ccle/home; see also, Barretina et al., The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, *Nature* 483:603-607 (2012)). In addition, single-cell RNA-seq gene expression data from CTCs isolated from human patients with breast, prostate, and pancreatic cancers were analyzed (GEO accession numbers GSE51827, GSE60407, and GSE67980) (Aceto et al., Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis, *Cell*, 158:1110-1122 (2014); Ting et al., Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells, *Cell Rep*, 8:1905-1918 (2014); and Miyamoto et al., RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance, *Science* 349:1351-1356 (2015). Tumor specific transcripts identified through these databases were then compared to human leukocyte RNA-Seq gene expression data (GEO accession numbers GSE30811, GSE24759, GSE51808, GSE48060, GSE54514, and GSE67980). Transcripts that displayed significant differential expression, with high expression in tumors and low or undetectable expression in leukocytes, were then selected for further downstream analysis. Moreover, a literature search was performed to select additional candidate tumor-specific transcripts. Between 50 and 100 candidate genes were selected for each type of human cancer.

For each candidate gene within each specific cancer type, two to four sets of PCR primers were designed to span regions across the target transcript. Primers are

synthesized by IDT (Integrated DNA Technologies), probes are labeled with FAM or HEX, ZEN, and IABkFQ to create a probe targeting the middle of the amplicon.

Unique features of our PCR primer design methodology necessary for the successful application of digital PCR-based mRNA transcript detection in human CTCs include the following: 1) the specific targeting of the 3' end of each mRNA transcript, given the proclivity of cellular mRNA transcripts to degrade from the 5'-end, particularly in unfixed, fragile cells such as CTCs; 2) the design of primers to generate amplicons that span introns in order to exclude the unintentional amplification of contaminating genomic DNA, for example from excess contaminating leukocytes in the enriched CTC mixture; and 3) the design of primers to inclusively amplify multiple splice variants of a given gene, given the uncertainty in some cases regarding the clinical relevance of specific splice variants.

The specificity of the primers was first tested by qRT-PCR using cDNA derived from cancer cell lines (representing breast, prostate, lung, pancreas, and liver cancers and melanoma). For each type of human cancer, 2 to 5 established cancer cell lines were cultured and used for initial testing to evaluate PCR primer performance and assess for expression of the target transcript in the specified cancer. To provide an initial test of specificity, the same primers were used to evaluate expression of the target transcript in leukocytes from healthy individuals who do not have a diagnosis of cancer. Leukocytes from a minimum of five different healthy individuals were tested in this phase of testing (mixture of male and female individuals – this was dependent on the type of cancer; i.e. candidate prostate cancer and breast cancer genes required the use of male or female healthy donors only, respectively).

Leukocytes from healthy individuals were isolated from whole blood using Cell Preparation Tubes with Sodium Heparin (CPT) (Becton, Dickinson, and Co., NJ) following product insert instructions. RNA extraction and first-strand cDNA synthesis was performed for cancer cell lines and isolated leukocytes using standard methods. The specificity of expression of each gene (using 2 to 4 distinct sets of primers for each gene) was tested using qRT-PCR (cell line cDNA as positive controls, leukocyte cDNA from healthy donors as negative controls, and water as an additional negative control). Transcripts present in cancer cell lines, but absent in leukocytes based on qRT-PCR testing were then selected for further validation by

droplet digital PCR. The selection criteria to pass this stage of testing were highly stringent, and required qRT-PCR signal to be present in at least one cancer cell line and absent in all healthy donor leukocyte samples tested.

5 Target transcripts and specific primer pairs that passed the qRT-PCR stage of testing were further validated using droplet digital PCR. For this stage of testing, the CTC-iChip (see, e.g., Ozkumur et al., "Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells," *Sci Transl Med*, 5, 179ra147 (2013) was used to process whole blood samples donated by healthy individuals. The CTC-iChip performs negative depletion of red blood cells, platelets, and leukocytes
10 from whole blood, and generates a sample product that is enriched for cells in the blood that do not express leukocyte markers, including CTCs (which should not be present in healthy individuals). For each blood sample, the product from the CTC-iChip was supplemented with an RNA stabilization solution (RNAlater®, Life Technologies) and processed for RNA extraction and cDNA synthesis using standard
15 methods. Droplet digital PCR (Biorad, CA) was then used to quantitate the number of transcripts present in each sample based on the specific primer pairs being tested. Samples assessed by droplet digital PCR during this phase of testing included cDNA from cancer cell lines, leukocyte cDNA from healthy donors processed through the CTC-iChip (at least four healthy individuals per primer pair being tested), and water
20 as a negative control.

Criteria for passing droplet digital PCR testing were stringent, and included:
1) the presence of transcript signal in cancer cell lines (at least one cell line with > 10 positive droplets); 2) excellent signal-to-noise ratio represented by separation of
signal between positive and negative (empty) droplets; 3) minimal or absent droplet
25 signal in healthy donors (< 3 droplets per healthy donor); and 4) absent droplet signal in water (0 positive droplets).

Primers that amplified transcripts specifically in cell lines and not in leukocytes in the above droplet digital PCR testing were then subjected to detailed testing of sensitivity of signal. Using single cell micromanipulation, precise numbers
30 of cancer cells (1, 5, 10, 25, and 50 cells) were spiked into whole blood donated by healthy individuals, and then processed through the CTC-iChip. Each sample was

then processed as above for testing with droplet digital PCR, and evaluated for sensitivity to ensure the signal was sufficient for the desired clinical application.

The above stringent procedure of evaluating candidate genes and primers using qRT-PCR and droplet digital PCR resulted in a final primer list consisting of approximately 10% of the initial list of 50-100 candidate genes for each type of cancer (total of approximately 400 initial candidate genes). These primers are then further evaluated for signal in patient CTCs using blood samples donated by cancer patients undergoing cancer treatment at the MGH Cancer Center, collected under an IRB-approved clinical protocol. Key to this portion of the evaluation is a comparison with blood collected from healthy individuals without a diagnosis of cancer. The following **Table 1** lists the primers and probes for that have been developed thus far using these methods for the specific detection of CTCs from patients with prostate, breast, hepatocellular, pancreatic, lung, and melanoma cancers using droplet digital PCR.

While a single gene for each cancer type could be used, the presence of multiple genes within each panel is useful both for sensitivity (CTCs are heterogeneous even within individual patients in their expression patterns) and specificity (detection of multiple gene signals confers added confidence that this represents a true cancer cell signature).

The gene list provided below in **Table 1** includes transcripts that are unique to specific types of cancer (e.g., highly specific markers of prostate or breast or liver cancers), as well as genes that are shared by several cancer types, e.g., all epithelial cancer types (and thus may serve as pan-cancer markers), and genes that are induced in certain conditions (e.g., active androgen signaling in prostate cancer or active estrogen signaling in breast cancer). Thus, each type of cancer was assigned a specific panel of genes that is designed for optimal sensitivity, specificity, and clinically actionable information for the given cancer type.

In addition, primers described in Table 2 are designed to pre-amplify some of the genes listed in Table 1, while maintaining their high specificity. If STA is a method of choice, these nested primers become additional components of each cancer panel.

Gene Lists for Different Types of Cancers

The following Table 1 provides a list of names of genes (with (Genbank ID) and Sequence Identification numbers (SEQ ID NO)), along with cancer types for which they are selective (Br: breast, Lu: lung, Li: liver, Pr: prostate, Panc: pancreatic, Mel: melanoma). In addition, optimized primer sets are listed for each gene (primers 1 and 2), along with the composition of the fluorescent primer probes (e.g., 6-FAMTM (blue fluorescent label) or HEXTM (green fluorescent label) for tagged probes, and ZEN-31ABkFQ quencher) for optimal visualization of the digital PCR product.

TABLE 1

Gene	Disease Group	Seq ID	Primer 1	Seq ID	Primer 2	Seq ID	Probe
AGR2 (10553)	Br, Lu, Li, Pr	1	CTG ACA GTT AGA GCC GAT ATC AC	2	CAA TTC AGT CTT CAG CAA CTT GAG	3	/56-FAM/ATG CTT ACG /ZEN/AAC CTG CAG ATA CAG CTC /31ABkFQ/
ALDH1A3 (220)	Br, Lu, Panc	4	GGT GGC TTT AAA ATG TCA GGA A	5	TGT CGC CAA GTT TGA TGG T	6	/56-FAM/TTT TCA CTT /ZEN/CTG TGT ATT CGG CCA AAG C /31ABkFQ/
CADPS2 (93664)	Br, Li, Lu, Mel	7	CTC TGF ATT TTT GGA CAT AGG AG	8	GCC TTG CAC TTC CAT TAT GAC	9	/56-FAM/TCC GAC GTG /ZEN/GTA CTG TCA TTC ACC T /31ABkFQ/
CDH11 (1009)	Br, Lu, Panc	10	GAG GCC TAC ATT CTG AAC GC	11	GTG GTT CTT TCT TTT GCC TTC TC	12	/56-FAM/CAT CCT CGC /ZEN/CTG CAT CGT CAT TCT /31ABkFQ/
CDH3 (1004)	Br, Li, Mel	13	GTT TCA TCC TCC CTG TGC TG	14	GCT CCT TGA TCT TCC GCT TC	15	/56-FAM/CTG CTG GTG /ZEN/CTG CTT TTG TTG GT /31ABkFQ/
COL8A1 (1295)	Br, Lu	16	GAT GCC CCA CTT GCA GTA	17	CCT CGT AAA CTG GCT AAT GGT	18	/56-FAM/AGT ATC CAC /ZEN/ACC TAC CCC AAT ATA TGA AGG AAA /31ABkFQ/
EGFR (1956)	Br, Lu, Li, Panc	19	CTG CTG CCA CAA CCA GT	20	TTC ACA TCC ATC TGG TAC GTG	21	/56-FAM/CTG CCT GGT /ZEN/CTG CCG CAA ATT C /31ABkFQ/
FAT1 (2195)	Br, Lu, Li, Mel, Pr, Panc	22	GAT CCT TAT GCC ATC ACC GT	23	ATC AGC AGA GTC AAT CAG TGA G	24	/56-FAM/TCT TGT CAG /ZEN/CAG CGT TCC CGG /31ABkFQ/
FAT2 (2196)	Br, Lu	25	CCT GGA TGC TGA CAT TTC TGA	26	TCC TCC ACT CAT CTC CAA CT	27	/56-FAM/ACC TGC TAC /ZEN/ATC ACA GAG GGA GAC C /31ABkFQ/
FOLH1 (2346)	Pr	28	CAA TGT GAT AGG TAC TCT CAG AGG	29	TGT TCC AAA GCT CCT CAC AA	30	/56-FAM/ATG AAC AAC /ZEN/AGC TGC TCC ACT CTG A /31ABkFQ/
HQXB13 (251729)	Br, Lu, Pr	31	CAG CCA GAT GTG TTG CCA	32	CTG TAC GGA ATG CGT TTC TTG	33	/56-FAM/CAG CAT TTG /ZEN/CAG ACT CCA GCG G /31ABkFQ/
KLK2 (3817)	Pr	34	GCT GTG TAC AGT CAT GGA TGG	35	GTG TTC AGG CTC AAA CAG GT	36	/56-FAM/TGG CTA TTC /ZEN/TTC TTT AGG CAA TGG GCA /31ABkFQ/
KLK3 (354)	Pr	37	GTG TGC TGG ACG CTG GA	38	GTG ATA CCT TGA AGC ACA CCA TTA C	39	/56-FAM/AAA GCA CCT /ZEN/GCT CGG GTG ATT CT /31ABkFQ/
LSAMP (4045)	Mel	40	CAC ATT TGA GTG AAG CTT GTC G	41	GCG GAT GTC AAA CAA GTC AAG	42	/56-FAM/TCC AAG AGC /ZEN/AAT GAA GCC ACC ACA /31ABkFQ/
MAGEA6-RM1 (4105)	Mel	43	GAA GGA GAA GAT CTG CCA GTG	44	GCT GAC TCC TCT GCT CAA G	45	/56-FAM/TTG CCC TGA /ZEN/CCA GAG TCA TCA TGC /31ABkFQ/

Gene	Disease Group	Seq ID	Primer 2	Seq ID	Primer 1	Seq ID	Probe
ME1 {4233}	Br, Li, Lu, Panc	46	CCA GTA GCC TGA TTG TGC AT	47	TGT CAG TGA TTC TGT TCA AGG A	48	/56-FAM/AGT CAT AGG /ZEN/AAG AGG GCA TTT TGG TTG T/3IABkFQ/
MLANA {2315}	Mel	49	ACT CTT ACA CCA CGG CTG A	50	CCA TCA AGG CTC TGT ATC CAT	51	/56-FAM/AAG ACT CCC /ZEN/AGG ATC ACT GTC AGG A/3IABkFQ/
NPY1R {4986}	Br, Lu	52	GGA TCT GAG CAG GAG AAA TAC C	53	GAA TTC TTC ATT CCC TTG AAC TGA	54	/56-FAM/AGC AGG AGC /ZEN/GAA AAA GAC AAA TTC CAA AG/3IABkFQ/
OCLN {100506658 }	Br, Lu, Li	55	AAG ATG GAC AGG TAT GAC AAG TC	56	ACT CTT TCC ACA TAG TCA GAT GG	57	/56-FAM/TGC AGA CAC /ZEN/ATT TTT AAC CCA CTC CTC G/3IABkFQ/
PDZRN3 {23074}	Mel	58	TGT CTT GGC TGT TCA TTC TG	59	TGG ATC CCT ATC TCT TGC CA	60	/56-FAM/AGC TCC TCC /ZEN/CTG TCC ATC TCC T/3IABkFQ/
PGR {5241}	Br	61	GGC AAT TGG TTT GAG GCA A	62	GGA CTG GAT AAA TGT ATT CAA GCA	63	/56-FAM/ACA AGA TCA /ZEN/TGC AAG TTA TCA AGA AGT TTT GTA AGT T/3IABkFQ/
PKP3 {11187}	Br, Li, Lu, Panc	64	CTG GTG GAG GAG AAC GG	65	GGT CGC TGG ATG AAA GGT T	66	/56-FAM/AGT GTC CGC /ZEN/AGC AGC TCG AA/3IABkFQ/
PMEL {6490}	Mel	67	CAG GCA TCG TCA GTT TCC T	68	ACA CAA TGG ATC TGG TGC TAA	69	/56-FAM/TTT GGC TGT /ZEN/GAT AGG TGC TTT GCT G/3IABkFQ/
PPL {5493}	Br, Lu, Li	70	GAG GAG AGA ATC AAC AAA CTG C	71	AGG TTC AGG TAC TCC TTC CAG	72	/56-FAM/AGG AAC TCC /ZEN/ATT GAG GCG CAC AT/3IABkFQ/
RXRG {6258}	Mel	73	ATA CTT CTG CTT GGT GTA GGC	74	AGC CAT TGT ACT CTT TAA CCC A	75	/56-FAM/CTC TGA GGT /ZEN/GGA GAC TCT GCG AGA /3IABkFQ/
RND3 {398}	Br, Lu, Li, Mel, Panc	76	CCG AGA ATT ACG TTC CTA CAG TG	77	GCG GAC ATT GTC ATA GTA AGG A	78	/56-FAM/ACG GCC AGT /ZEN/TTT GAA ATC GAC ACA C/3IABkFQ/
S100A2 {6273}	Br, Lu, Li, Panc	79	CTG CCT TGC TCT CCT TCC	80	CTT ACT CAG CTT GAA CTT GTC G	81	/56-FAM/ACC TGG TCT /ZEN/GCC ACA GAT CCA TG/3IABkFQ/
SCGB2A1 {4246}	Br	82	ACT TCC TTG ATC CCT GCC A	83	GTC TTT TCA ACC ATG TCC TCC A	84	/56-FAM/CCA TGA AGC /ZEN/TGC TGA TGG TCC TCA /3IABkFQ/
SFRP1 {6422}	Mel	85	CAA TGC CAC CGA AGC CT	86	CTT TTA TTT TCA TCC TCA GTG CAA AC	87	/56-FAM/TGT GAC AAC /ZEN/GAG TTG AAA TCT GAG GCC /3IABkFQ/
SODX10 {6663}	Mel	88	CTT GTC ACT TTC GTT CAG CAG	89	CTT CAT GGT GTG GGC TCA	90	/56-FAM/TTG TGC AGG /ZEN/TGC GGG TAC TGG /3IABkFQ/
Gene	Disease Group	Seq ID	Primer 2	Seq ID	Primer 1	Seq ID	Probe
SCHLAP1/SET4 {101069767}	Pr	91	TCC TTG GAT GAC TCT CCC TAC	92	AGA TAC CAC CTC CCT GAA GAA	93	/56-FAM/CCA ATG ATG /ZEN/AGG AGC GGG ATG GAG /3IABkFQ/
SCHLAP1 SET 5	Pr	94	AGA GGT TTA ATG GGC TCA CAG	95	CTC TGG TCT GTC GTC ATG TAA G	96	/56-FAM/ACA TGC CTT /ZEN/TCA CCT TCT CCA CCA /3IABkFQ/
AMACR {23600}	Pr	97	CAC ACC ACC ATA CCT GGA TAA T	97	TCA CTT GAG GCC AAG AGT TC	99	/56-FAM/AGA AAC GGA /ZEN/GGT CCA GCC AAG TTC /3IABkFQ/
AR Variant 7/SET1 {367}	Pr	100	CTT TCT TCA GGG TCT GGT CAT T	101	CTT GTC GTC TTC GGA AAT GTT ATG	102	/56-FAM/AAG CAG GGA /ZEN/TGA CTC TGG GAG AAA /3IABkFQ/
AR Variant 7 SET 3	Pr	103	GAG GCA AGT CAG CCT TTC T	104	TGT CCA TCT TGT CGT CTT CG	105	/56-FAM/TGA AGC AGG /ZEN/GAT GAC TCT GGG AGA /3IABkFQ/
AR Variant 12 SET 1	Pr	106	GCT CAC CAT GTG TGA CTT GA	107	TGG GAG AGA GAC AGC TTG TA	108	/56-FAM/TGA TTG CGA /ZEN/GAG AGC TGC ATC AGT /3IABkFQ/
AR Variant 12 SET 4	Pr	109	GAA AGT CCA CGC TCA CCA T	110	GCA GCC TTG CTC TCT AGC	111	/56-FAM/TGA TTG CGA /ZEN/GAG AGC TGC ATC AGT /3IABkFQ/
UGT2B15 SET 1 {7366}	Pr	112	CTC TGC ACA AAC TCT TCC ATT TC	113	TTT CCT CGC CCA TTC TTA CC	114	/56-FAM/TTG GCT GGT /ZEN/TTA CAG TGA AGT CCT CC/3IABkFQ/
UGT2B15 SET 5	PR	115	GGA AGG AGG GAA CAG AAA TCC	116	GTG AGC TAC TGG CTG AAC TAT T	117	/56-FAM/TGG CTA CAC /ZEN/ATT TGA GAA GAA TGG TGG A/3IABkFQ/
AFP SET 1 {174}	Li	118	AGG AGA TGT GCT GGA TTG TC	119	TCT GCA TGA ATT ATA CAT TGA CCA C	120	/56-FAM/AAT GCT GCA /ZEN/AAC TGA CCA CGC TG/3IABkFQ/
AFP SET 2	Li	121	ACT GCA GAG ATA AGT TTA GCT GAC	122	TCA CCA TTT TGC TTA CTT CCT TG	123	/56-FAM/TTG CCC AGT /ZEN/TTG TTC AAG AAG CCA C/3IABkFQ/

Gene	Strain	Seq ID	Primer 2	Seq ID	Primer 1	Seq ID	Probe
SILAP2 (261729)	Br, Lu, Pr, Panc	124	CAT GTT GCC TAC AGC CTC T	125	TCT CCA AAC TTC TTC CTC ATT CC	126	/56-FAM/ACA TGG CTT /ZEN/ATC AGC AGG TTC ATG CA/3IABKFIQ/
TEAD3 (7805)	Br, Lu, Li	127	GAA GAT CAT CCT GTC AGA CGA G	128	CTT CCG AGC TAG AAC CTG TAT G	129	/56-FAM/AGC GTG CAA /ZEN/TCA ACT CAT TTC GGC /3IABKFIQ/
TFAP2C (7022)	Br, Lu, Mel	130	GAT CAG ACA GTC ATT CGC AAA G	131	GAC AAT CTT CCA GGG ACT GAG	132	/56-FAM/ACA GGG GAG /ZEN/GTT CAG AGG GTT CTT /3IABKFIQ/
TMPRSS 2 (7113)	Pr	133	CCC AAC CCA GGC ATG ATG	134	TCA ATG AGA AGC ACC TTG GC	135	/56-FAM/ACC CGG AAA /ZEN/TCC AGC AGA GCT /3IABKFIQ/
GPC3 (2719)	Li	136	TGC TGG AAT GGA CAA GAA CTC	137	GCT CAT GGA GAT TGA ACT GGT	138	/56-FAM/TCC TTG CTG /ZEN/CCT TTT GGC TGT ATC T/3IABKFIQ/
ALB (219)	Li	139	CTT ACT GGC GTT TTC TCA TGC	140	CCA ACT CTT GTA GAG GTC TCA AG	141	/56-FAM/ACA TTT GCT /ZEN/GCC CAC TTT TCC TAG GT/3IABKFIQ/
G6PC SET 1 (2538)	Li	142	GGA CCA GGG AAA GAT AAA GCC	143	GCA AGG TAG ATT CGT GAC AGA	144	/56-FAM/ACA GCC CAG /ZEN/AAT CCC AAC CAC AAA /3IABKFIQ/
G6PC SET 2	Li	145	CAT TTT GTG GTT GGG ATT CTG G	146	GAT GCT GTG GAT GTG GCT	147	/56-FAM/CTG TCA CGA /ZEN/ATC TAC CTT GCT GCT CA/3IABKFIQ/
PRAME (23532)	Mel	148	GCC TTG CAC TTC CAT TAT GAC	149	CTC TGC ATT TTT GGA CAT AGG AG	150	/56-FAM/CAA GCG TTG /ZEN/GAG GTC CTG AGG C/3IABKFIQ/
AHSG (197)	Li	151	ATG TGG AGT TTA CAG TGT CTG G	152	AGC TTC TCA CTG AGT GTT GC	153	/56-FAM/CCA CAG AGG /ZEN/CAG CCA AGT GTA ACC /3IABKFIQ/
GPR143 (4915)	Mel	154	ACG GCT CCC ATC CTC CT	155	CCA CTA TGT CAC CAT GTA CCT G	156	/56-FAM/TTC GCC ACG /ZEN/AGA ACC AGC AGC /3IABKFIQ/
PTPRZ1 (5803)	Mel	157	TGC TCT GAC AAC CCT TAT GC	158	GGC TGA GGA TCA CTT TGT AGA	159	/56-FAM/AGG CCA GGA /ZEN/GTC TTT GCT GAC ATT /3IABKFIQ/

Gene	Accession Group	Seq ID	Primer 2	Seq ID	Primer 1	Seq ID	Probe
MUCL1 (118430)	Br	160	CAT CAG CAG GAC CAG TAG C	161	TGT CTG TGC TCC CTG ATC T	162	/56-FAM/ACT CCC AAG /ZEN/AGT ACC AGG ACT GCT /3IABkFQ/
PIP (5304)	Br	163	TCA TTT GGA CGT ACT GAC TTG G	164	CTT GCT CCA GCT CCT GTT C	165	/5HEX/CCT GCT CCT /ZEN/GGT TCT CTG CCT G/3IABkFQ/
PGR (3241)	Br	166	GGT GTT TGG TET AGG ATG GAG	167	ACT GGG TTT GAC TTC GTA GC	168	/56-FAM/AGT GGG CAG /ZEN/ATG CTG TAT TTT GCA C/3IABkFQ/
TFAP2C (7022)	Br, Lu	169	GTG ACT CTC CTG ACA TCC TTA G	170	CCA TCT CAT TTC GTC CTC CAA	171	/56-FAM/TTC GGC TTC /ZEN/ACA GAC ATA GGC AAA GT/3IABkFQ/
SCGB2A1 (4246)	Br	172	ACT CTG AAA AAC TTT GGA CTG ATG	173	TCT AGC AAT CAA CAG ATG AGT TCT	174	/56-FAM/TAG CCC TCT /ZEN/GAG CCA AAC GCC /3IABkFQ/
FAT1 (2195)	Br, Lu, Pr	175	AGC TCC TTC CAG TCC GAA T	176	GTC TGC TCA TCA ATC ACC TCA	177	/56-FAM/ATC CCA GTG /ZEN/ATA CCC ATT GTC ATC GC/3IABkFQ/
FAT2 (2196)	Br, Lu, Pr	178	GGA CAG AGA GAA CAA GGA TGA AC	179	TGT GGG AGA ATA TAG GTG GAT TG	180	/56-FAM/TGG AGG TGA /ZEN/CTG TGC TGG ACA ATG /3IABkFQ/
RND3 (390)	Br, Lu	181	GCT TTG ACA TCA GTA GAC CAG AG	182	CTG TCC GCA GAT CAG ACT TG	183	/56-FAM/ACA GTG TCC /ZEN/TCA AAA AGT GGA AAG GTG A/3IABkFQ/
SFTPB (6439)	Lu	184	CCT GGA AAA TGG CCT CCT T	185	CAT TGC CTA CAG GAA GTC TGG	186	/56-FAM/CCG ATG ACC /ZEN/TAT GCC AAG AGT GTG AG/3IABkFQ/
SCGB3A2 (117156)	Lu	187	CCA GAG GTA AAG GTG CCA AC	188	TCC CAG ATA ACT GTC ATG AAG C	189	/56-FAM/AAG GCA GTA /ZEN/GCA GAG TAA CTA CAA AGG C/3IABkFQ/
SERPINA3 (12)	Br, Lu	190	CCT CAA ATA CAT CAA GCA CAG C	191	GGA AGE CTT CAC CAG CAA	192	/56-FAM/TAG CAG TCT /ZEN/CCC AGG TGG TCC A/3IABkFQ/
SFRP2 (8423)	Br, Lu	193	TTG CAG GCT TCA CAT ACC TT	194	GCC CGA CAT GCT TGA GT	195	/56-FAM/TTT CCC CCA /ZEN/GGA CAA CGA CCT TT/3IABkFQ/
CRABP2 (1382)	Br, Lu	196	CTC TTG CAG CCA TTC CTC TT	197	CCC TTA CCC CAG TCA CTT CT	198	/56-FAM/TTT CTT TGA /ZEN/CCT CTT CTC TCC TCC CCT /3IABkFQ/
AQP4 (361)	Lu	199	TGG ACA GAA GAC ATA CTC ATA AAG G	200	GGT GCC AGC ATG AAT CCC	201	/56-FAM/CCG ATC CTT /ZEN/TGG ACC TGC AGT TAT CA/3IABkFQ/
TM6PS54 (56649)	Br, Lu	202	ATC TTC CCT CCA TTC TGC TTC	203	CAG TTC CCA CTC ACT TTC TCA G	204	/56-FAM/CTC ACT CCA /ZEN/GCC ACC CCA CTC /3IABkFQ/
GREM1 (26585)	Lu	205	TTT TGC ACC AGT CTC GCT T	206	GCC GCA CTG ACA GTA TGA G	207	/56-FAM/CCT ACA CGG /ZEN/TGG GAG CCC TG/3IABkFQ/

Gene	Accession Group	Seq ID	Primer 2	Seq ID	Primer 1	Seq ID	Probe
FOXF1 (2294)	Lu	208	CGA CTG CGA GTG ATA CCG	209	CTC TCC ACG CAC TCC CT	210	/56-FAM/CTG CAC CAG /ZEN/AAC AGC CAC AAC G/3IABkFQ/
NKX2-1 (7080)	Lu	211	TGC CGC TCA TGT TCA TGC	212	CAG GAC ACC ATG AGG AAC AG	213	/56-FAM/CCC GCC ATC /ZEN/TCC CGC TTC A/3IABkFQ/
NKX2-1 (7080)	Lu	214	AAG ATG TCA GAC ACT GAG AAC G	215	CGA AGE CCG ATG TGG TC	216	/56-FAM/ATG TCG ATG /ZEN/AGT CCA AAG CAC ACG A/3IABkFQ/
AFP (174)	Li	217	AGGAGATGTGCTGGATTGTC	218	TCCTCATGSAATTATACATTGAC CAC	219	/56-FAM/AAT GCT GCA /ZEN/AAC TGA CCA CGC TG/3IABkFQ/
AH5G (197)	Li	220	ATGTGGAGTTTACAGTCTCTG G	221	AGCTTCTCTAGTGGTTTGC	222	/56-FAM/CCA CAG AGG /ZEN/CAG CCA AGT GTA ACC /3IABkFQ/
ALB (213)	Li	223	GAG ATC TGC TTG AAT GTG CTG	224	CAA CAG AGG TTT TTC ACA GCA T	225	/56-FAM/AGA TAT ACT /ZEN/TGG CAA GGT CCG CCC /3IABkFQ/
ALB (213)	Li	226	CAT GGT AGG CTG AGA TGC TTT	227	GAC GAT AAG GAG ACC TGC TTT G	228	/56-FAM/ACT TGT TGC /ZEN/TGC AAG TCA AGE TGC /3IABkFQ/
ALB (213)	Li	229	GCG CAT TCT GGA ATT TGT ACT C	230	GCT ATG CCA AAG TGT TCG ATG	231	/56-FAM/ACC TCT TGT /ZEN/GGA AGA GCC TCA GAA /3IABkFQ/
APOL1 (350)	Li	232	TGA TGG ATA TTC TCT GGA TGG C	233	CCT GAA TCT TTA CTC TCT CTC CTT G	234	/56-FAM/CCA GTT TCC /ZEN/CAG TTT GGT ACA TTC TAT TTC TTC C/3IABkFQ/
FABP1 (2108)	Li	235	GCA CTT CAA GTT CAC CAT CAC	236	ACC AGT TTA TTG TCA CCT TCC A	237	/56-FAM/AAC CAC TGT /ZEN/CTT GAC TTT CTC CCC TG/3IABkFQ/
FGB (2244)	Li	238	ACA TCT ATT ATT GCT ACT ATT GTG TGT T	239	TGG GAG CTT CTT CTC TCT TC	240	/56-FAM/ACC CTC CTC /ZEN/ATT GTC GTT GAC ACC /3IABkFQ/
FGG (2266)	Li	241	TTC ATT TGA TAA GCA CAC AGT CTG	242	ACC TTG AAC ATG GCA TAG TCT G	243	/56-FAM/TGC CAT TCC /ZEN/AGT CTT CCA GTT CCA C/3IABkFQ/
GPC3 (2719)	Li	244	AATCAGCTCGGTTCCTTG	245	TGCTATCTCGTGTGCTCTCG	246	/56-FAM/TTT CAG GCG /ZEN/CAT CAT CCA CAT CC/3IABkFQ/
RBP4 (5950)	Li	247	CAG AAG CGC AGA AGA TTG TAA G	248	TCT TTC TGA TCT GCC ATC GC	249	/56-FAM/AGG CTG ATC /ZEN/GTC CAC AAC GGT T/3IABkFQ/
TF (7018)	Li	250	AGA AGE GAG TCC GAC TGT	251	CAC TGC ACA CCA TCT CAC A	252	/56-FAM/CCA GAC ACA /ZEN/GCC CCA GGA CG/3IABkFQ/

Note that PRAME is also named MAPE (Melanoma Antigen Preferentially Expressed In Tumors), OIP4 (Opa-Interacting Protein OIP4), and CT130 (Cancer/Testis Antigen 130).

5

The following Table 2 lists nested primers designed to specifically pre-amplify the regions targeted by primers listed in Table 1.

Table 2

Primer name	Seq ID	Nested Forward	Seq ID	Nested Reverse
FAT1	253	CAG ATG GAG GAG GAA GAT TCT G	254	GTA TAC TGC CTG GAG TTC TCT G
FAT2	255	CTG GTT CAG GTC TCC ATT ACA G	256	GCT GTG ACT CTG AGC AAG TA
AGR2	257	TGT CCT CCT CAA TCT GGT TTA TG	258	GAC AGA AGG GCT TGG AGA TTT
PKP3	259	CGG TGG CGT TGT AGA AGA T	260	AGA AGA TCT CTG CCT CCG A
RND3	261	CAA GAT AGT TGT GGT GGG AGA C	262	AGG GTC TCT GGT CTA CTG ATG
TFAP2C	263	TTTGGATTACCGCTTGGG	264	GACTCCAGTGTGGGAGAG
S100A2	265	GGG CCC ACA TAT AAA TCC TCA C	266	CTG CTG GTC ACT GTT CTC ATC
PRAME	267	CTTCGCGGTGTGGTGAA	268	GCTGTGTCTCCGTCAAA
PIP	269	CTG GGA CAC ATT GCC TTC T	270	CCA CCA TGC ATT CTT TCA ATT CT
PGR	271	AAA CCC AGT TTG AGG AGA TGA G	272	CCC TGC CAA TAT CTT GGG TAA T
SCGB2A1	273	ACA GCA ACT TCC TTG ATC CC	274	GCG GCA TCA CTG TCT ATG AA
MUC11	275	CCT TGC CTT CTC TTA GGC TTT	276	AGC AGT GGT TTC AGC ATC A
PGR	277	CAG ATA ACT CTC ATT CAG TAT TCT TGG	278	CTC TAA TGT AGC TTG ACC TCA TCT
TFAP2C	279	GAG AAG TTG GAC AAG ATT GGG	280	GCT GAG AAG TTC TGT GAA TTC TTT A
SCGB2A1	281	GTT TCC TCA ACC AGT CAC ATA GA	282	AGT TGT CTA GCA GTT TCC ACA TA

10

Primer name	Seq ID	Nested Forward	Seq ID	Nested Reverse
FAT1	283	GGG AAA GCC TGT CTG AAG TG	284	TCG TAG CCT CCA GGG TAA TAG
FAT2	285	GTT ACA GGT CTC CTA TCT ACA GC	286	GCT CAG CCT CTC TGG AAG
RND3	287	CTC TCT TAC CCT GAT TCG GAT G	288	GGC GTC TGC CTG TGA TT
SFTP8	289	CCT GAG TTC TGG TGC CAA AG	290	GGG CAT GAG CAG CTT CAA
SCGB3A2	291	CCA CTG GCT TGG TGG ATT T	292	TCA ACA GAA ATG CCC AGA GTT
SERPINA3	293	CTT CTC CAG CTG GGC ATT	294	TGC TGT GGC AGC AGA TG
SFRP2	295	CGG TCA TGT CCG CCT TC	296	GCG TTT CCA TTA TGT CGT TGT C
CRABP2	297	CCC TCC TTC TAG GAT AGC G	298	AAC CCG GAA TGG GTG AT
AQP4	299	AAACGGACTGATGTCCTGG	300	TGGACAGAAGACATACTCATAAAGG
TMPR554	301	CCCACTGCTTCAGGAAACATA	302	GTCAGACATCTTCCCTCCATTC
GREM1	303	GCCGCACTGACAGTATGA	304	CAGAAGGAGCAGGACTGAAA
FOXF1	305	AGC GGC GCC TCT TAT ATC	306	GCG TTG AAA GAG AAG ACA AAC T
NKX2-1	307	CTA CTG CAA CGG CAA CCT	308	GGG CCA TGT TCT TGC TCA
NKX2-1	309	CAG ACT CGC TCG CTC ATT T	310	CCT CCA TGC CCA CTT TCT T
PIP	311	CCCAAGTCAGTACGTCCAAAT	312	GCCTAATTCCEGAATAACATEAAC
AGR2	313	GCT TTA AAG AAA GTG TTT GCT G	314	CTG TAT CTG CAG GTT CGT AAG

Primer name	Seq ID	Nested Forward	Seq ID	Nested Reverse
SOX10	315	AAG TTC CCC GTG TGC ATC	316	CTC AGC CTC CTC GAT GAA
MAGEA6	317	GTGAGGAGGCAAGGTTCTG	318	GGCTCCAGAGAGGGTAGTT
TFAP2C	319	TTTGGATTACCGCTTGGG	320	GACTCCAGTGTGGGAGAG
PRAME	321	CTTCGCGGTGTGGTGAA	322	GCTGTGTCTCCCGTCAAA
GPR143	323	ATC CTG CTG TAT CAC ATC ATG	324	CTG ACA GGT TTC AAA GAA CCT
PMEL	325	CCAGTGCCCTTGTTGCT	326	CAAGAGCCAGATGGGCAAG
MLANA	327	TGCCAAGAGAAGATGCTCAC	328	CATTGAGTGCCAACATGAAGAC
PTPRZ1	329	AAG AAG CTG CCA ATA GGG AT	330	TGT CCA GAG AGG TGG ATG

Multiplex Digital Analysis of Gene Transcripts from CTC-Chip Products

To improve the detection of tumor-specific mRNA from minimal amounts of RNA derived from CTCs, we established a multiplex assay capable of testing many different gene transcripts from a minute amount of CTC-Chip product. This
5 combines the higher sensitivity/specificity of using multiple independent genes, with the fact that the amount of input template is limited (and hence should not be diluted into multiple reactions). Our assay includes 4 genes per reaction, with each gene being resolved uniquely in 2-dimensional space by selecting different ratios of fluorescent conjugated primers. Thus, in a single reaction, we can independently
10 measure 4 gene transcripts without having to dilute the template. For different cancers, we have gone as far as up to 4 different reactions (i.e., up to 20 different gene transcripts), and with application of nested RT-PCR digital assays, there is no limit to the number of reactions that can be performed.

This multiplex strategy achieves the ideal balance between analyzing multiple
15 transcripts (and hence ensuring against heterogeneous variation in cancer cell expression patterns), but not diluting the input material by performing multiple independent PCR reactions. Depending on tumor types and the number of genes required for optimal signal, we have developed assays ranging from 2-4 multiplex reactions (each multiplex reaction testing for 4-genes). Thus, without undue dilution
20 of input template, we can interrogate the product of a single CTC for expression of anywhere from 8 to 16 different genes. It is important to the assay to be able to add the signal from all of these genes (i.e. cumulative signal), while also having individual gene results (to optimize signal/noise at the individual gene level, and also gather information from specific signaling pathways that each gene interrogates – for
25 example androgen signaling in prostate CTCs).

To display the results of the multiplex reaction in a single view (and hence differentiate amplification of each gene is isolation), we varied the concentrations of the two fluorescent probes (FAM (blue) and HEX (green)). By doing this, each individual gene amplification reaction has a unique combination of FAM/HEX signal
30 that reflects the composition of the gene-specific primers, and hence identifies the gene-specific PCR product. In 2-dimensional space, we can illustrate the signal position of 4 different gene amplification products produced from a single multiplex

reaction. As applied to digital PCR using droplets to encapsulate each PCR reaction, this method separates the targets into individual clusters by modifying the binary signal amplitude of positive droplets, which are displayed quantitatively. As predicted, this method allows both cumulative scoring of total signal for multiple genes (e.g., 16 markers in a total of 4 reactions), while also retaining the ability to quantify the signal from each individual gene target.

Specific results of testing are detailed in the examples below.

Applications of the d-CTC Assay Methods

The early detection of epithelial cancers at a time when they can be surgically resected or irradiated provides the best chance of cure, and the administration of adjuvant chemotherapy in the setting of minimal cancer dissemination is far more effective in achieving cure than the treatment of established metastatic disease. However, current efforts at early cancer detection suffer from lack of specificity. For instance widespread screening of men for prostate cancer, using serum PSA measurements is effective in uncovering early cancers, but it also identifies a much larger number of non-malignant prostate conditions (e.g., benign hypertrophy of the gland) or even cancers that are indolent and never destined to become invasive. As such, broad PSA screening is not recommended by public health organizations, because the number of complications (including deaths) from over-diagnosis match or even outweigh the calculated benefit in early cancer detection.

For other cancers, such as breast cancer, mammography is considered effective, but even then a large number of breast biopsies are performed to diagnose each true malignancy. For lung cancer, the recently recommended low dose CT scanning of individuals with a heavy cigarette smoking history is also likely to detect hundreds of innocent radiographic abnormalities for each true malignancy.

It is in this context that the addition of a blood-based ultra-sensitive readout for the presence of cancer cell-derived signatures would provide the required specificity. The d-CTC assays described herein can be used for both initial screening and as a confirmation of earlier screenings at a later time. For example, in some cases the assays can be used as a second-line test to validate a highly sensitive, but nonspecific screening test (e.g., PSA in prostate cancer). In other settings for which a

cancer is highly lethal, but no screening approach currently exists (e.g., pancreatic cancer), routine periodic blood screening using the assays described herein may become the norm to monitor a patient's status or condition over time.

5 The new d-CTC readouts are also highly relevant to the serial monitoring of patients, e.g., seemingly healthy patients with a family history and/or genetic markers of a specific type of cancer, or patients with advanced or metastatic cancer. Imaging of CTCs is expensive and relatively insensitive, in that intact cells that stain appropriately for all required markers produce a single signal. The use of the new d-CTC assays described herein, in which each CTC (no matter how intact or pre-
10 apoptotic) can give rise to hundreds of molecular signals, dramatically enhances the ability to detect and monitor CTCs in patients with known cancer, and to quantitatively monitor and analyze their response to therapeutic interventions. Beyond scoring for cell numbers through molecular markers, specific interrogation of mutations or cancer-associated rearrangements (e.g., EML4-ALK in lung cancer) can
15 be achieved with comparable sensitivity.

 In addition to providing a digital (quantitative) measure of CTCs present within a blood sample, the new d-CTC assay also allows analysis of specific signaling pathways that are unique to the tumor cells in the blood. For instance, a subset of prostate lineage-specific genes are driven by androgen signaling (such as PSA), while
20 another subset are repressed by androgen signaling (such as PSMA). By analyzing these genes together, we can ascertain the status of androgen signaling within CTCs. Similarly, in breast cancer, expression of estrogen-responsive genes (such as progesterone receptor) provides a measure of the status of the estrogen-responsive pathway within CTCs. These measurements are particularly important in that
25 therapeutic interventions in both prostate and breast cancers are derived to target the androgen and estrogen receptors, respectively. Thus, defining the total number of CTC signal in the blood, simultaneously with information about the effectiveness of the therapeutic agent in targeting and shutting off the critical pathway is important for therapeutic monitoring.

30 As discussed in the examples below, the new methods described herein are illustrated in prostate cancer, where the anti-androgenic agent abiratorone (e.g.,

ZYTIGA®) is effective in suppressing cancer progression, particularly in tumors that are still dependent on the androgen pathway.

EXAMPLES

5 The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

Example 1 - Preliminary Testing and Validation of the Digital CTC Assay

10 To test the feasibility of CTC-Chip-Droplet assay, we first selected several transcripts that are specifically expressed in prostate tumor cells, but are absent in contaminating leukocytes. These were the prostate lineage specific markers KLK3 (kallikrein-related peptidase; aka Prostate Specific Antigen, or PSA), FOLH1 (Folate Hydrolase; aka Prostate Specific Membrane Antigen, or PSMA) and AMACR (alpha-methylacyl-CoA racemase), as well as EpCAM (Epithelial Cell Adhesion Molecule).
15 PCR conditions were optimized using intron-spanning primers and ZEN double-quenched FAM-labelled probes from Integrated DNA Technologies (Coralville, Iowa) following standard qPCR protocols. These conditions were first tested with encapsulated cDNA from admixtures of cancer cells and leukocytes in order to explore the dynamic range of the system. Next, using manual isolation techniques for
20 individually selecting cells, 0, 3, 6, 12, 25, and 125 prostate cancer LNCaP cells were progressively spiked into individual 5 ml aliquots of HD blood, followed by CTC-iChip processing, RT-PCR and droplet encapsulation using the RainDrop system. We chose KLK3 as the target transcript for this experiment as it is predicted to be modestly abundant. Using an intensity threshold of 5,000, we found that as few as 3
25 cells worth of KLK3 transcript were readily detected at approximately 250 droplets.

 Based on these preliminary data, we tested the CTC-Chip Droplet assay in patients with metastatic and localized prostate cancer versus healthy controls. Each sample was run through the iChip, then CTC-containing product was run through droplet RT-PCR using the four prostate markers mentioned above: KLK3, AMACR,
30 FOLH1 and EpCAM. Patients with either local or metastatic prostate cancer produced significantly higher positive droplet counts as compared to HD controls.

FIG 1A shows cDNA dilutions prepared from total RNA of LNCaP prostate cancer cells, mixed with leukocytes and analyzed by droplet PCR using two different prostate primer sets. The results represent several purities and show good response of positive droplet number across this range.

5 FIG. 1B shows manually isolated LNCaP cells spiked into HD blood samples, run through the iChip, and subjected to droplet RT-PCR (KLK3 primer set). The results show excellent sensitivity down to low numbers of target cells.

FIG. 1C shows the analysis of blood samples from healthy controls, patients with localized (resectable) prostate cancer and metastatic prostate cancer, processed
10 through the CTC-iChip, subjected to RT-PCR and droplet analysis using three prostate-specific and one epithelial-specific biomarkers (KLK3, AMACR, FOLH1, EpCAM). The results are shown for the total number of droplets/ml for all four markers combined.

These results suggest that the application of a droplet-based PCR readout to
15 the CTC-iChip greatly enhances its sensitivity in detecting virtually all CTCs present in a biological specimen. Taken together, the CTC-iChip and Droplet-PCR represent two powerful microfluidic technologies that are highly compatible with each other and can be integrated in-line to create a new and highly sensitive and accurate biological assay.

20 **Example 2 – Digital CTC Assay Protocol**

This example provides a general digital CTC assay protocol that can be used for the methods described herein. Different aspects of this general protocol were used in some of the Examples described herein. For example, Approach 1 of Step 3 of the
25 protocol described below (relating to RNA purification to cDNA synthesis), was used to generate data for FIGs. 15A to 15C. Approach 2 in Step 3 was used to generate data for FIGs. 19A to 24B.

1. Patient blood is run through I-Chip, version 1.3M or 1.4.5T. Sample is collected in a 15 mL conical tube on ice.

30 2. Sample is spun down at 4C. Supernatant is decanted and SUPERase™ In (DTT independent RNase inhibitor) + RNALater® Stabilization Solution

(prevents RNA degradation by inhibiting RNAses) is added to the pellet. Sample is flash frozen and placed at -80 until further processing. Samples are stable at -80.

3. There are two different processing protocols for RNA purification to cDNA synthesis that were used in the examples described below.

5 **Approach 1**

- a. Sample was thawed on ice.
- b. Direct lysis of sample using detergents (NP40, Tween20).
- c. Lysed sample was taken straight for cDNA synthesis (Superscript III).

- 10 d. After cDNA synthesis sample was purified via SPRI (Agencourt AMPure® XP beads) clean-up to clean up detergents and any nucleotides <100 bps.

Approach 2

- a. Sample was thawed on ice.

- 15 b. Sample was processed on RNeasy Qiagen Micro Kit. Protocol has some slight variations compared to traditional Qiagen recommendations. Higher volumes of Buffer RLT (Lysis buffer) were used as well as higher ETOH concentrations. These modifications were made because of RNALater® addition to the sample.

- 20 c. After cDNA synthesis – sample was purified via SPRI (Agencourt AMPure XP beads) clean-up to clean up detergents and any nucleotides < 100 bps.

4. cDNA (synthesized from Approach 1 or 2) can be processed in two different ways:

- 25 a. cDNA was used directly for ddPCR; or
- b. cDNA was amplified used a Fluidigm BioMark™ Nested PCR approach (primers from genes used for nested PCR have been pre-validated). Amplified cDNA was diluted.

5. cDNA (from step 4a or 4b), Biorad Supermix™ for probes, primer or primers (for gene of interest; up to 4 different primers (FAM and HEX) can be multiplexed) were added in a total volume of 22 µl.
- 30

- 6. Droplets were generated (~15,000-18,000 droplets per well).

7. Droplet Sample were put in a PCR machine. The PCR conditions were different than Biorad recommendations. We used a step-down rather than a slow ramp to ensure that all droplets reach the same temperature. This is different than what both RainDance and Biorad uses. Better results (i.e., more signal and more separation
5 between positive and negative droplets) can be obtained with the step-down rather than the gradient.

8. After the PCR, positive droplets were counted in a ddPCR machine.

9. Data is collected and analyzed using TIBCO® Spotfire® analysis software.

10 The reagents, reagent concentrations, and reaction volumes are provided below:

Reagents:

- Biorad ddPCR™ Supermix for Probes (No dUTP)
- IDT primers/probes (20x or 40x)
- cDNA (1ng/ul for cell lines)
- Nuclease free water
- Eppendorf semi-skirted 96 well plate (Only these plates work with the machine)

Testing Relevant Cell Lines

Per single reaction:

ddPCR Supermix	11.0 µl
Primer (20x)	1.10 µl
cDNA (1ng/ul)	1.10 µl
<u>Water</u>	<u>8.80 µl</u>
TOTAL	22.0 µl per well

A master-mix containing ddPCR supermix, cDNA, and water were aliquoted into wells and 1.1 µl of each the primer was added to each well and mixed well.

Patient Samples

Per single reaction for Individual Genes

ddPCR Supermix	11.0 µl
Primer (20x)	1.1 µl
<u>cDNA (patient)</u>	<u>Up to 9.9 µl (Balance with water if less)</u>
TOTAL	22.0 µl per well

Per single multiplexed reaction for Multiple Genes

	ddPCR Supermix	11.0 μ l
	Primer 1 (40x)	.55 μ l
	Primer 2 (40x)	.55 μ l
5	Primer 3 (40x)	.55 μ l
	Primer 4 (40x)	.55 μ l
	<u>cDNA (patient)</u>	<u>8.8 μl</u>
	TOTAL	22.0 μ l per well

When testing multiple patients against a gene-specific primer or multiplexing
 10 primers against multiple genes, a master-mix, which includes the ddPCR supermix
 and primers, was aliquoted into wells followed by addition of patient cDNA to each
 well and mixed well.

Example 3 - Protocol for Gene Validation

15 The following protocol was used for selecting the specific marker genes listed
 in Table 1.

1. Transcripts that are unique to CTCs and not expressed in white blood cells
 (WBCs), leukocytes, etc. were mined bioinformatically – Primary tumor and
 CTC gene expression data was compared to WBC gene expression datasets to
 20 isolate transcripts that were present only in primary tumor and/or CTCs.
2. Transcripts that passed a threshold cutoff were validated by qPCR.
3. Primers were synthesized by IDT. Probes were labeled with
 25 FAM/ZEN/IBFQ.
4. qPCR validation required that every transcript be validated by at least two
 independent primer sets on two different cell lines, 5 healthy donors WBCs
 (isolated via CPT column) and water as a negative control. 50 cycles for qPCR
 30 were used to confirm that expression of a transcript was only present in cell
 lines and not in healthy donors.
5. Transcripts that passed qPCR validation were validated on ddPCR with cell
 lines and healthy donors passed through the CTC-iChip (with and without cell
 35 spiking).
6. Panels of transcripts were multiplexed (up to 4 different genes per reaction)
 depending on disease of interest.

40 The validity of this strategy is shown below in a spiked cell experiment, in
 which a carefully measured number of tumor cells (from the LNCAP prostate cancer

cell line) are individually micro-manipulated, added to control blood specimens, passed through the CTC-iChip and then analyzed by d-CTC assay as above.

Increasing numbers of spiked cells show increasing numbers of digital signal as

shown in FIG. 2, which illustrates the power of this protocol. FIG. 2 demonstrates the

5 use of a single gene transcript (KLK3, also known as PSA, for prostate cancer) as a

probe (in the assay, we use from 8-24 gene transcripts, thereby further increasing

sensitivity). Here, we spike a calculated number of cancer cells (each cell is micro-

manipulated, picked and introduced into 10ml of control blood specimen). The blood

is then processed through the CTC-Chip and subjected to digital readout as described

10 above. No signal is observed in blood that has not been spiked with a single cancer

cell. Introduction of 2 cells/10 ml of blood generates clear signal (65 positive

droplets). In this case, the 10 CTC product was divided into 4 and run in

quadruplicate, so the 64 droplets actually represent the digital signal derived from $\frac{1}{4}$ of a tumor cell.

15 This assay is both highly sensitive and reproducible. As shown in FIG. 3, the digital signal in these spiked cell experiments shows high reproducibility (2

independent replicates shown here), and the same amount of signal is seen when cells

are spiked into buffer (rather than blood) and directly analyzed (without CTC-Chip

processing). Thus, there is virtually no loss of signal when a tumor cell is diluted into

20 billions of normal blood cells and then "re-isolated" using the CTC-Chip prior to

digital readout.

Example 4 - Multiplex Digital Analysis of Gene Transcripts from CTC-Chip Product

25 We established a multiplex assay capable of testing many different gene

transcripts from a minute amount of CTC-Chip product. This combined the higher

sensitivity and specificity of using multiple independent genes, with the fact that the

amount of input template is limited (and hence should not be diluted into multiple

reactions). The new assays include multiple genes, e.g., 2, 3, 4, 6, 8, 10, or more

30 genes per reaction, with each gene being resolved uniquely in 2-dimensional space by

selecting different ratios of fluorescent conjugated primers. Thus, in a single reaction,

one can independently measure 2, 3, 4, or more gene transcripts without having to

dilute the template. For different cancers, one can run and analyze multiple different

reactions (e.g., up to 20 different gene transcripts in four runs), and with application of nested RT-PCR digital assays, there is no limit to the number of reactions that can be performed.

To display the results of the multiplex reaction in a single view (and hence differentiate amplification of each gene is isolation), we varied the concentrations of the two fluorescent probes (FAM and HEX). By doing this, each individual gene amplification reaction has a unique combination of FAM/HEX signal that reflects the composition of the gene-specific primers, and hence identifies the gene-specific PCR product. In 2-dimensional space, we can illustrate the signal position of 4 different gene amplification products produced from a single multiplex reaction. As applied to digital PCR using droplets to encapsulate each PCR reaction, this method separates the targets into individual clusters by modifying the binary signal amplitude of positive droplets, which are displayed quantitatively. As predicted, this method allows both cumulative scoring of total signal for multiple genes (e.g., 16 markers in a total of 4 reactions), while also retaining the ability to quantify the signal from each individual gene target.

Probe 1: 100% FAM

Probe 2: 100% HEX

Probe 3: Mixture of FAM and HEX – sum up to 100%

Probe 4: Mixture of FAM and HEX – sum up to 100%

As shown in Tables 3 to 7, the following probe mixtures were used in the multiplex reactions:

TABLE 3
Multiplexing primers against 4 genes per reaction
(Melanoma)

Reaction 1

FAM	HEX	Primer	FAM Int	HEX Int
100%	0	Sox10	6000	0
70%	30%	SFRP1	4000	2500
30%	70%	RND3	4500	5500
0%	100%	TFAP2C	0	6000

Reaction 2

FAM	HEX	Primer	FAM Int	HEX Int
100%	0	PRAME	11000	0
70%	30%	MLANA	8000	4000
30%	70%	MAGEA6	5000	6000
0%	100%	PMEL	0	5500

Reaction 3

FAM	HEX	Primer	FAM Int	HEX Int
100%	0	PMEL	7000	0
70%	30%	MLANA	6000	3000
30%	70%	MAGEA6	4000	5000
0%	100%	MET	0	4500

TABLE 4

**Multiplexing primers against 4 genes per reaction
(Pan-Cancer/lineage)**

FAM	HEX	Primer	Exp. FAM Int	Exp. HEX Int
100	0	TFAP2C	9000	0
60	40	PGR	5100	1800
35	65	SCGB2A1	2205	7800
0	100	CADPS2	0	5000

5

TABLE 5

**Multiplexing primers against multiple genes per reaction
(AR status in Prostate)**

**Multiplexing primers against 4 genes per reaction
(Prostate)**

Reaction 1

FAM	HEX	Primer	Exp. FAM Int	Exp. HEX Int
100	0	TMPR2	5500	0
65	35	FAT1	5525	1837.5
40	60	KLK2	2440	2580
0	100	STEAP2	0	4300

Reaction 2

FAM	HEX	Primer	Exp. FAM Int	Exp. HEX Int
100	0	KLK3	6600	0
70	30	HOXB13	4340	1320
50	50	AGR2	4050	3050
0	100	FOLH1	0	5200

TABLE 6

Multiplexing primers against 4 genes per reaction Epithelial-Mesenchymal Transition (EMT)

Reaction 1

FAM	HEX	Primer	Exp. FAM Int	Exp. HEX Int
100	0	PKP3	8000	0
75	25	OCLN	6000	1625
40	60	CDH11	4000	3600
0	100	S100A2	0	5000

Reaction 2

FAM	HEX	Primer	Exp. FAM Int	Exp. HEX Int
100	0	FAT1	8000	0
65	35	FAT2	5200	1750
40	60	COL8A1	3200	3900
0	100	CDH3	0	6000

TABLE 7

Multiplexing primers against multiple genes per reaction

Reaction	Gene-primer set	Avg intensity (FAM)	Avg intensity (HEX)	AR status
1	TMPRSS2	5500		ON
1	FAT1	8500	5250	?
1	KLK2	6100	4300	ON
1	STEAP2	3350	4300	ON
2	KLK3	6600		ON
2	FOLH1	6200	5200	OFF
2	AGR2	8100	6100	OFF
2	HOXB13	6500	4400	OFF

Validation and Testing

To validate and demonstrate the effectiveness of this multiplex strategy, we illustrated both the concept (using spiked cell experiments) and patient-derived samples. FIG. 4 shows the results of processing a normal control blood sample from a healthy donor (HD) through the CTC-Chip and subjected to d-CTC assay for 4 different gene transcripts, all of which are negative (i.e., blank droplets).

On the other hand, FIG. 5 is a representation of data from spiked cell experiments, prostate cancer cell lines introduced into blood and processed through the CTC-Chip, followed by digital assay, showed positive signal (fluorescent droplets) for each of the 4 lineage transcripts. These appeared at separate locations within the 2-Dimensional plot, based on differential fluorescence of two probes (color coded in picture). As the sample is overloaded with tumor cells, some droplets contained signal from more than one gene transcript (multiple genes per droplet are shown in gray).

The strategy of representing four different genes within each reaction was applicable to multiple different cancers, with specific lineage markers substituted for each tumor type. For instance, in prostate cancer, we predicted (theoretical model) a multiplex reaction with four quadrants (one gene per quadrant) for each of 2 reactions (total of 8 gene markers). The spiked cell experiment (prostate cancer cells introduced into control blood and processed through the CTC-iChip) precisely recapitulated the predicted results.

Furthermore, FIGs. 6A-6B and FIGs. 7A-7B show that when assembled together, our analytic program integrated all positive signals within quadrants, just as predicted from modeling, and allowing us to develop methods to score the specific gene signals. Multi-dimensional space analysis of signal allowed for automated analysis and scoring with high level accuracy. FIGs. 6A and 6B show the theoretical model and actual results, respectively, for a prostate cancer cell line for Reaction 1, and FIGs. 7A and 7B show the theoretical model and actual results, respectively, for the same prostate cancer cell line for Reaction 2.

FIGs. 8A-8B (breast and lung cancer theoretical and actual results, Reaction 1), 9A-9B (breast and lung cancer theoretical and actual results, Reaction 2), 10A-10B (same, Reaction 3), 11A-11B (same Reaction 4), 12A-12B (same, Reaction 5),

and 13A-13B (same, Reaction 6) illustrate the results when the same approach was use with breast cancer and lung cancer. We can establish a multi-cancer panel that is effective in identifying markers shared by most adenocarcinomas (i.e., grouping breast and lung cancer togher), as 6 reactions (4 gene markers within each reaction for a total of 24 markers), as shown below (theoretical vs validation using spiked cell experiments with both breast and lung cancer cells).

These figures show the results when the same approach of testing multiple gene transcripts in multiplex fashion (4 genes per reaction) was applied to breast cancer. Six different reactions were performed of the same CTC chip product (enabling a total of 24 gene transcripts to be tested independently), with each one having a designated signal position (predicted in upper panel) and observed in spiked cell validation experiments (observed in lower panel).

Example 5 - Target-Specific Pre-Amplification to Improve Detection of Tumor-Specific mRNA

To improve the detection of tumor specific RNAs, a nested PCR strategy was optimized for each of the gene-specific amplifications. To achieve this, cDNA derived from the CTCs was first amplified with gene-specific primers which are situated a few base pairs external to the gene-specific primers used for d-CTC assay. For each gene, two to three primer sets were tested, and the primer set that is compatible with the gene-specific d-CTC assay primer and tests negative in HD blood was chosen for analysis of patient samples.

As described above, the target specific amplification protocol was first tested in cell lines derived from the different cancers. The primer combinations that are specific for tumor cells (and absent in leukocytes) were then tested with a mixture of cancer cell lines mixed into blood and enriched through the CTC-iChip. HD blood processed through the CTC-iChip was used as control. Key to this strategy is the design of the nested PCR conditions to enhance the signal from minute amounts of CTC-derived cDNAs, without increasing the minimal baseline signal from normal blood cells. This selectivity was achieved by careful optimizing of PCR primer sequences and assay conditions, as well as balancing the cycle number for the external and internal PCRs. All conditions are validated first with purified nucleic acids, then

with individual tumor cells that are spiked into control blood samples and processed through the CTC-iChip, then with large panels (> 10) of different healthy blood donors, and ultimately with patient-derived blood samples from patients who have either metastatic or localized cancers of the prostate, breast, melanoma, liver, lung or pancreas.

Reagents

- DNA Suspension Buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA) (TEKnova, PN T0221)
- 0.5 EDTA, pH 8.0 (Invitrogen, PN Am9260G)
- TaqMan PreAmp Master Mix (Applied Biosystems, PN 4391128)
- Nuclease-free Water (TEKnova, PN W330)

Preparing 10X Specific Target Amplification (STA) Primer Mix

- 1.) In a DNA-free hood, 0.5 μ L of each of 200 μ M primer pairs (0.5 μ L Forward primer and 0.5 μ L Reverse primer) were mixed.
- 2.) Each primer was diluted in 1X DNA Suspension Buffer to a final concentration of 500 nM. (Ex: If pooled primer volume equals 8 mL, add 192 mL DNA Suspension Buffer)
- 3.) The mix was vortexed for 20 seconds and spun down for 30 seconds.
- 4.) 10X STA Primer Mix can be stored at 4 °C for repeated use for up to six months or stored frozen at -20 °C for long-term usage.

Preparing STA Reaction Mix

- 1.) For each well of a 96-well PCR plate, prepare the following mix.

Component	Per 9 μ L Sample (μ L)	96 Samples with overage (μ L)
TaqMan® PreAmp Master Mix	7.5	780.0
10X STA Primer Mix (500 nM)	1.5	156.0
0.5M EDTA, pH 8.0	0.075	7.8
Total Volume	9.0	943.8

- 2.) 6 μ L cDNA was added to 9 μ L STA reaction mix
- 3.) Thermocycling conditions listed below were used with 18 cycles of

denaturation and annealing/extension steps rather than 20 cycles. (Note: 18 cycles were used to compare TSA Pre-Amplification protocol to Whole Transcriptome Amplification).

Condition	Enzyme Activation	10 to 18 Cycles		Hold
		Denaturation	Annealing/Extension	
Temperature	95°C	96°C	60°C	4°C
Time	10 minutes	5 seconds	4 minutes	Infinity

5 1 µl of the pre-amplified product is loaded in each droplet PCR reaction.

FIG. 14 shows the droplet PCR signal for 7 markers (PIP, PRAME, RND3, PKP3, FAT1, S100A2, and AGR2) from 1 ng of non-amplified cell-line cDNA and from 1 µl of pre-amplified product after 10, 14, and 18 cycles of pre-amplification. Additional cycles of pre-amplification result in signal increase. Of note, PRAME, a
10 marker expressed at very low levels in this cell line is detected only after 18 cycles of pre-amplification, demonstrating the utility of the technique.

Example 6 - Clinical Data and Assay Validation

The assays described herein have been validated using actual patients samples from clinical studies. These include patients with metastatic cancer (lung, breast,
15 prostate and melanoma), as well as patients with localized cancer (prostate). The assays are conducted as described in Examples 2 through 5.

FIGs. 15A, B, and C show a summary of clinical assays from patients with metastatic cancers of the lung (6 patients; FIG. 15A), breast (6 patients; FIG. 15B) and prostate (10 patients; FIG. 15C) showed that virtually all patients have positive
20 signal, whereas healthy controls have none. In this assay, all positive scores were added (cumulative score). However, as described below, the scores can also be broken down by individual genes, as shown in FIG. 16.

FIG. 16 illustrates the cumulative analysis of data from multiple probes, and shows a positive signal in 10/11 metastatic prostate cancer patients (91% on a per
25 patient basis) versus 0/12 (0%) of healthy controls. On a per sample basis, 24 of 28

samples had a positive signal, indicating an 86% detection rate. In addition, some individual markers were also fairly effective, e.g., AGR2 (9/10 detection for metastatic cancer, and 0/3 for localized cancer), TMPRSS2 (5/10 and 1/3), KLK2 (6/10 and 0/3), STEAP2 (1/10 and 1/3), FAT1 (2/10 and 1/3), and FOLH1 (3/10 and 1/3)

5 As illustrated above, one can also break down the individual gene markers for independent validation and quantitation, using the multiplex fluorescence color scheme described above. In this example below, a patient with metastatic prostate cancer had multiple positive markers, a patient with localized prostate cancer has a smaller number of positive scores within fewer markers, and a healthy control is
10 negative for all markers.

FIG. 17 shows clinical data from three representative patient samples. In two separate reactions with four gene transcripts each (8 probes total), a blood sample from a patient with metastatic prostate cancer showed multiple signals (all probes are positive to various degrees). In contrast, a blood sample from a patient with localized
15 (curable) prostate cancer showed weaker (but clearly detectable) signal. Whereas probes 1 (TMPRSS2), 5 (KLK3), 6 (HOXB13), 7 (AGR2) had the strongest signal in the metastatic cancer patient, probes 2 (FAT1) and 4 (STEAP2) were most positive in the localized cancer patient. This result clearly illustrates the heterogeneity in signal among cancer cells in the blood and the importance of dissecting the differential
20 signals within the assay. Blood from a HD control (processed identically to the cancer patient samples) had a complete absence of signal.

Example 7 – Measurement of Signaling Pathways within CTCs

In addition to providing a digital (quantitative) measure of CTCs present
25 within a blood sample, our d-CTC assay also allowed analysis of specific signaling pathways that are unique to the tumor cells in the blood. For instance, a subset of prostate lineage-specific genes were driven by androgen signaling (such as PSA), while another subset was repressed by androgen signaling (such as PSMA). By analyzing these genes together, we can ascertain the status of androgen signaling
30 within CTCs. Defining the total number of CTC signal in the blood, simultaneously with information about the effectiveness of the therapeutic agent in targeting and shutting off the critical pathway is important for therapeutic monitoring.

We have illustrated this concept in prostate cancer, where the anti-androgenic agent abiratorone is effective in suppressing cancer progression, particularly in tumors that are still dependent on the androgen pathway. Below, we showed the results of a patient with “Castrate Resistant Prostate Cancer (CRPC)” who is no longer responding to first line leuprolide and was treated with abiratorone. The androgen response markers (green) were initially suppressed by the therapy as it shows initial efficacy, but subsequently returned as the tumor becomes resistant and the patient experiences disease progression on this drug.

FIG. 18 provides the results of a clinical study of a patient with metastatic prostate cancer. The subset of signals from “androgen receptor-induced genes (AR-On)” is shown in green at the top of the bars in this bar graph, while the subset of signals from “androgen-repressed genes (AR-Off) is shown in red at the bottom of each bar. As the patient is treated with the androgen pathway inhibitor abiratorone (e.g., ZYTIGA® (abiraterone acetate), the AR-On signal is greatly reduced, indicating effective suppression of the androgen pathway within cancer cells in the blood. By cycle 4 of drug treatment, however, the androgen pathway appears to be reactivated in cancer cells (increasing green signal), indicative of drug resistance. Serum PSA measurements taken at these time points are consistent with failure of drug treatment.

Example 8 – Non-Specific Pre-Amplification to Improve Detection of Tumor-Specific mRNA

Similar to Example 5, non-specific whole transcriptome amplification (WTA) can be used to increase the detection rate of CTC-specific transcripts. This method relies on the use of random primers that amplify not only the targets of interest but all messages found in the product. In this example, the SMARTer™ Ultra Low RNA kit protocol (Clontech) was used as described below:

Transfer RNA to PCR tubes or plate

- 1) Add 1 uL of 1:50,000 diluted ERCC Spike-In Mix 1 to each sample
- 2) Bring the volume of each sample up to 10 uL
- 3) Add 1 uL of 3' SMART CDS Primer IIA to each sample
- 4) Run “72C” thermocycler program:

72°C 3 min

4°C forever

First Strand cDNA Master Mix (FSM):

- 5 1x 4 uL 5x First-Strand Buffer
 0.5 uL DTT
 1 uL dNTP Mix
 1 uL SMARTer IIA Oligonucleotide
 0.5 uL RNase Inhibitor
 10 2 uL SMARTScribe RT
 9 uL per sample
 5) Prepare the 10% excess FSM for your sample number, then add 9
 uL of FSM to each sample and pipet to mix
 6) Run “cDNA” thermocycler program:
 15 42°C 90 min
 70°C 10 min
 4°C forever

Second Strand Synthesis and Amplification (SSM):

- 20 1x 25 uL 2x SeqAmp PCR Buffer
 1 uL Primer IIA – v3
 1 uL SeqAmp DNA Polymerase
 3 uL Nuclease-free water 30 uL per sample
 7) Prepare the 10% excess SSM for your sample number, then add 30
 25 uL of SSM to each sample and pipet to mix
 8) Run “PCR” thermocycler program:
 95°C 1 min
 X cycles
 98°C 10 sec
 30 65°C 30 sec
 68°C 3 min
 72°C 10 min

4°C forever

The number of cycles can be adjusted depending on RNA input (e.g., 18 cycles for single cells or 9 cycles for 10 ng of RNA input). In addition, the 4 degree stopping point is overnight.

5

Solid Phase Reversible Immobilization (SPRI) Purification:

Transfer PCR product to lo-bind 1.5 mL Eppendorfs and label a second set of tubes with sample IDs; run the SPRI protocol at RT until the final elution

9) Incubate AMPure™ XP beads [4 deg] at RT for at least 30 minutes

10) Ensure that a sufficient amount of Elution Buffer is thawed and at RT

11) Make 80% ethanol (at least 400 uL per sample)

12) Vortex beads well before adding 50 uL of beads to each sample, pipetting up and down 5-10 times to mix well. Note: When pipetting beads, it's advisable to use RPT tips for better control of the volumes added and less residual bead binding in the tips

13) Incubate samples at RT for 5 minutes

14) Place samples on the magnet and let sit for 5 minutes

15) Pipet out the supernatant (~95 uL) without disturbing the beads (check for brown color in the pipet tip and put back in tube if there's a significant amount of bead loss)

16) Wash twice with 200 uL of 80% ethanol – do not mix or disturb the bead pellet. Simply submerge the bead pellet in the ethanol for 30 seconds and then remove the ethanol. Try not to let the bead pellet dry between ethanol washes.

17) Air-dry the samples on the magnetic rack until the bead pellets are no longer shiny but before they crack. Pipet off any residual ethanol that pools at the bottom while drying (Note: The drying time can vary greatly depending on the DNA concentration after amplification). Single-cell level RNA inputs generally take 3-5 minutes to dry, while other IFD product samples have taken up to an hour.

- 18) Elute pellets in 17 uL of Elution Buffer as they begin to crack. Remove a sample from the magnet and pipet the buffer over the pellet repeatedly until all of the beads are in solution; then pipet mix to fully resuspend the beads (this will work to varying degrees for each sample). Try not to mix too vigorously as this creates many bubbles, which tends to decrease the attainable elution volume.
- 19) Let the resuspended samples incubate at RT for at least 2 minutes, then quick spin all of the samples.
- 20) Put the samples back on the magnetic rack for 5 minutes.
- 21) Pipet off ~15 uL of your eluted amplified cDNA and check for beads in the pipette tip. If beads are present, pipet the solution back over the bead pellet and let sit for ~ 1 minute before attempting another elution. Otherwise, store in a new lo-bind 1.5-mL Eppendorf, PCR tube, or 96-well PCR plate. Note: If you are repeatedly getting beads in the elution product, the only solution may be to decrease your aspiration volume to 14 uL or lower.

This whole transcriptome amplification (WTA) approach was first tested in cell lines derived from different cancers. FIGs. 19A and 19B show three different replicates of SMARTer-preamplified cDNA (18 cycles) from a liver cancer cell line (HEPG2) analyzed with 12 probes from the liver cancer panel. As shown in FIG. 19A, while the amplification efficiency for each target region is different, it is consistent among the three replicates (WTA1, WTA2, WTA3), demonstrating the reproducibility of this approach. As shown in FIG. 19B, these methods using 18 cycles of SMARTer pre-amplification provide an increase in signal of approximately four orders of magnitude (10^8 vs 10^4), providing a great boost in detection.

Example 9 – Multiplexed vs. Individual Marker Assays for Liver Cancer

For each sample, 10-20 mL of blood was collected from each patient. Blood was processed within 3 hours of arrival on a CTC-iCHIP running in negative depletion mode. RNA was extracted from the product using a Qiagen RNeasy™ plus Micro kit, and 5 uL of the available 17 uL amplified using ClonTech's v3 SMARTer™ whole-transcriptome amplification (WTA) strategy. 1% of the WTA

product was then loaded into each well of a digital PCR plate, and 500 nM Taqman™ primer/probe combinations used to determine the transcript concentration for each gene of interest. Transcript counts were normalized to blood volume and compared between HCC, HD, and CLD patients. HCC patients are defined as biopsy-confirmed non-resected hepatocellular carcinoma, CLD patients are patients with liver disease of varying etiologies (alcohol-mediated, HBV, HCV) who have negative ultrasound/MRI. HD are healthy donors external to the lab who donate 10-20 mL of blood.

FIGs. 20A to 20C show the total droplet numbers in 21 hepatocellular carcinoma (HCC) patients (FIG. 20A), 13 chronic liver disease (CLD) patients (FIG. 20B) and 15 healthy donors (HDs)(FIG. 20C). HCC patients show higher number of droplets compared to both CLDs and HDs, suggesting that the panel is very clean in the high risk CLD group and can be used to screen those patient for the development of liver cancer. This is an important result given the low specificity of screening methods c=for liver cancer currently available in the clinic. Among CLD patients the American Association of Liver Disease recommends ultrasound (US) every 6 months, with a detailed algorithm dependent on the size of liver lesion detected. A prospective combined AFP gene marker-ultrasound screening in China demonstrated a 37% mortality benefit for those who were screened compared to those who were not, even when the screened population only maintained a compliance rate of 60%.

The sensitivity and specificity of each assay are dependent on the threshold values chosen to define "diseased" vs. "non-diseased," but using 20 ug/L, the AFP gene marker has a sensitivity between 50-80% and a specificity between 80-90%. In a study using 20 ng/ml as the cut-off point, the sensitivity rose to 78.9%, although the specificity declined to 78.1% (Taketa, Alpha-fetoprotein, J. Med. Technol., 1989;33:1380). On the other hand, the overall detection rate of the present assay was 76% when taking into account the clinical history of the patients and correcting for the ones that received curative resection or liver transplant with 100% specificity.

In addition, while all 11 markers of the liver cancer assay used herein contributed to the 76% sensitivity, the top 5 markers (AHSG, ALB, APOH, FGB and FGG) by themselves have 70% sensitivity, while the top 3 markers alone (ALB, FGB, FGG) result in 67% sensitivity. ALB alone detected 56% of the cases.

Example 10 – Multiplexed vs. Individual Marker Assays for Lung Cancer

Blood samples from 8 metastatic lung patients and 8 healthy donors were processed through the CTC-chip as previously described. Samples were spun down, treated with RNeasyTM and stored at -80C. RNA was purified and cDNA was synthesized as described. STA was performed on each sample using 6 µl cDNA and the nested primers corresponding to the probes listed in the figure. 1 µl of STA product was loaded per each droplet PCR reaction.

Droplet numbers were normalized to blood volume. As shown in FIG. 21A and 21B, the multiplexed lung gene marker panel was able to detect 100% (8/8) metastatic lung cancer patient samples above the background of the 8 healthy donors. The sensitivity of each marker of the lung panel was also determined and the results show that SFRP had a detection rate of 8/8, FAT1 Probe 2 had a detection rate of 7/8, TMPRSS4 had a detection rate of 6/8, FOXF1 and ARG2, Probe 2 had a detection rate of 5/8, FAT1 had a detection rate of 4/8, FAT2 and AGR2 had a detection rate of 3/8, and FAT2, Probe 2 had a detection rate of 2/8.

Assays for SERPINA3 and SFRP2 indicated that SFRP2 is effective for both lung and breast cancer detection, whereas the former seems more specific for breast cancer detection, but also detects some lung cancer samples.

Example 11 – Multiplexed vs. Individual Marker Assays for Breast Cancer

Blood samples from 9 metastatic breast cancer patient, 5 localized breast cancer patients, and 15 healthy donors were processed through the CTC-Chip. Products were pelleted, treated with RNeasyTM and stored at -80C. RNA and cDNA from each sample were prepared as previously described. 6 µl cDNA from each sample was STA amplified using nested primers corresponding to the probes listed in FIG. 22 (FAT2, SCGB2A1, PGR, PRAME, TFAP2C, S100A2, FAT1, AGR2, PKP3, RND3, and PIP). Droplet numbers were normalized to blood volumes and the highest healthy donor value for each marker was subtracted from the patient sample values.

FIG. 22 shows the above-background signal for each patient. These methods detected 7/9 (78%) of metastatic samples and 2/5 (40%) of localized samples. The sensitivity of each marker alone varied from 1/14 to 6/14, with the two most relevant

markers being AGR2 (6/14) and FAT1 (5/14), and the next four most relevant markers being RND3, PKP3, PRAME, and SCGB2A1 (3/14 each).

Example 12 – AVR7 Detection in Metastatic Breast Cancer

5 Blood samples from 10 metastatic breast cancer patient and 7 healthy donors were processed through the CTC-Chip. Products were pelleted, treated with RNAlater™ and stored at -80C. RNA and cDNA from each sample were prepared as previously described. 6 µl of non-amplified cDNA were loaded into each droplet PCR reaction. The samples were analyzed with probes against the v7 isoform of the
10 androgen receptor (ARv7, sequence in Table 1). Droplet number was normalized to blood volume.

 As shown in FIG. 23A, ARv7 was detected in 5/10 patients (50%) at above background (HD) levels, demonstrating that the assay is successful at detecting ARv7 from liquid biopsy. One of the patients had a triple negative breast cancer, suggesting
15 utility of ARv7 as a marker even in the triple negative breast cancer (TNBC) context (e.g., patients who do not express genes for any of the three most common breast cancer markers, the estrogen receptor (ER), HER2/neu, and the progesterone receptor (PR) marker).

Example 13 – Multiplexed vs. Individual Marker Assays for Melanoma

20 Blood samples from 34 metastatic or unresectable melanoma patients, each with multiple draw points (total draw points: 182), and 15 healthy donors were processed through the CTC-Chip. Products were pelleted, treated with RNAlater™ and flash frozen at -80C. RNA and cDNA from each sample were prepared as
25 previously described. 12 µl cDNA from each sample was amplified by specific target amplification (10 cycles) using nested primers corresponding to the probes listed along the bottom of the graph in FIG. 24A (individual markers PMEL, MLANA, MAGEA6, PRAME, TFAP2C, and SOX10)). Droplet numbers were normalized to blood volumes. FIG. 24B shows a dot plot distribution of droplet signals detected in
30 melanoma patients as compared to healthy donors. The detection sensitivity was 81% for all patient draw points (a patient draw is scored positive if any 1 of 6 markers shows droplet signals above the highest background signal in HD for that particular

marker). Of the individual markers, PMEL and MLANA showed the highest detection rate.

OTHER EMBODIMENTS

- 5 It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method for detection of cancer with ultra-high sensitivity and specificity in a subject, the method comprising
 - isolating circulating tumor cells (CTCs) from a blood sample from the subject;
 - converting the CTC-derived RNA into cDNA;
 - encapsulating the cDNA into individual droplets;
 - amplifying the cDNA in each droplet in the presence of a reporter group configured to bind specifically to cDNA from CTCs and not to cDNA from other cells in the blood; and
 - determining a total number of droplets positive for the reporter group to determine a presence of CTCs indicating a presence of cancer in the subject.
2. A method of analyzing circulating tumor cells (CTCs) in a blood sample, the method comprising:
 - isolating from the blood sample a product comprising CTCs and other cells present in blood;
 - isolating ribonucleic acid (RNA) molecules from the product;
 - generating cDNA molecules in solution from the isolated RNA;
 - encapsulating cDNA molecules in individual droplets;
 - amplifying cDNA molecules within each of the droplets in the presence of one or more reporter groups configured to bind specifically to cDNA from CTCs and not to cDNA from other cells;
 - detecting droplets that contain the reporter groups as an indicator of the presence of cDNA molecules from CTCs in the droplets; and
 - analyzing CTCs in the detected droplets.
3. The method of claim 2, further comprising reducing a volume of the product before isolating RNA.
4. The method of claim 2 or claim 3, further comprising removing contaminants from the cDNA-containing solution before encapsulating the cDNA molecules.

5. The method of any one of claims 2 to 4, wherein generating cDNA molecules from the isolated RNA comprises conducting reverse transcription (RT) polymerase chain reaction (PCR) of the isolated RNA molecules.
6. The method of any one of claims 1 to 5, wherein amplifying cDNA or cDNA molecules within each of the droplets comprises conducting PCR in each droplet.
7. The method of any one of claims 2 to 6, wherein encapsulating individual cDNA molecules further comprises encapsulating PCR reagents in individual droplets with the cDNA molecules and forming at least 1000 droplets of a non-aqueous liquid.
8. The method of any one of claims 2 to 7, wherein the reporter groups comprise a fluorescent label.
9. The method of any one of claims 4 to 10, wherein removing contaminants from the cDNA-containing solution comprises the use of Solid Phase Reversible Immobilization (SPRI).
10. The method of claim 9, wherein the SPRI comprises
immobilizing cDNA in the solution with magnetic beads that are configured to specifically bind to the cDNA;
removing contaminants from the solution; and
eluting purified cDNA.
11. The method of any one of claims 7 to 10, wherein the non-aqueous liquid comprises one or more fluorocarbons, hydrofluorocarbons, mineral oils, silicone oils, and hydrocarbon oils.
12. The method of any one of claims 6 to 11, wherein probes and primers for use in amplifying the cDNA molecules within each of the droplets correspond to one or more probes and primers that relate to the selected cancer genes listed in Table 1.

13. The method of claim 12, wherein the selected cancer selective genes include prostate cancer-selective genes.
14. The method of claim 12, wherein the selected cancer genes include breast cancer-selective genes.
15. The method of claim 12, wherein the selected cancer genes include genes selective for one or more of lung cancer, pancreatic cancer, liver cancer, and melanoma.
16. The method of claim 12, wherein the selected cancer genes include one or more genes selective for two or more, three or more, four or more, or five or more different types of cancer.
17. The method of claim 16, wherein the genes are selective for breast cancer and lung cancer; breast cancer, lung cancer, and liver cancer; breast cancer, lung cancer, and pancreatic cancer; breast cancer, lung cancer, and prostate cancer; breast cancer, liver cancer, and melanoma; breast cancer, lung cancer, and melanoma; breast cancer, lung cancer, liver cancer, and prostate cancer; breast cancer, lung cancer, liver cancer, and melanoma; breast cancer, lung cancer, liver cancer, and pancreatic cancer; breast cancer, lung cancer, prostate cancer, and pancreatic cancer; breast cancer, lung cancer, liver cancer, melanoma, and pancreatic cancer; or breast cancer, lung cancer, liver cancer, melanoma, pancreatic cancer, and prostate cancer.
18. The method of any one of claims 1 to 17, wherein the CTCs arise from metastatic or primary/localized cancers.
19. The method of any one of claims 2 to 18, wherein analyzing the CTCs in the detected droplets comprises monitoring CTCs from blood samples taken over time from a patient with a known cancer, and testing, imaging, or both testing and imaging the CTCs to provide a prognosis for the patient.

20. The method of any one of claims 2 to 18, wherein analyzing the CTCs in the detected droplets comprises testing, imaging, or testing and imaging the CTCs from a blood sample from a patient to provide an indication of a response by the CTCs to a therapeutic intervention.

21. The method of any one of claims 2 to 19, wherein analyzing the CTCs in the detected droplets comprises determining a number or level of CTCs per unit volume of a blood sample from a patient to provide a measure of tumor burden in the patient.

22. The method of claim 21, further comprising using the measure of tumor burden in the patient to select a therapy.

23. The method of claim 22, further comprising determining the measure of tumor burden in the patient at a second time point to monitor the tumor burden over time, e.g., in response to a therapeutic intervention.

24. Use of probes and primers related to one or more selected cancer genes listed in Table 1 for amplifying and detecting cDNA molecules obtained from circulating tumor cells (CTCs) in a blood sample.

25. The use of claim 24, wherein the amplified CTCs are analyzed to detect cancer in a subject from whom the blood sample was obtained.

26. The use of claim 24, wherein the amplified CTCs are analyzed in multiple blood samples obtained over time from a patient with a known cancer, and testing, imaging, or both testing and imaging the CTCs to provide a prognosis for the patient.

27. The use of claim 24, wherein the amplified CTCs are analyzed to provide an indication of a response by the CTCs to a therapeutic intervention.

28. The use of claim 24, wherein the amplified CTCs are analyzed to provide a measure of tumor burden in a patient from whom the blood sample was obtained.

1/22

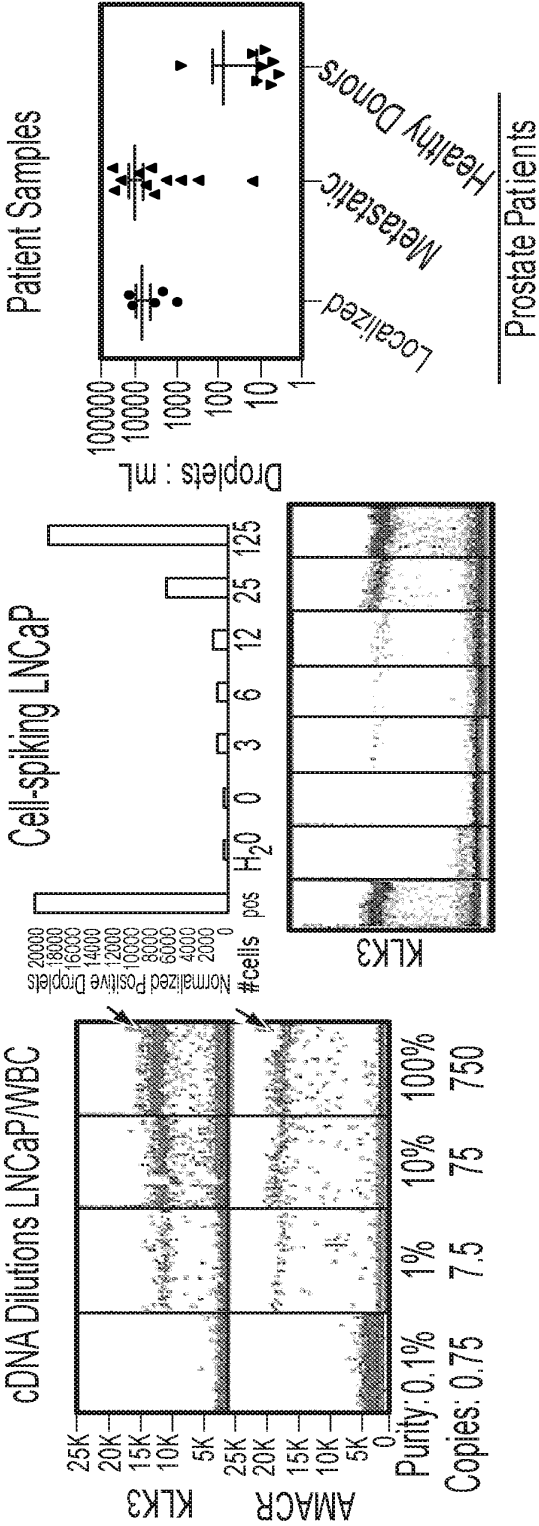


FIG. 1A

FIG. 1B

FIG. 1C

2/22

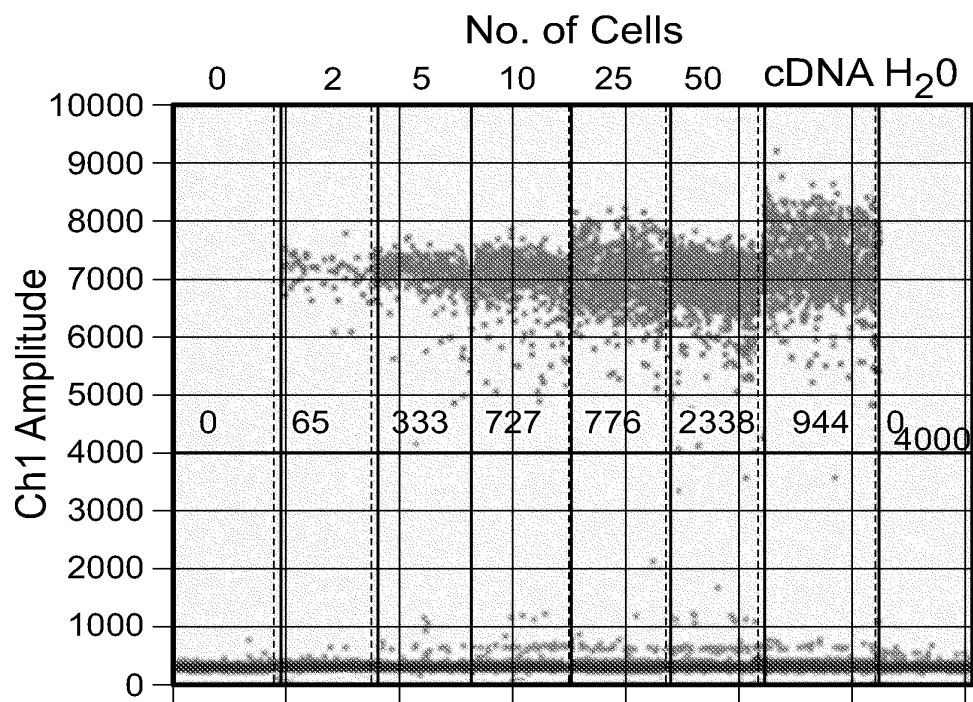


FIG. 2

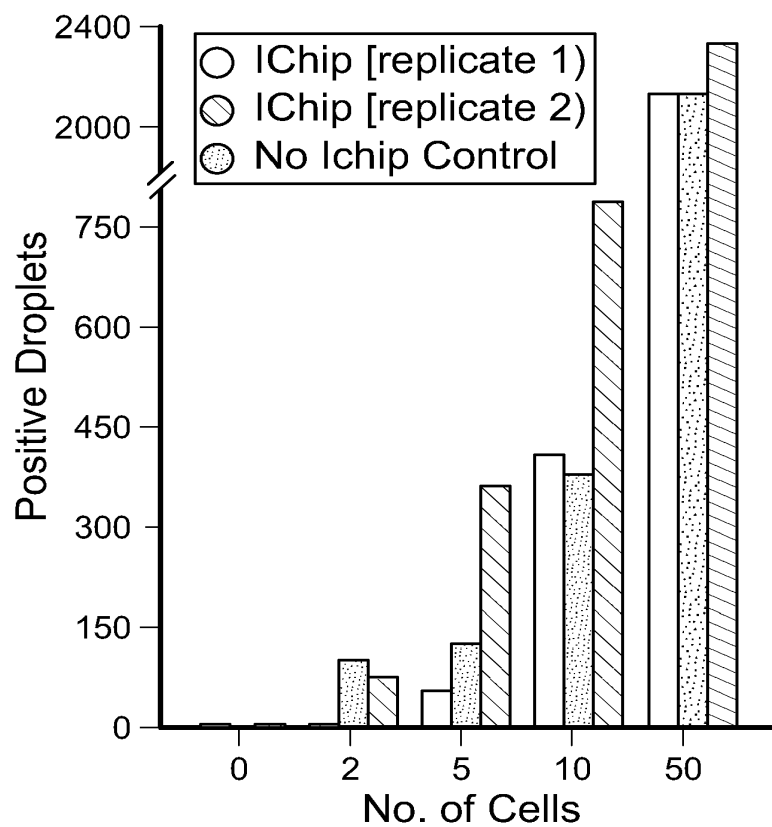


FIG. 3

SUBSTITUTE SHEET (RULE 26)

3/22

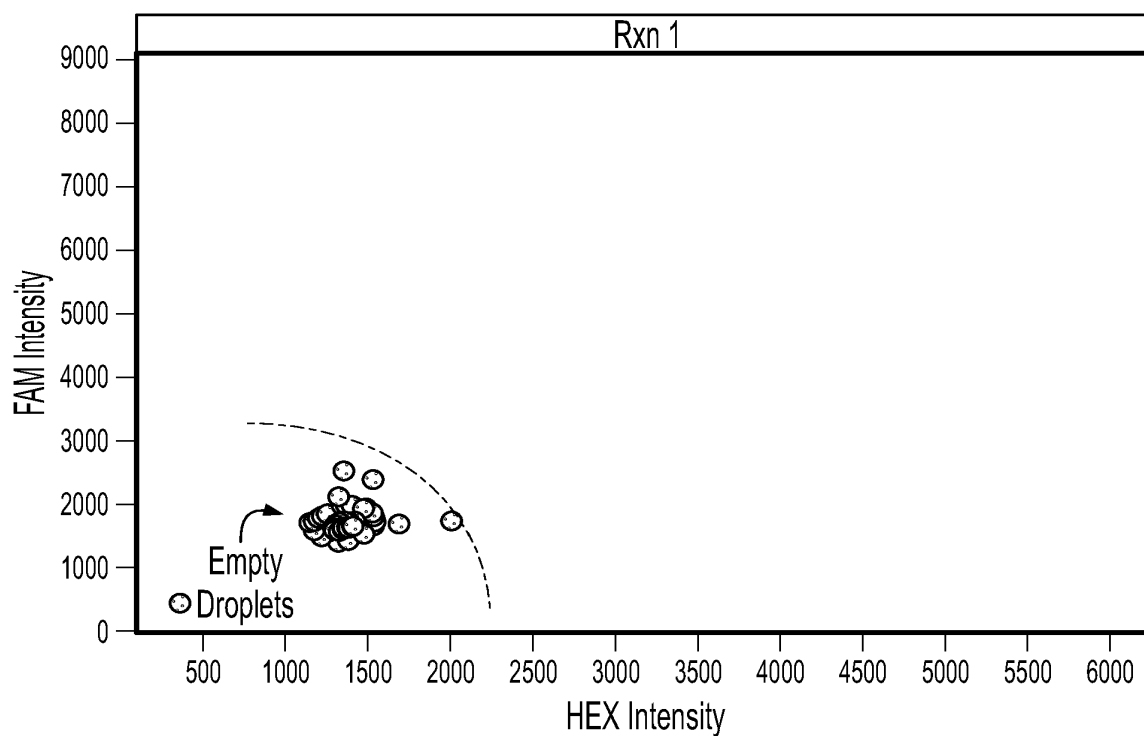


FIG. 4

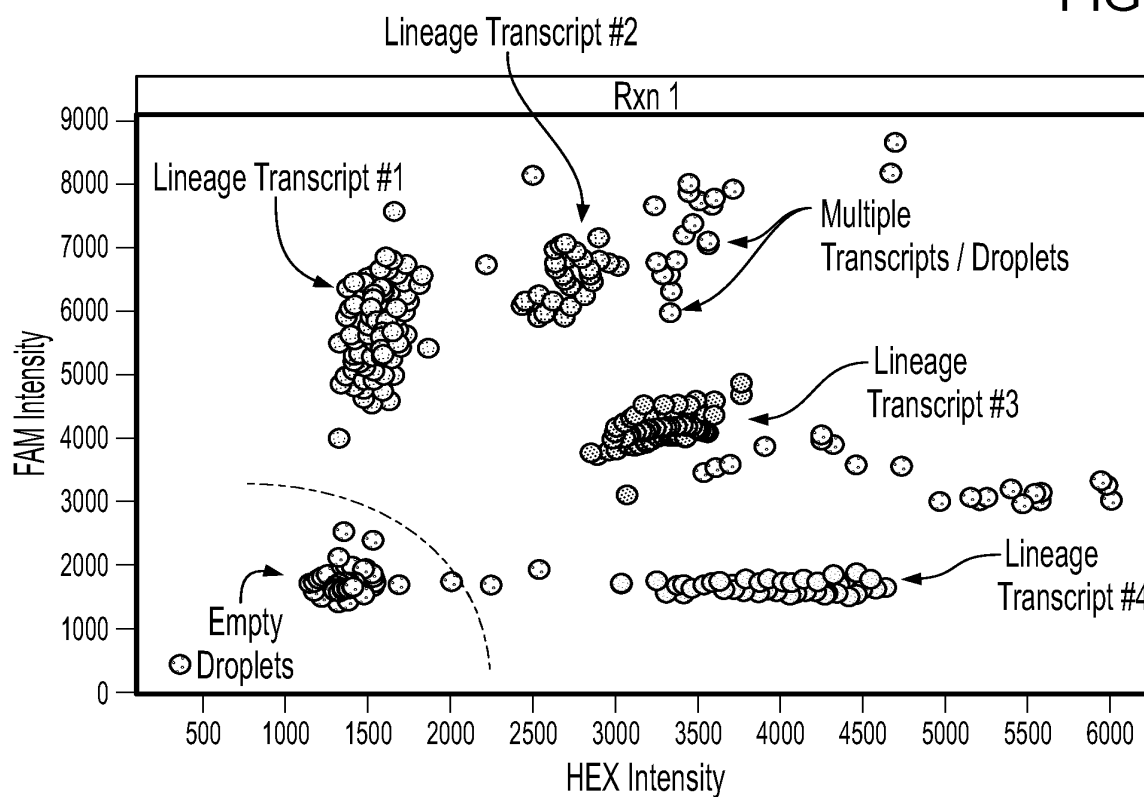


FIG. 5

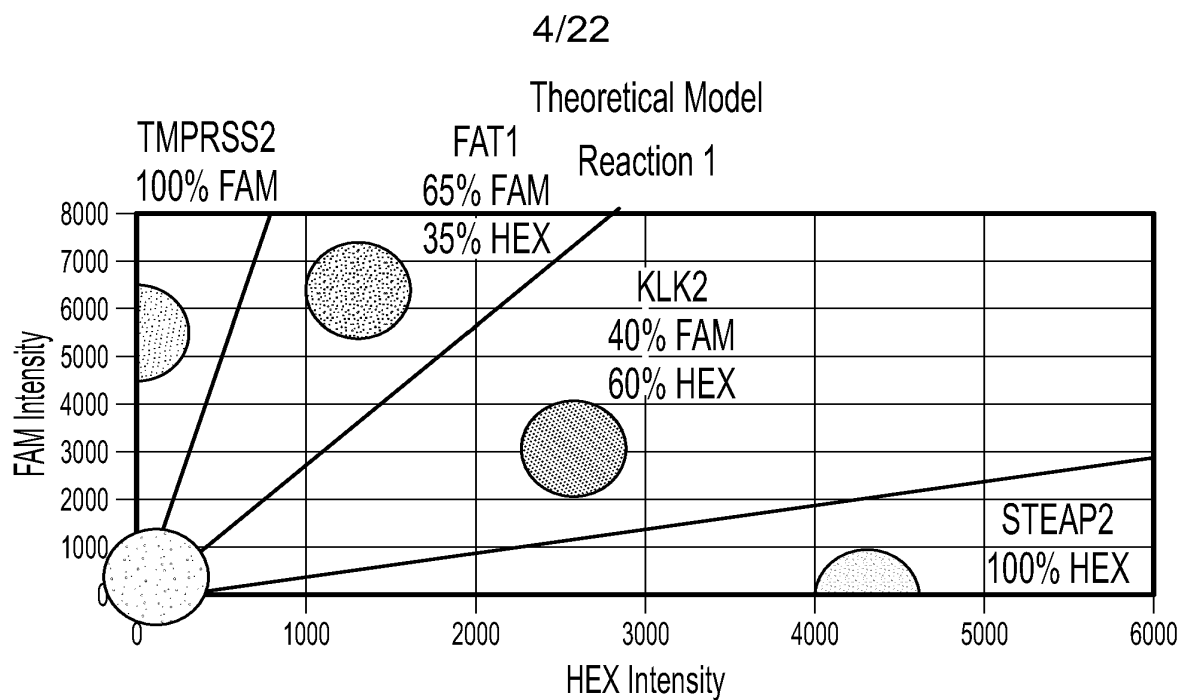


FIG. 6A

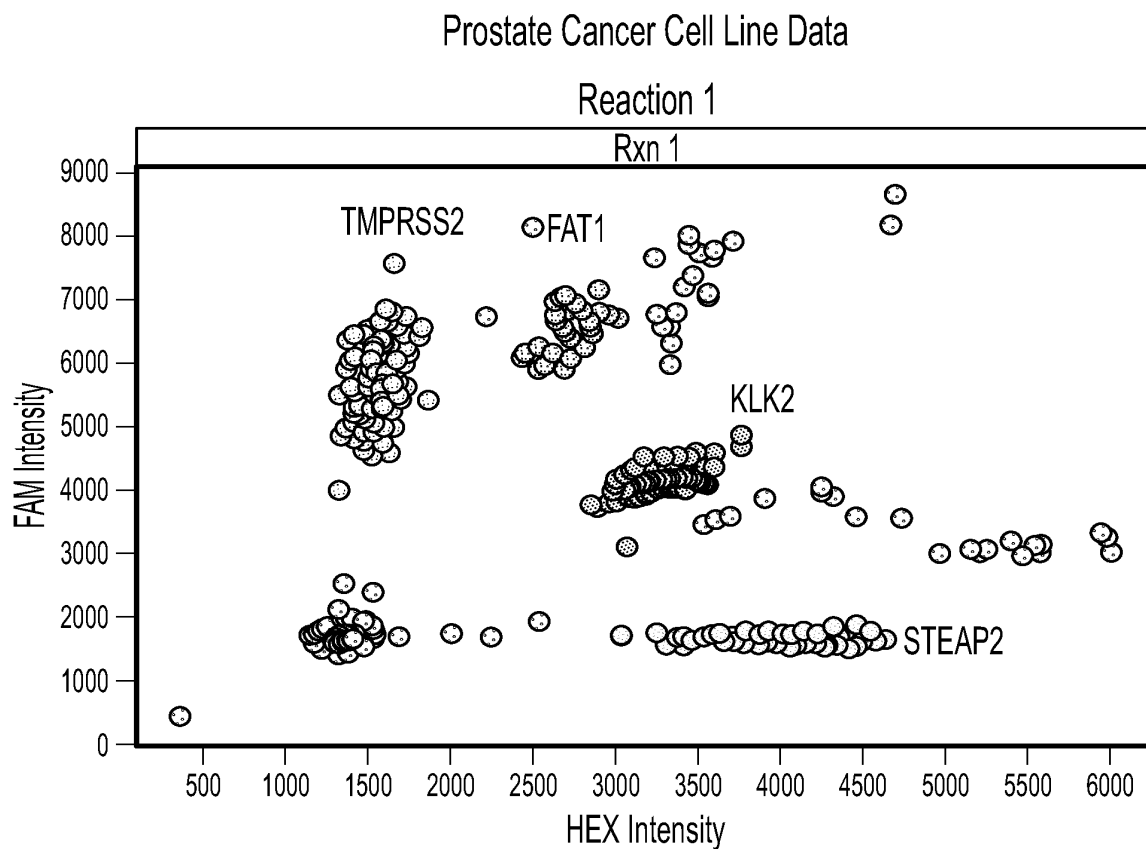


FIG. 6B

5/22

Theoretical Model

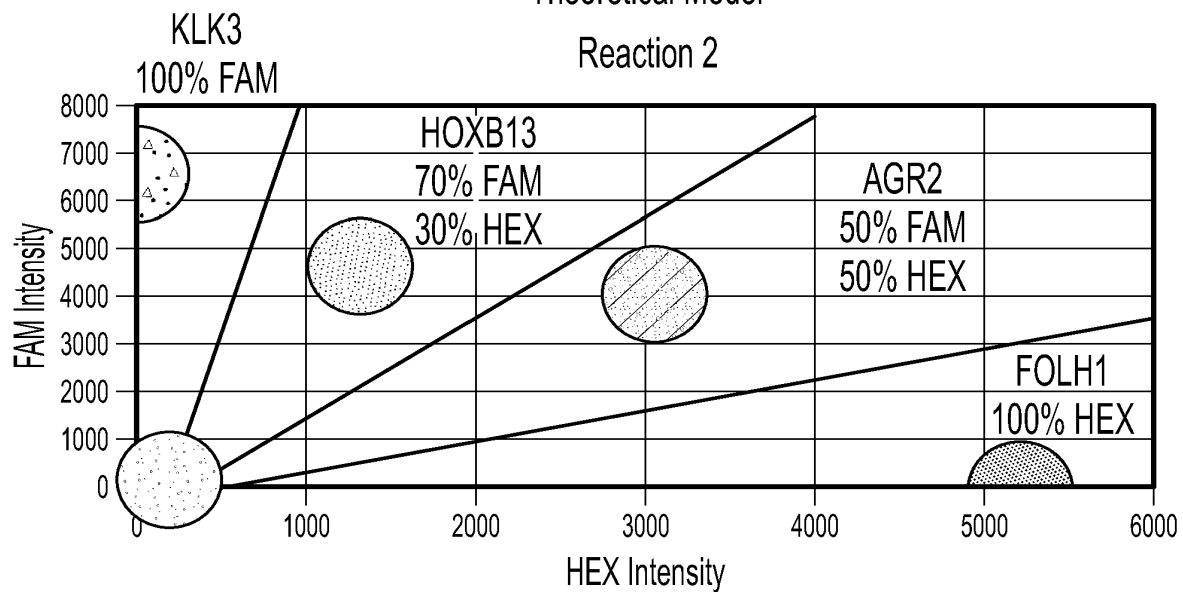


FIG. 7A

Prostate Cancer Cell Line Data

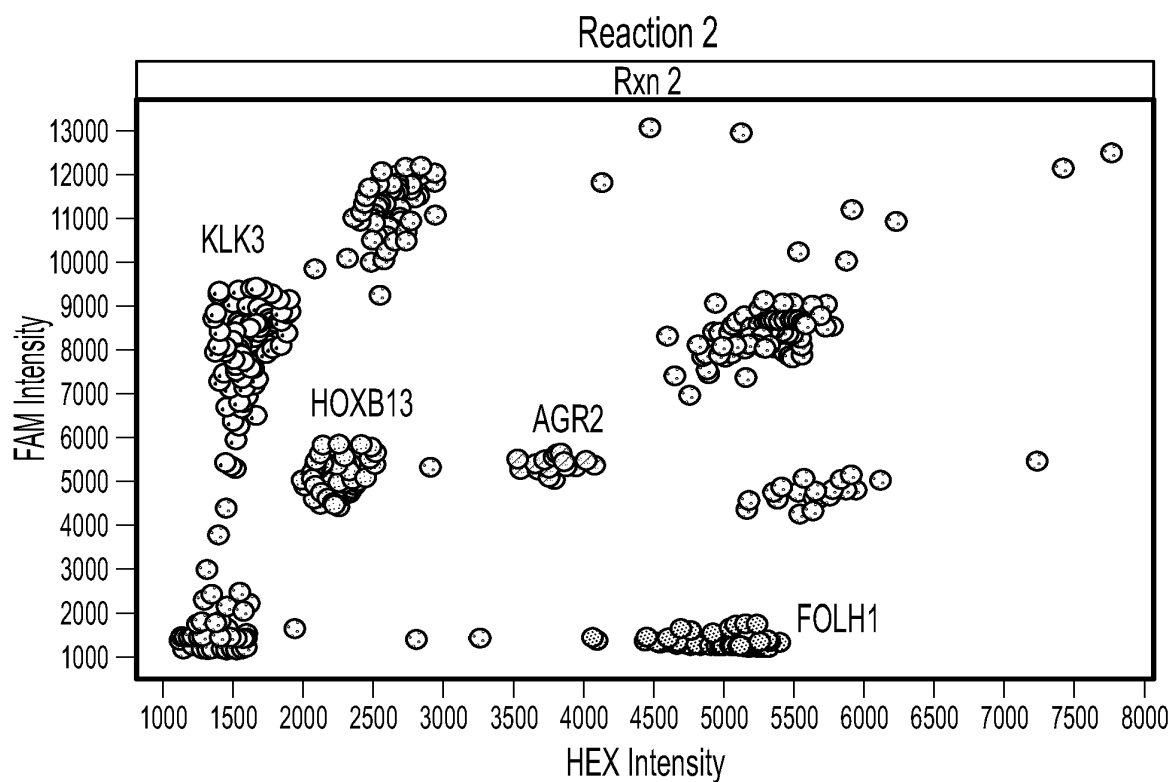


FIG. 7B

SUBSTITUTE SHEET (RULE 26)

IPR2024-01451

California Institute of Technology - Ex. 2003 Page 3346

6/22

Theoretical Model

Rxn 1

Pan-Cancer Rxn 1

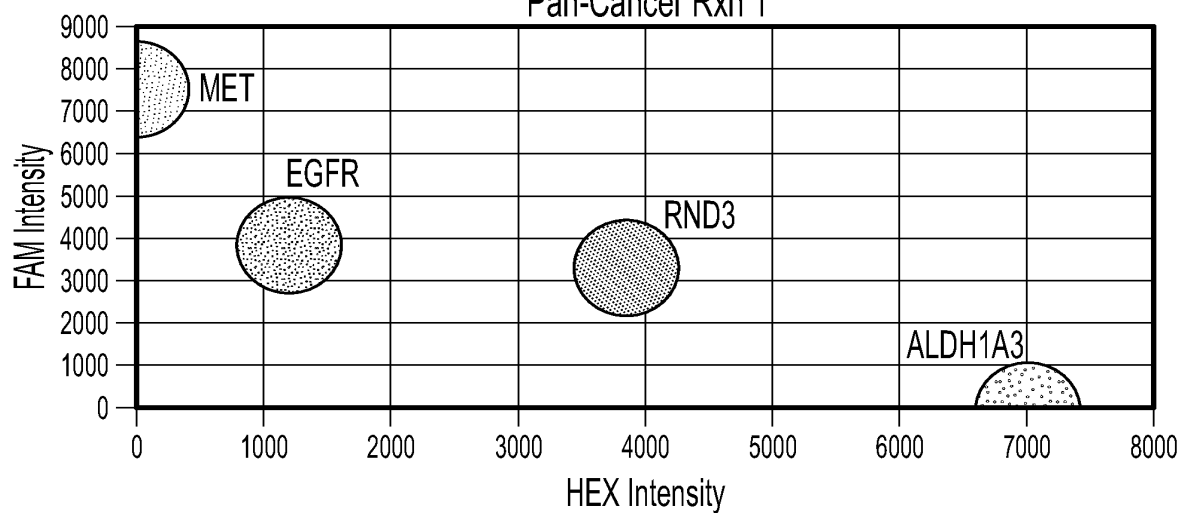


FIG. 8A

Breast/Lung Cancer Cell Line Experiment (Combo)

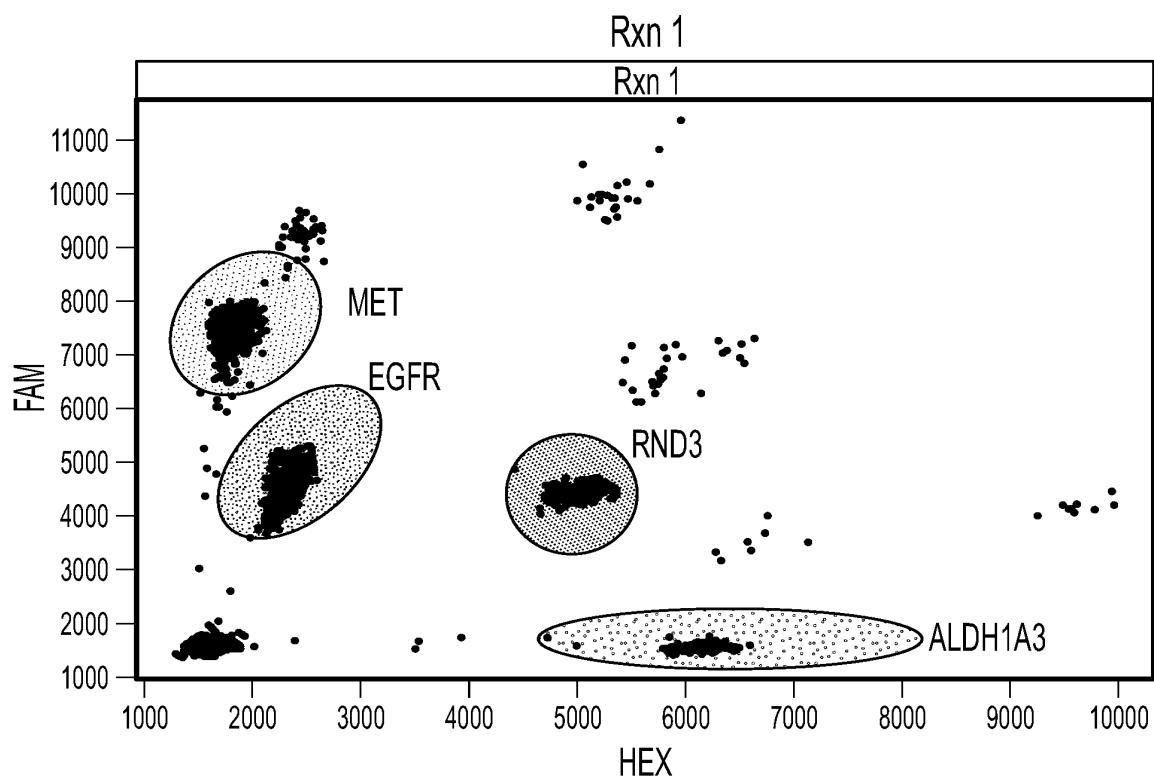


FIG. 8B

SUBSTITUTE SHEET (RULE 26)

IPR2024-01451

California Institute of Technology - Ex. 2003 Page 3347

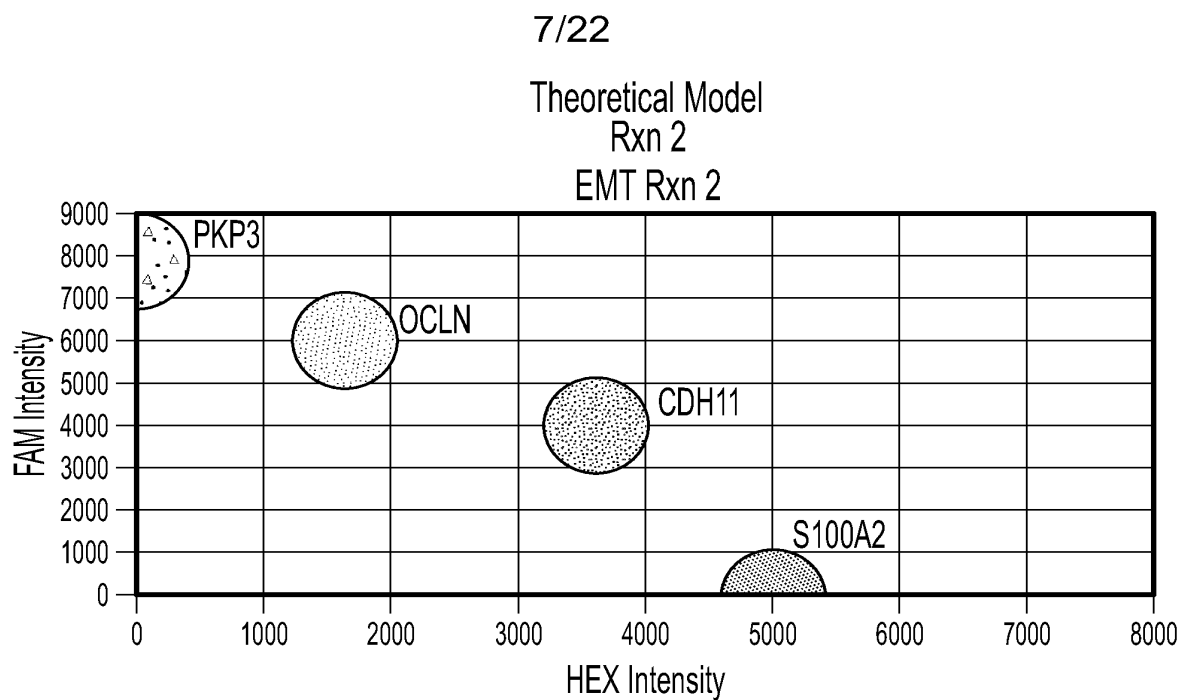


FIG. 9A

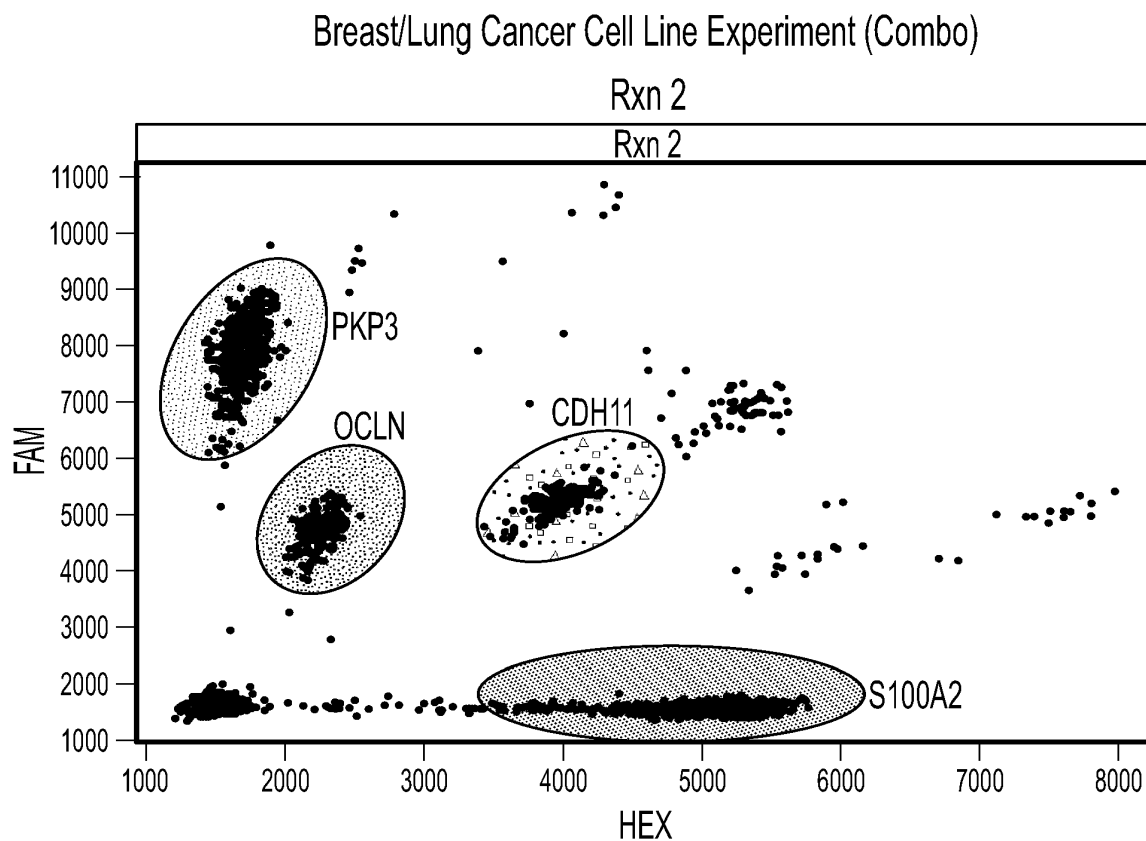


FIG. 9B

SUBSTITUTE SHEET (RULE 26)

8/22

Theoretical Model
Rxn 3
Pan-Cancer/EMT Cancer Rxn 3

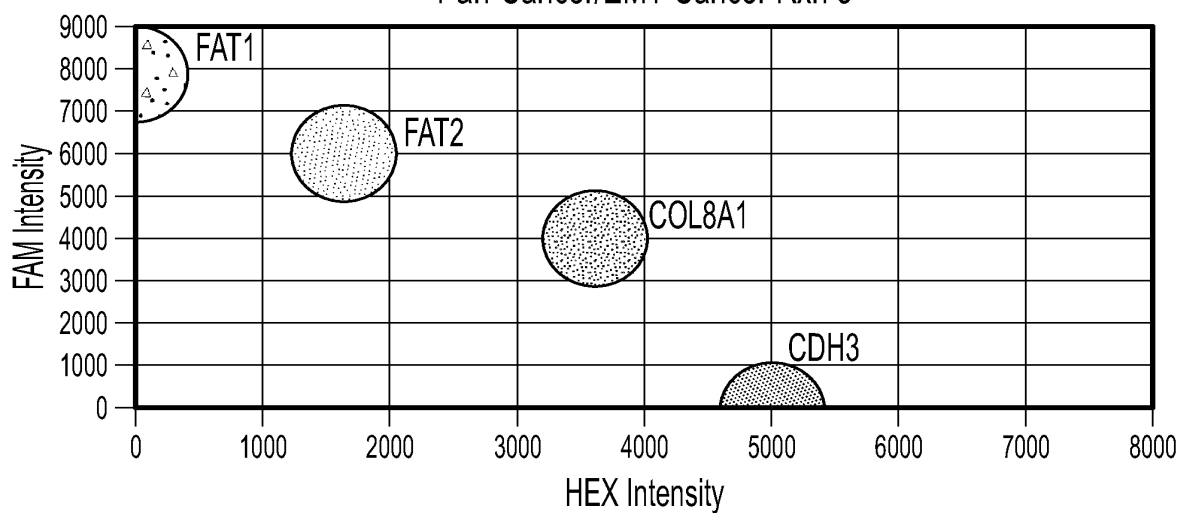


FIG. 10A

Breast/Lung Cancer Cell Line Experiment (Combo)

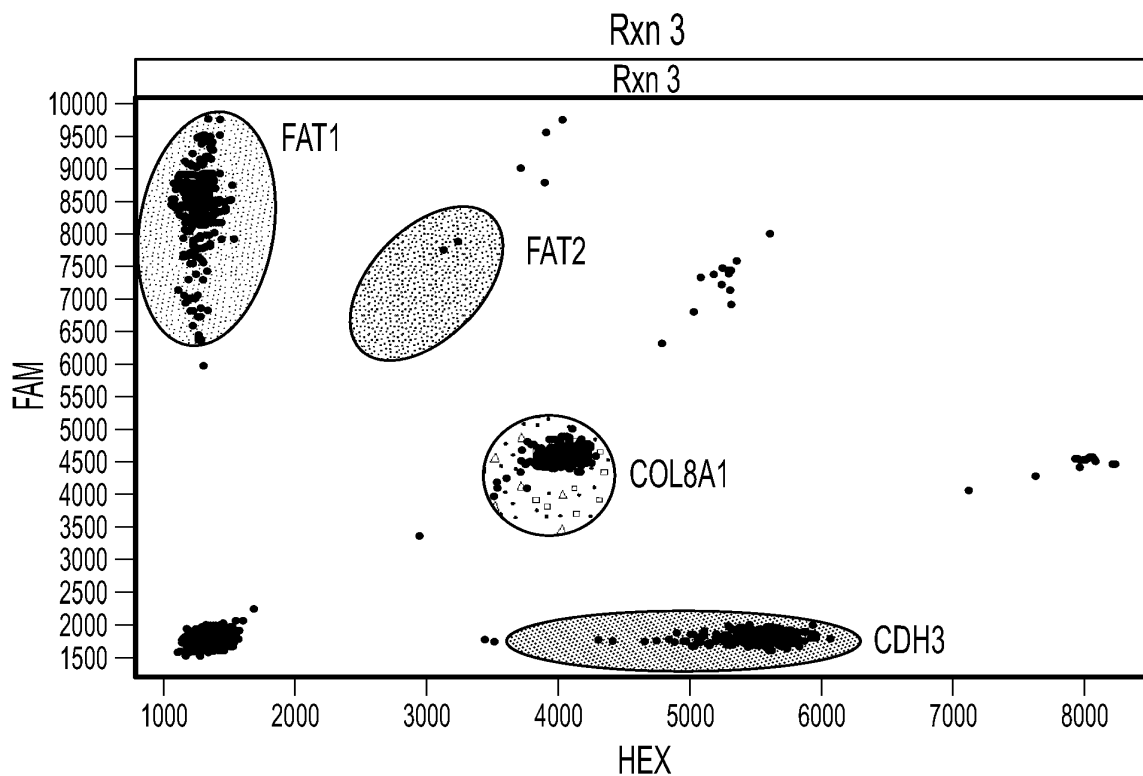


FIG. 10B

SUBSTITUTE SHEET (RULE 26)

9/22

Theoretical Model
Rxn 4
Pan-Cancer Rxn 4

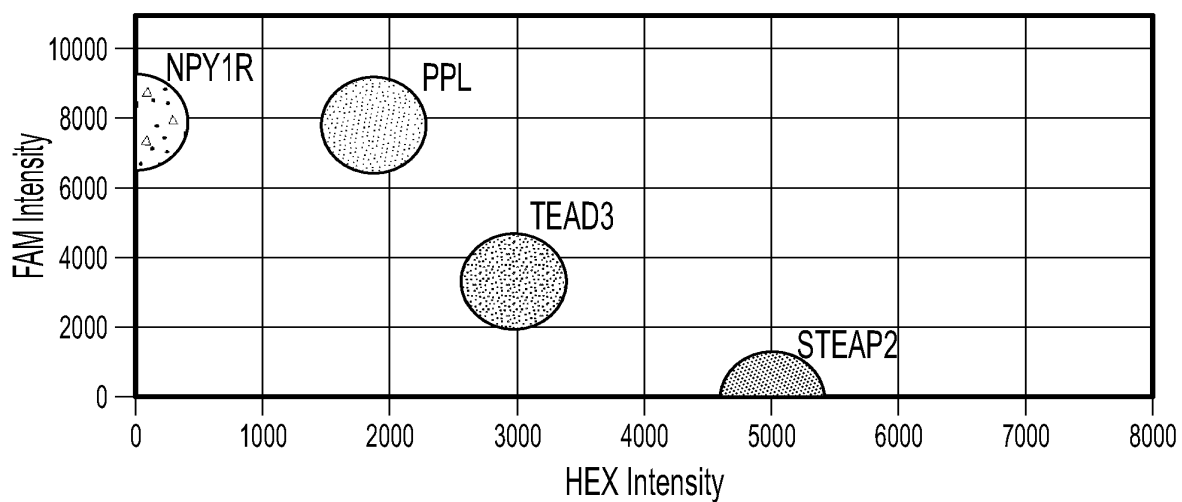


FIG.11A

Breast/Lung Cancer Cell Line Experiment (Combo)

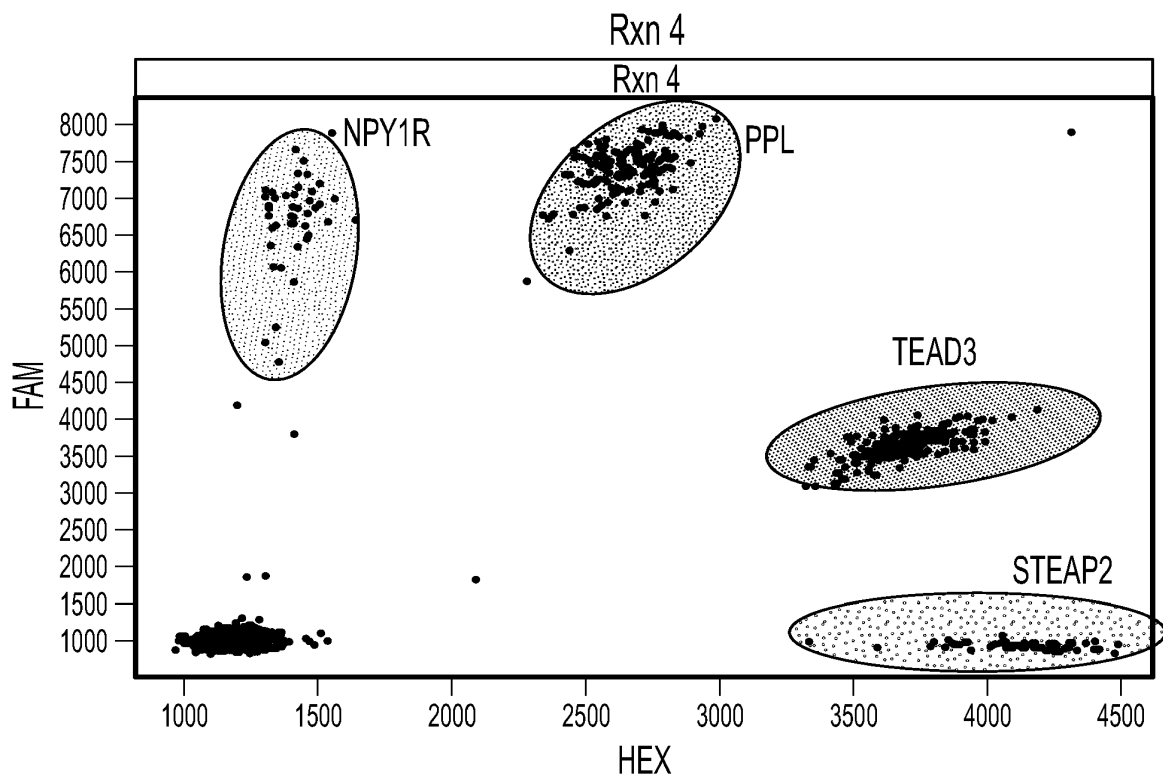


FIG. 11B

SUBSTITUTE SHEET (RULE 26)

10/22

Theoretical Model
Rxn 5
Pan-Cancer/Breast-Lineage

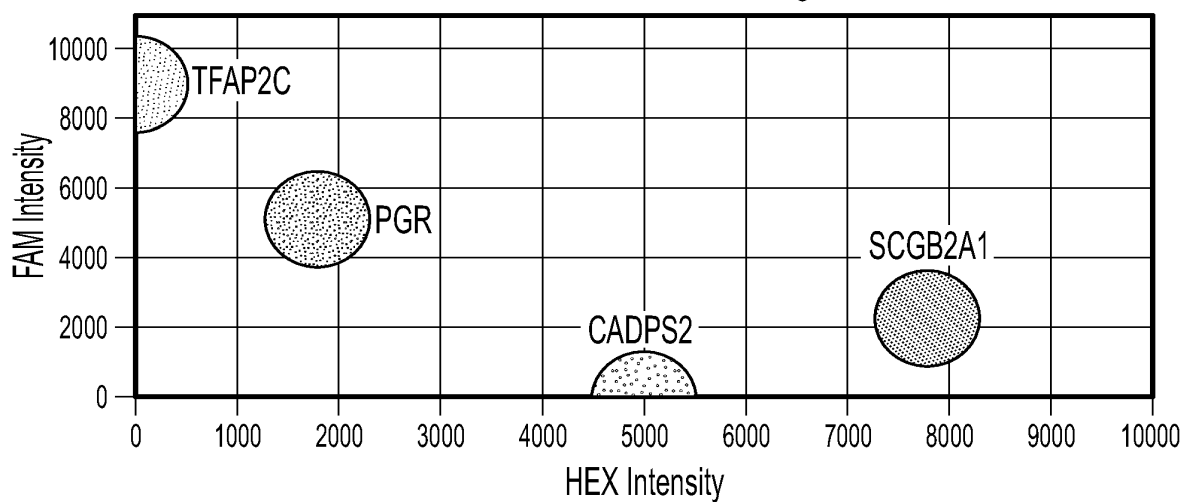


FIG.12A

Breast/Lung Cancer Cell Line Experiment (Combo)

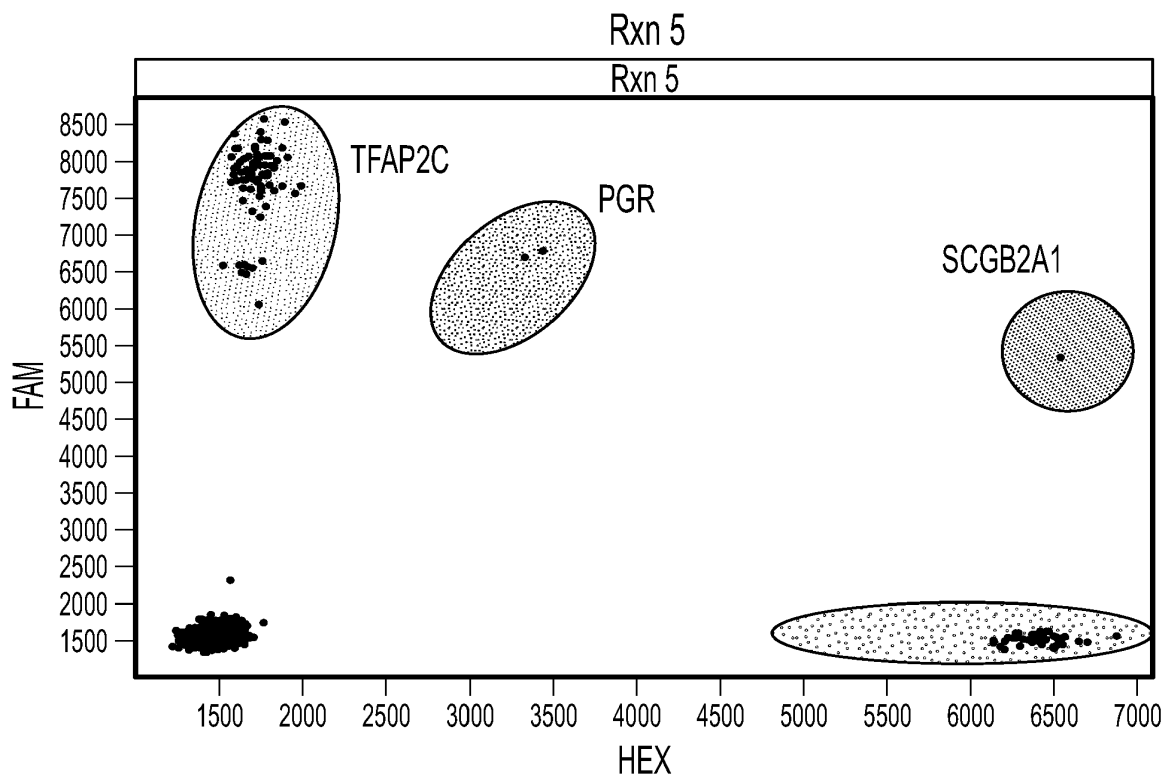


FIG. 12B

SUBSTITUTE SHEET (RULE 26)

11/22

Theoretical Model
Rxn 6
Pan-Cancer Rxn 6

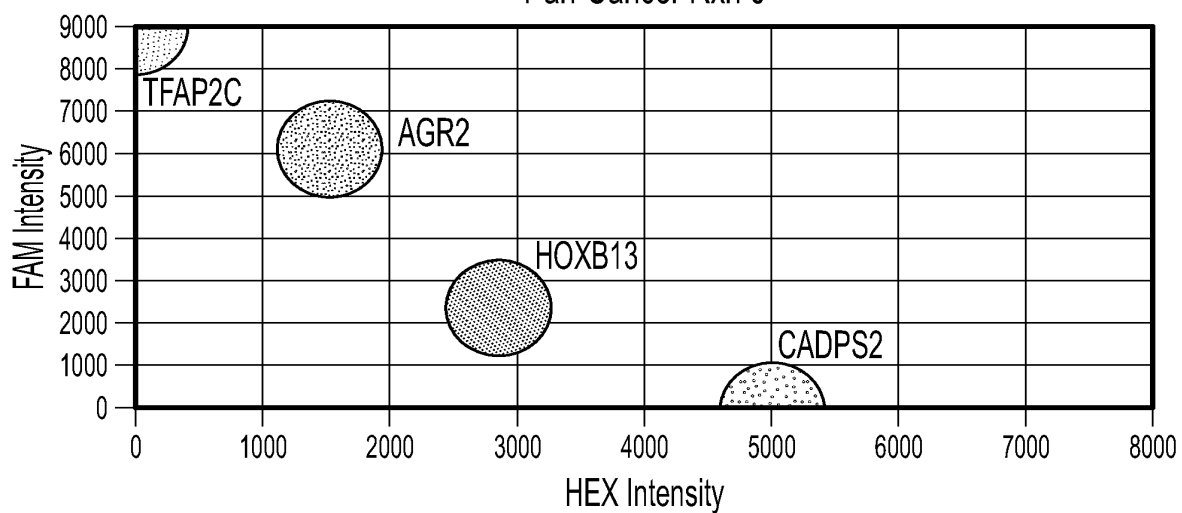


FIG.13A

Breast/Lung Cancer Cell Line Experiment (Combo)

Rxn 6

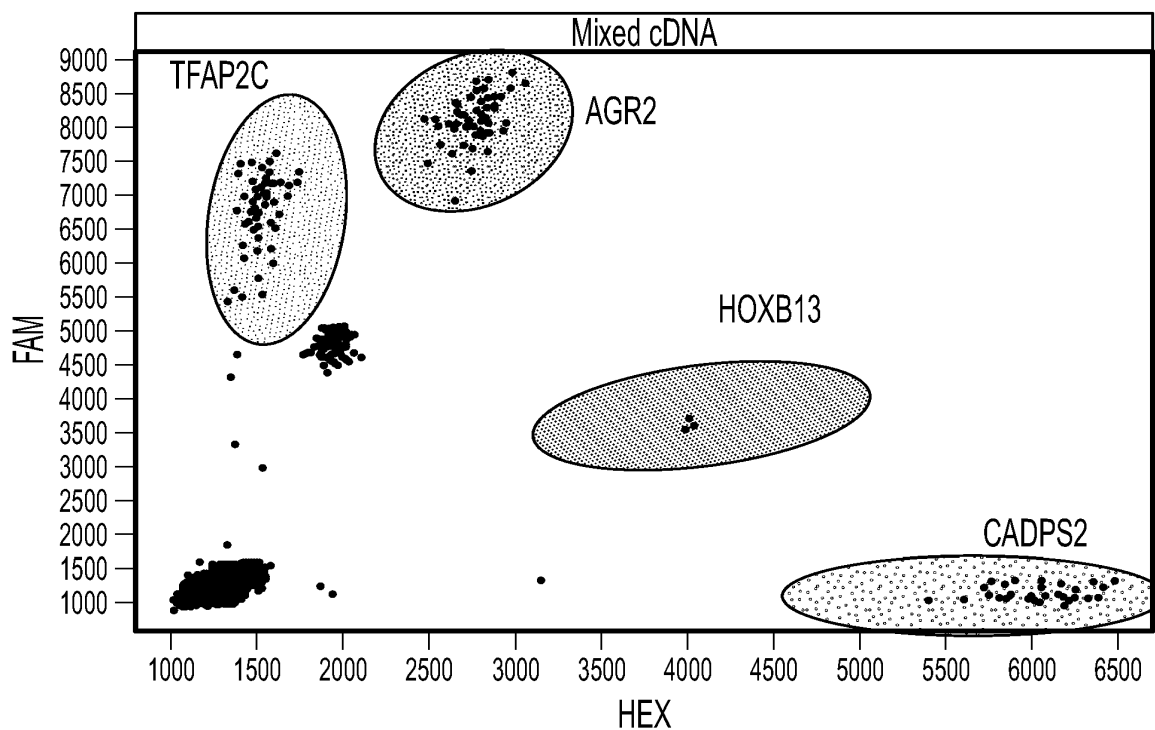


FIG. 13B

SUBSTITUTE SHEET (RULE 26)

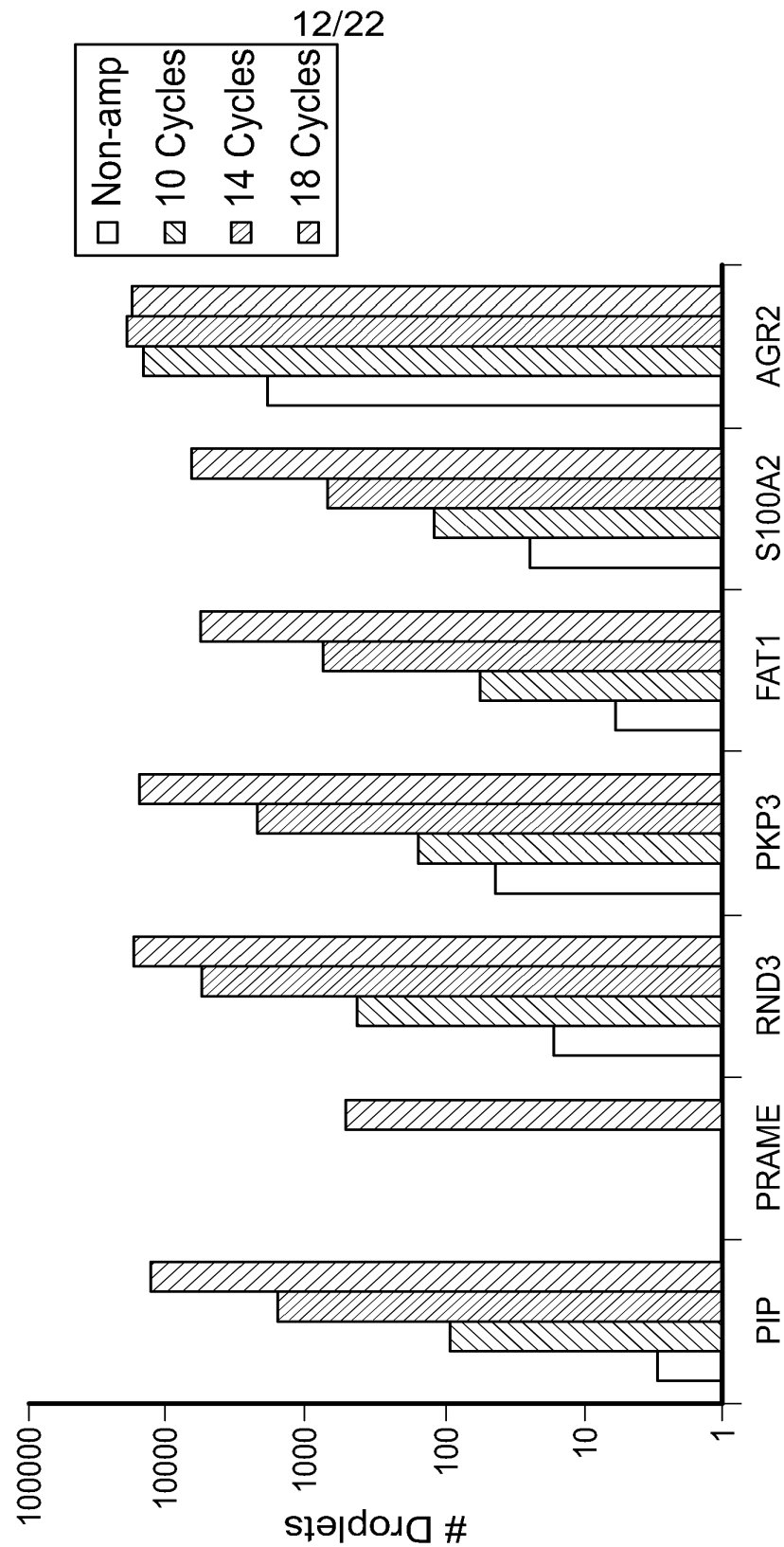


FIG. 14

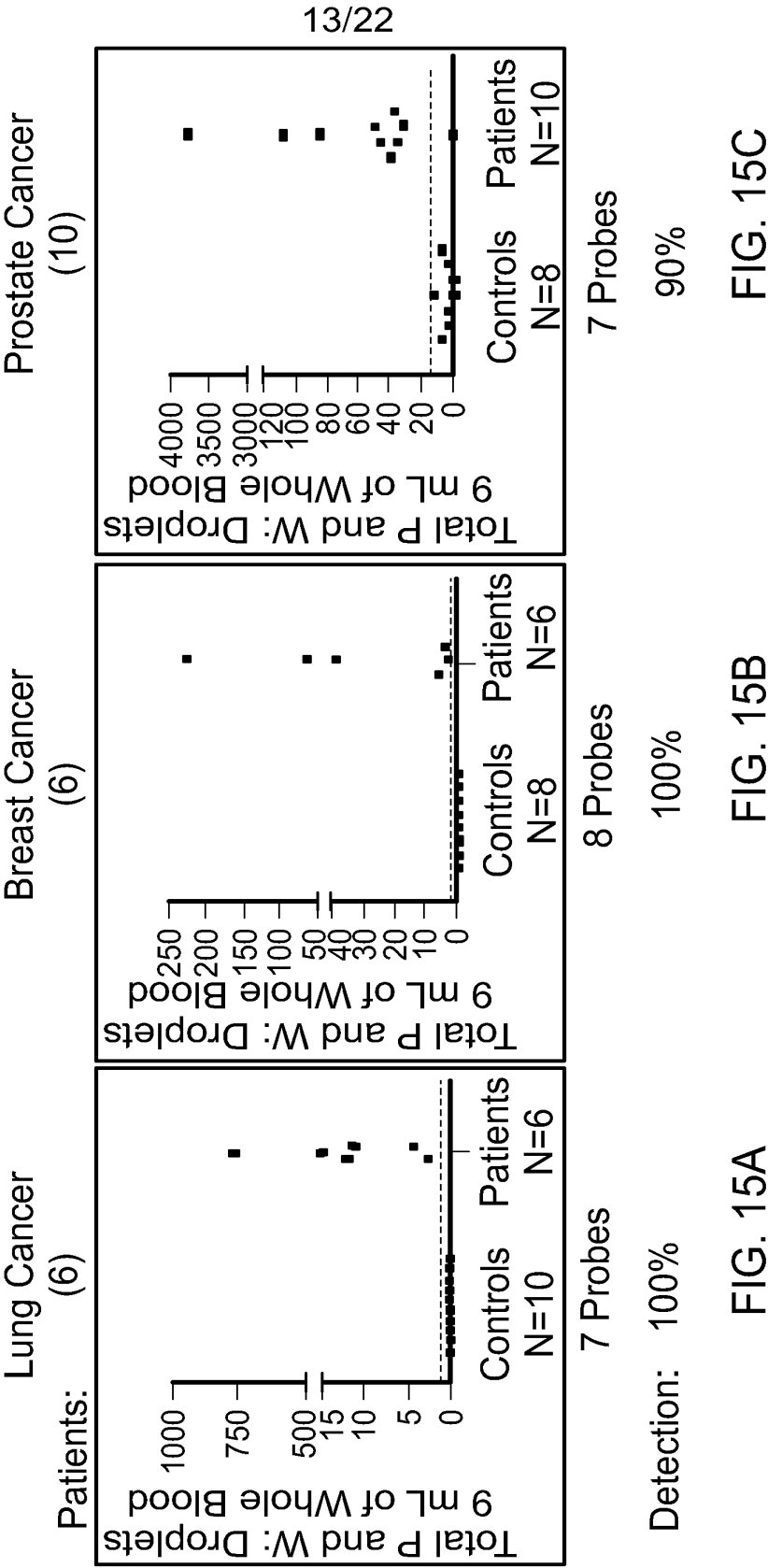


FIG. 15C

FIG. 15B

FIG. 15A

14/22

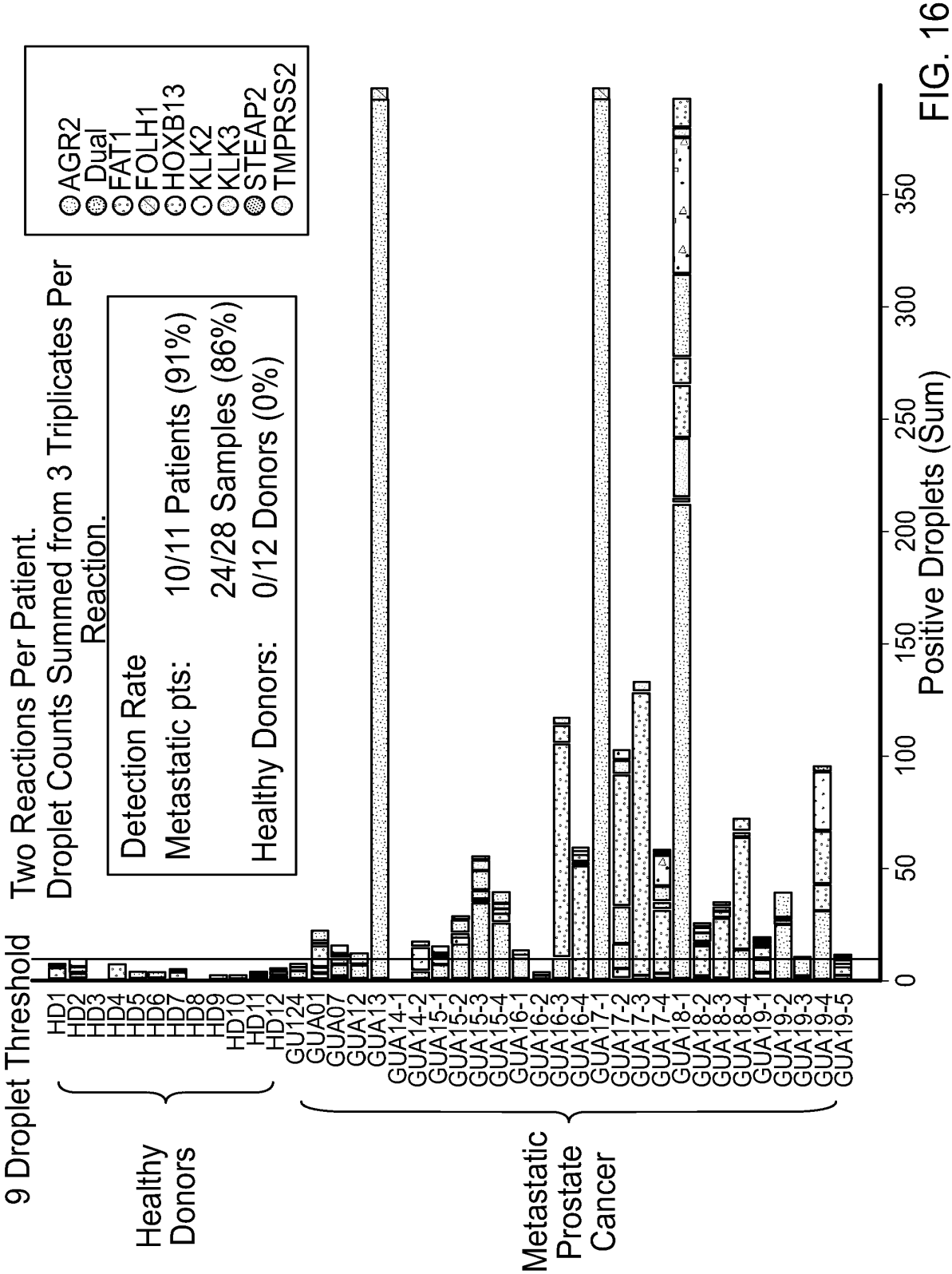


FIG. 16

15/22

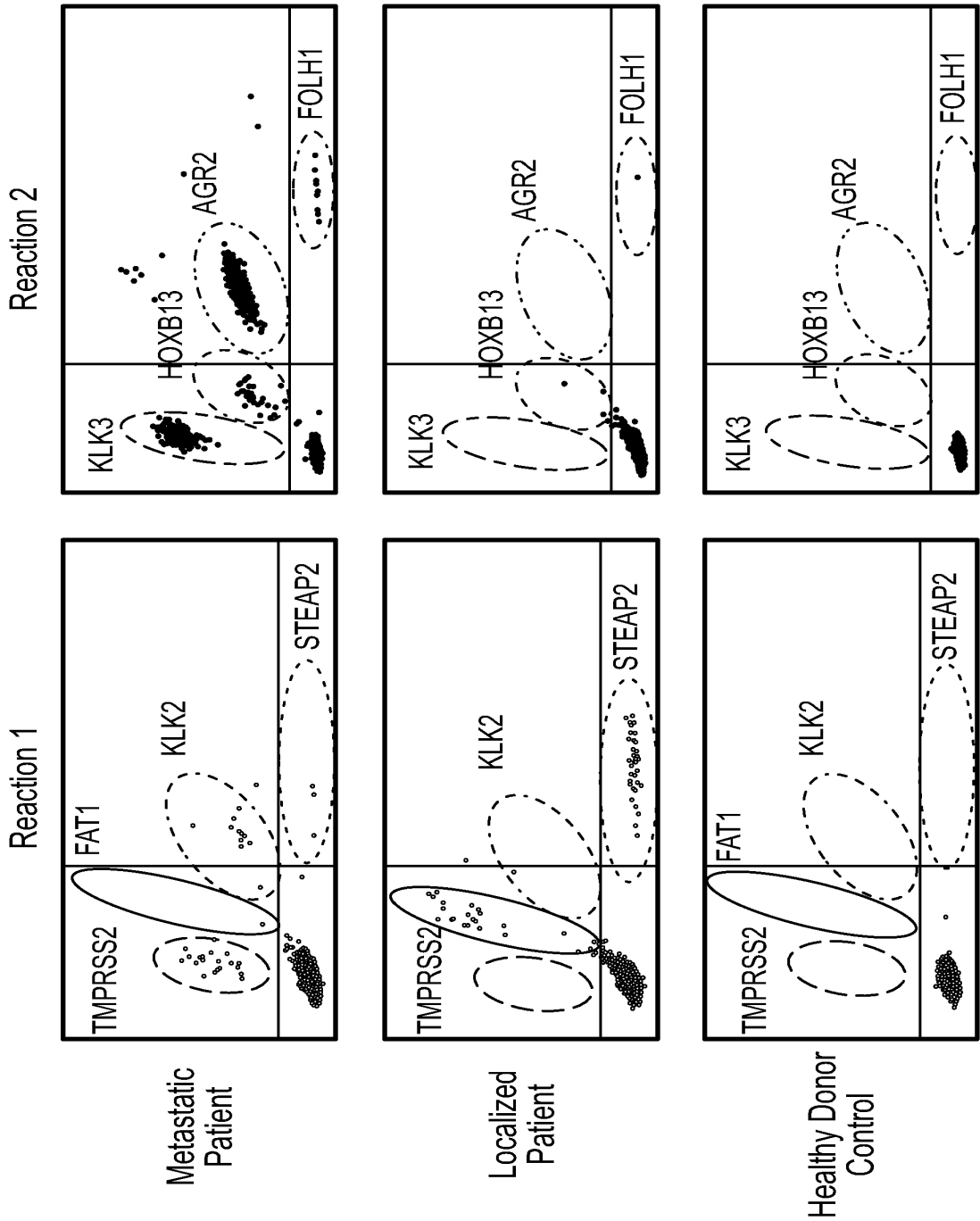


FIG. 17

16/22

Using Markers of Androgen Receptor Signaling to Readout Drug Response

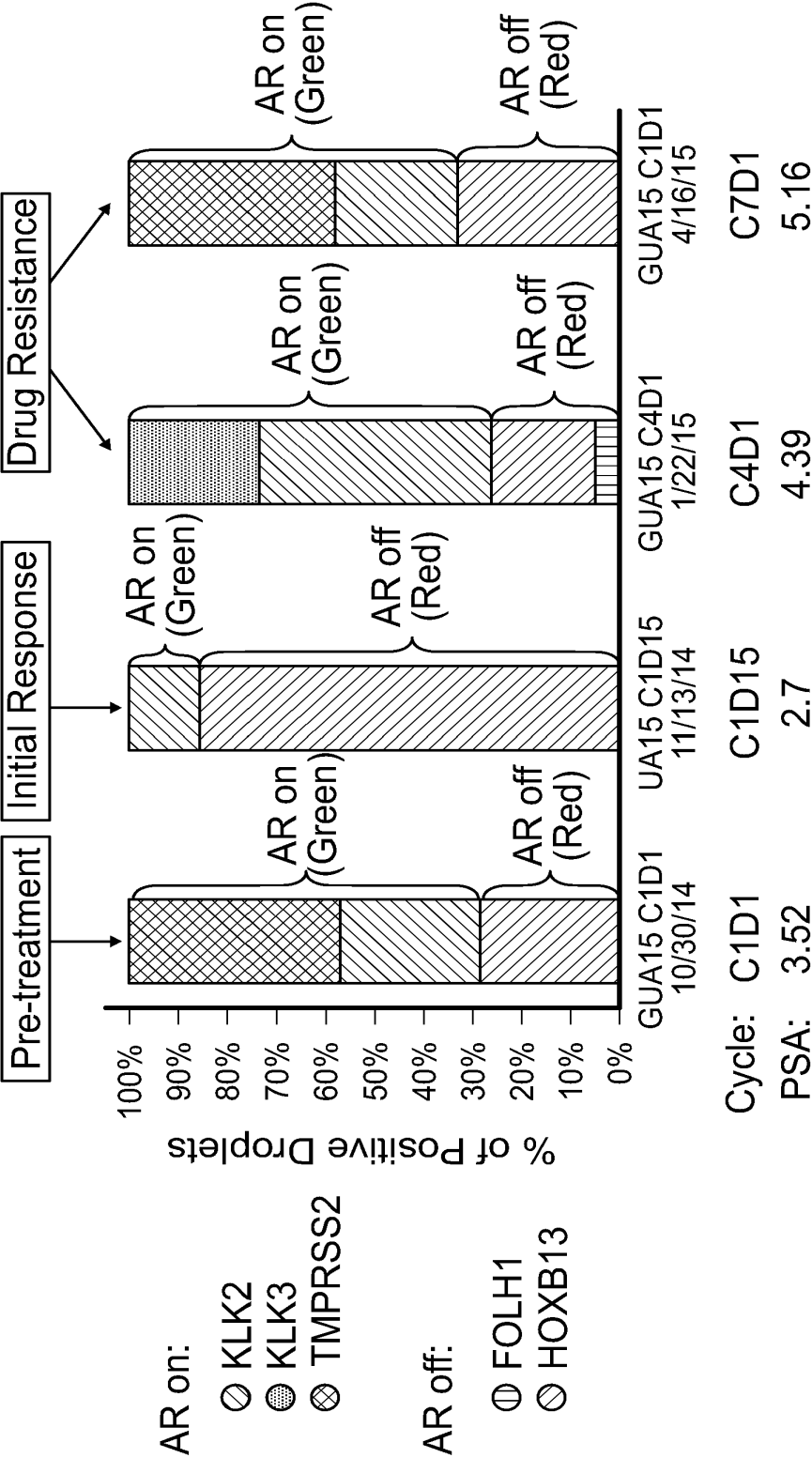


FIG. 18

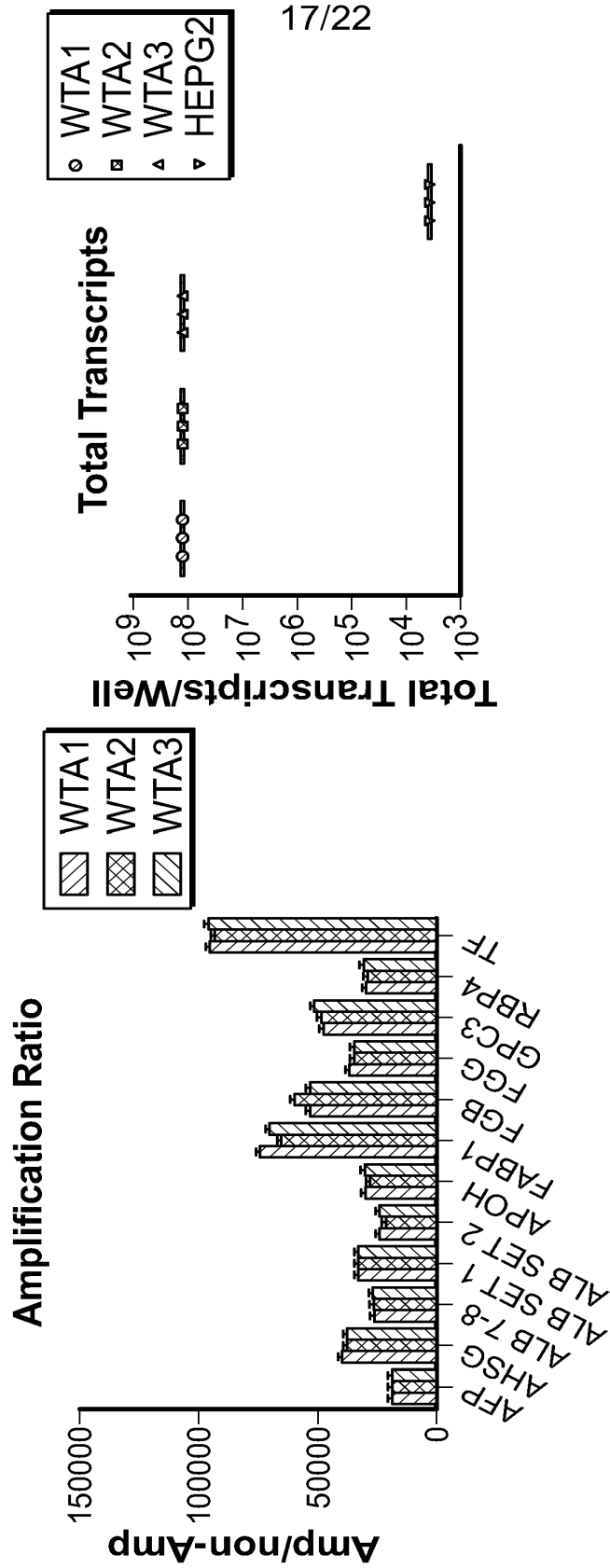


FIG. 19B

FIG. 19A

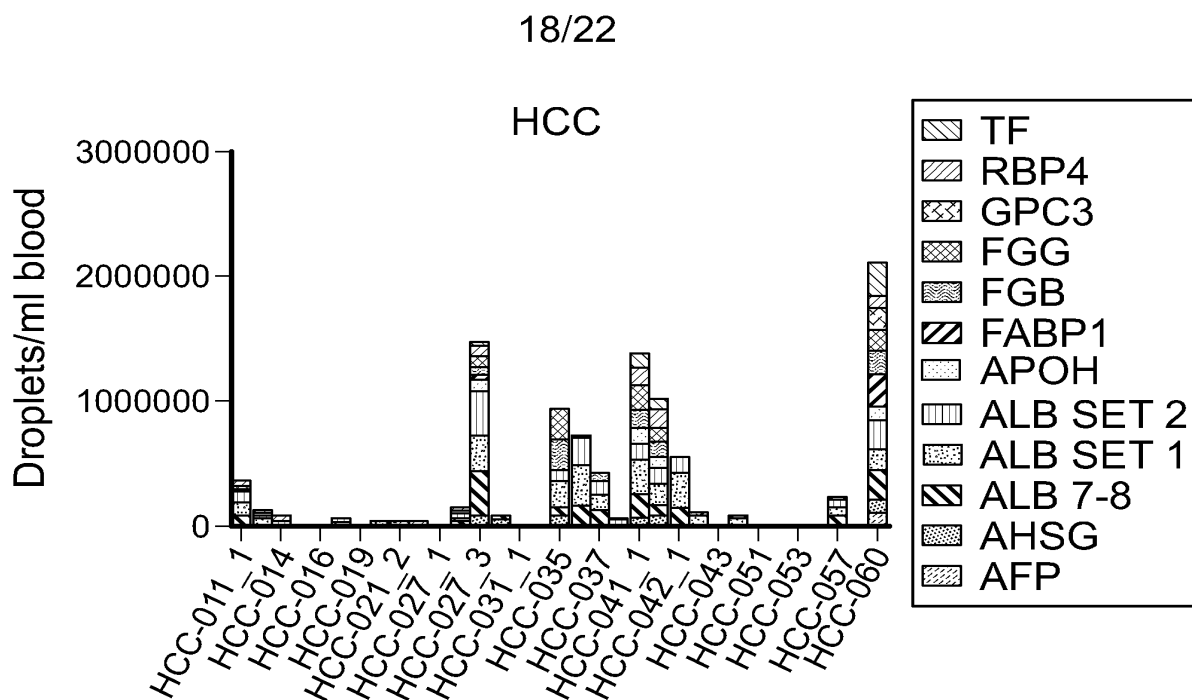


FIG. 20A

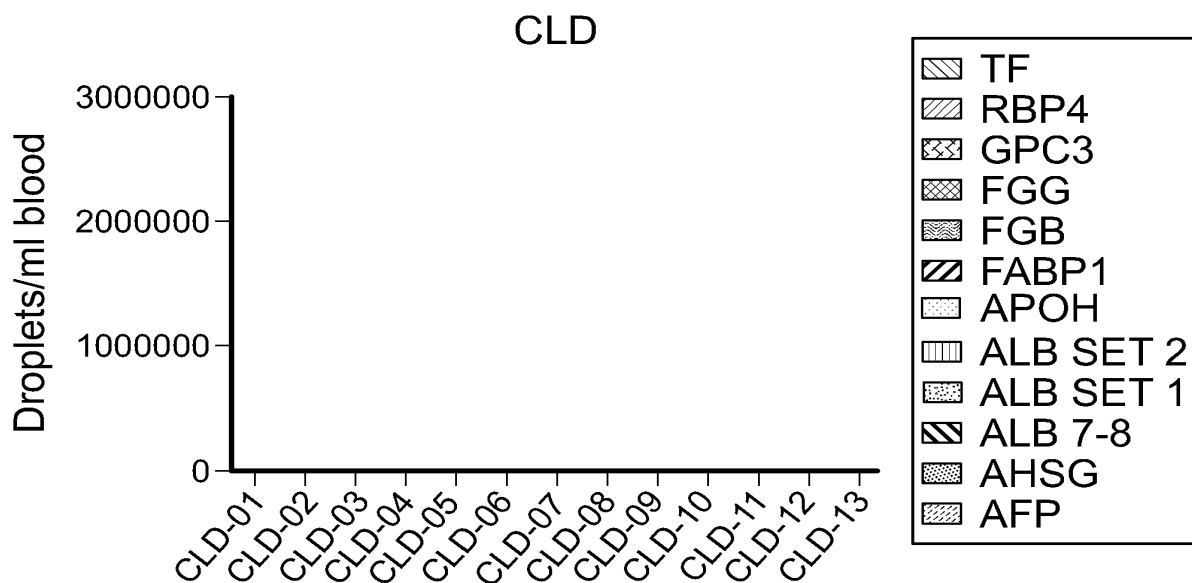


FIG. 20B

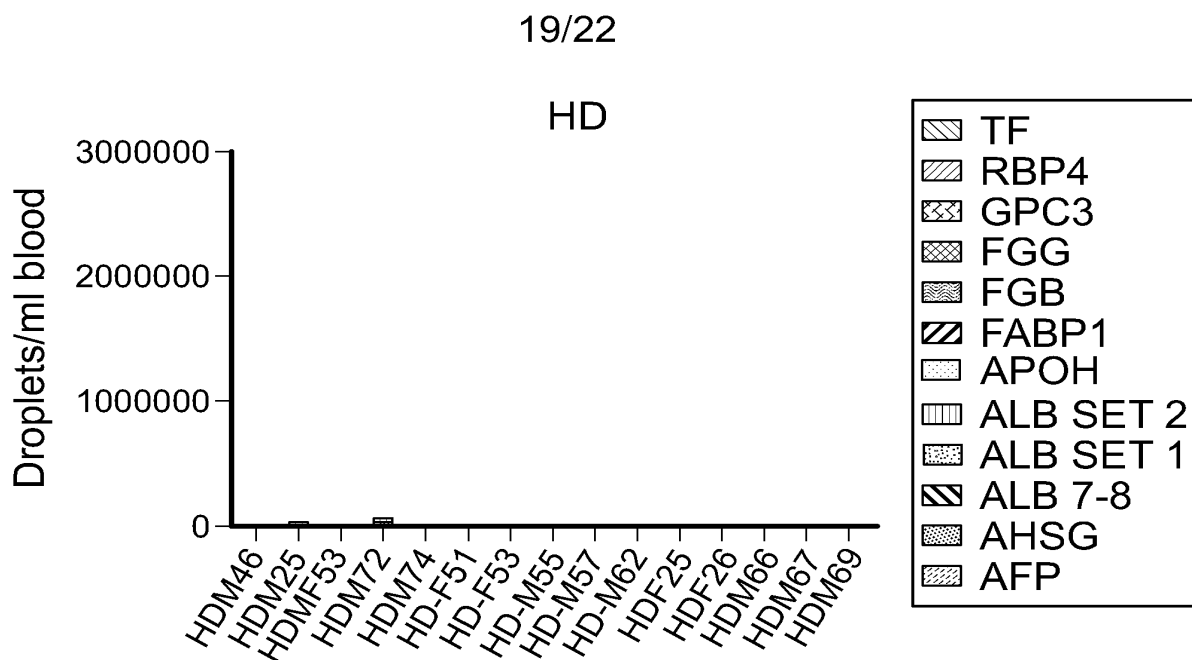


FIG. 20C

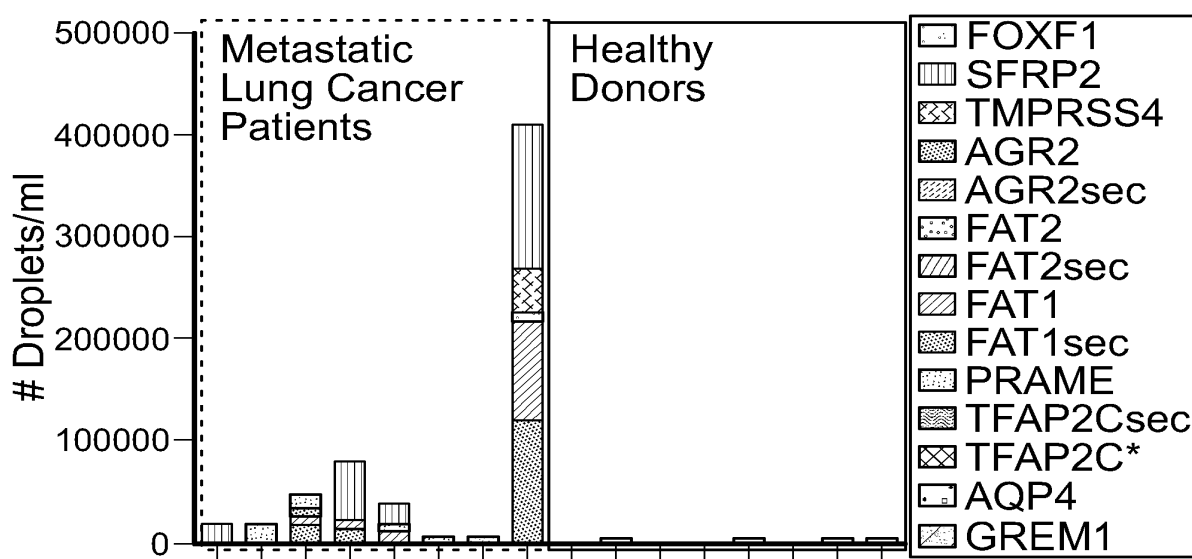


FIG. 21A

20/22

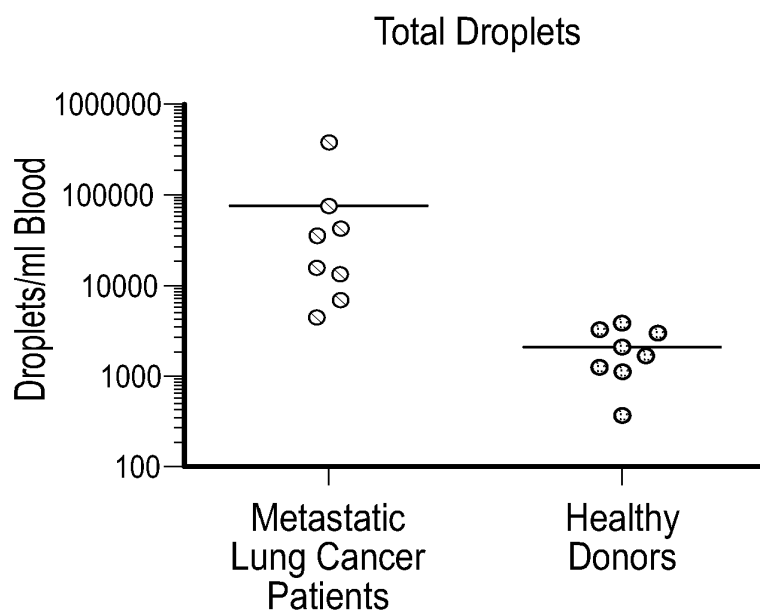


FIG. 21B

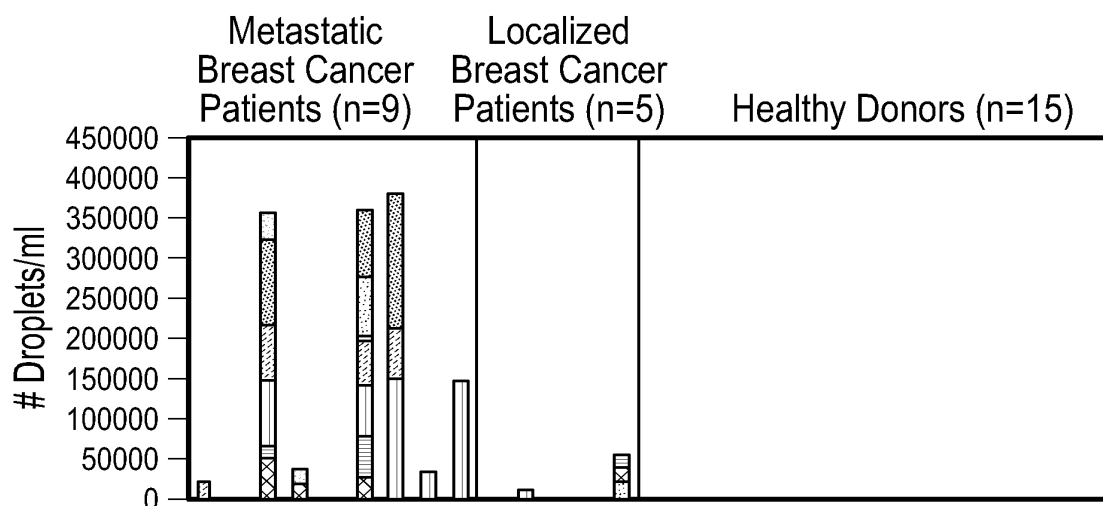
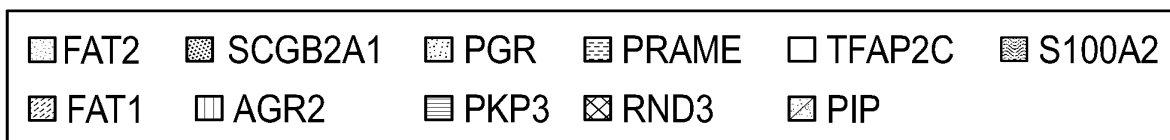


FIG. 22

SUBSTITUTE SHEET (RULE 26)

21/22

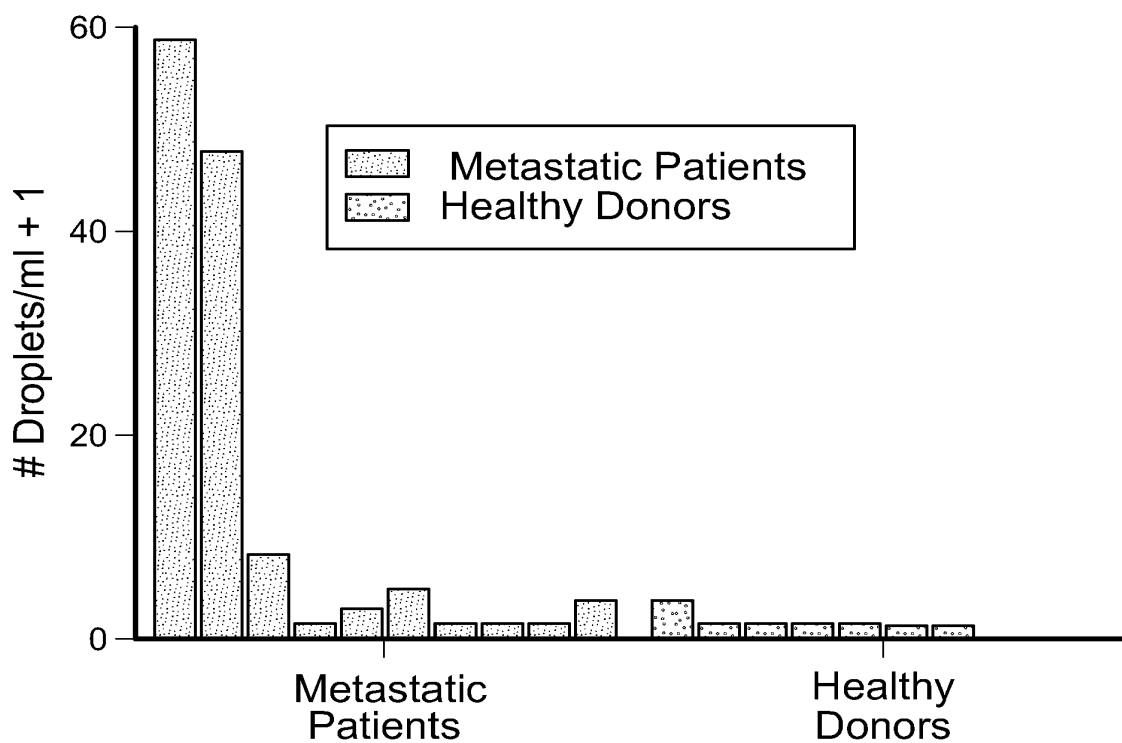


FIG. 23A

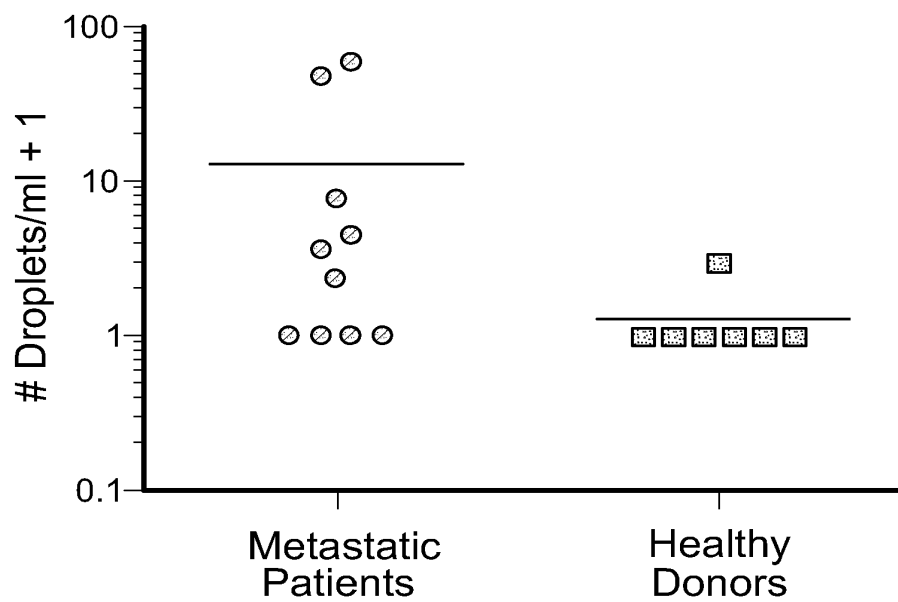


FIG. 23B

22/22

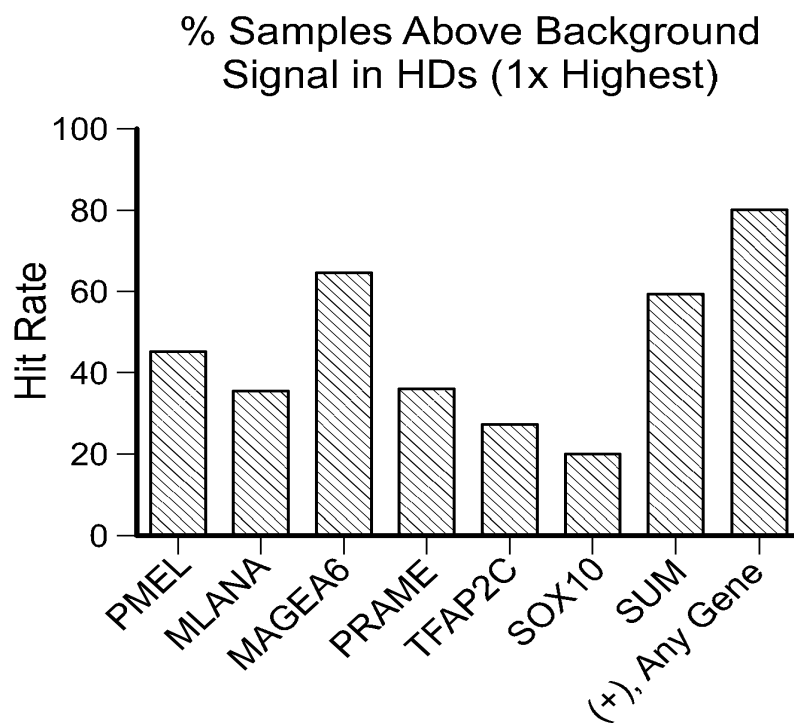


FIG. 24A

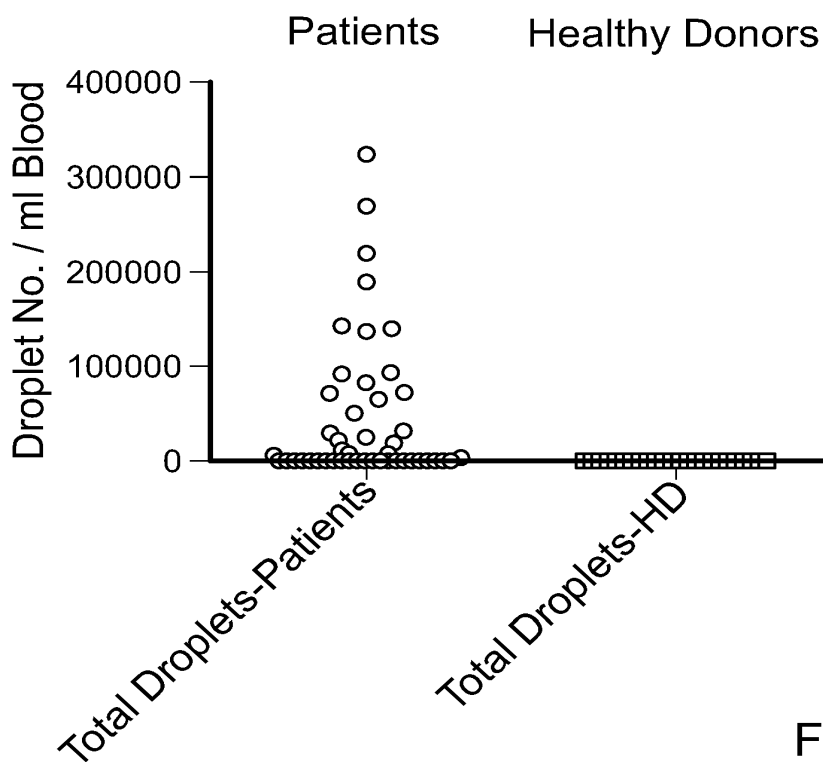


FIG. 24B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 16/24367

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68 (2016.01)

CPC - C12Q 1/6827, C12Q 1/6886

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C12Q 1/68 (2016.01)

CPC - C12Q 1/6827, C12Q 1/6886

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/6.12, 435/6.14

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

pubWEST; Google Scholar; PatBase

search terms - circulating tumor cells, encapsulate, encapsulated, droplets, cDNA, RNA, AGR2, FOLH1, HOXDB13, KLK2, KLK3 or SCHLAP1 or SET4 or AMACR or AR or UGT2B15 or SET1 or SET5 or STEAP2 or LSAMP or fat2 or fat1 or egfr or aldh1a\$ or cdh21

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2014/028378 A2 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 20 February 2015 (20.02.2015) para [0004]; [0009]; [0010]; [0023]; [0026]; [0036]; [0094]; [0110]; [0259].	1-4
Y	US 2012/0252015 A1 (HINDSON et al.) 04 October 2012 (04.10.2012) para [0244]; [0391]-[0395].	1-4
A	US 2012/0015835 A1 (FUCHS et al.) 19 January 2012 (19.01.2012) para [0010]; [0292]-[0296].	1

☐ Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

01 August 2016

Date of mailing of the international search report

22 AUG 2016

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

IPR2024-01451

California Institute of Technology - Ex. 2003 Page 3364

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 16/24367

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 5-23
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-4, directed to a method for detection of cancer with ultra-high sensitivity and specificity in a subject, or a method of analyzing circulating tumor cells (CTCs) in a blood sample.

Group II, claims 24-28, directed to use of probes and primers related to one or more selected cancer genes listed in Table 1.

The inventions listed as Groups I and II do not relate to a single special technical feature under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

--continued on first extra sheet attached hereto--

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-4

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

IPR2024-01451

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 16/24367

--- continued from Box No III Lack of Unity ---

Special technical features

Group I has the special technical feature of a method for detection of cancer in a subject, or a method of analyzing circulating tumor cells (CTCs) in a blood sample, comprising isolating CTCs, converting CTC-derived RNA into cDNA, and encapsulating the cDNA into individual droplets, that is not required by Group II.

Group II has the special technical feature of use of probes and primers related to one or more selected cancer genes listed in Table 1, that is not required by Group I.

Common technical features:

Groups I-II share the common technical feature of amplifying and detecting cDNA molecules obtained from circulating tumor cells (CTCs) in a blood sample. Groups I and II further share the common technical feature of cancer genes. However, this shared technical feature does not represent a contribution over prior art, because this shared technical feature is anticipated by US 2012/0015835 A1 to Fuchs et al., (hereinafter Fuchs).

Fuchs teaches amplifying and detecting cDNA molecules of cancer genes obtained from circulating tumor cells (CTCs) in a blood sample (para [0010] 'a gene encoding a polypeptide listed in Table 1...Gene expression analysis may be used to determine the tissue or tissues of origin of the cancer cells and may be performed on a single cancer cell'; [0292]-[0296] 'Purify CTCs from blood sample...Purify total RNA from CTCs...Convert RNA to cDNA using reverse transcriptase...Use resultant cDNA to perform first and second PCR reactions...Purify the nested PCR amplicon and use as a sequencing template to sequence EGFR exons 18-21').

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

Therefore, Groups I-II lack unity of invention under PCT Rule 13.

NOTE, continuation of number 4 above: claims 5-23 are held unsearchable because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).



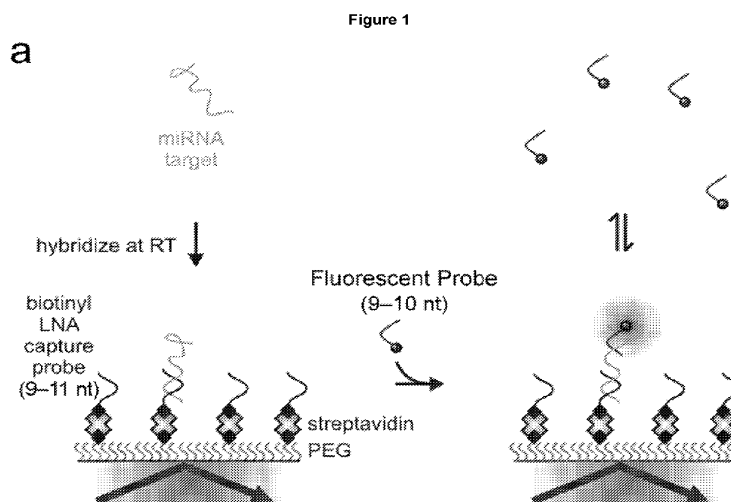
- (51) **International Patent Classification:**
C12N 9/22 (2006.01) *C40B 30/04* (2006.01)
C12P 19/34 (2006.01) *C40B 40/06* (2006.01)
- (21) **International Application Number:**
 PCT/US2017/016977
- (22) **International Filing Date:**
 8 February 2017 (08.02.2017)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
 62/293,589 10 February 2016 (10.02.2016) US
- (71) **Applicant:** THE REGENTS OF THE UNIVERSITY OF MICHIGAN [US/US]; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US).
- (72) **Inventors:** WALTER, Nils; c/o - The Regents Of The University Of Michigan, 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). TEWARI, Mun-eesh; c/o - The Regents Of The University Of Michigan, 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). JOHNSON-BUCK, Alexander; c/o - The Regents Of The University Of Michigan, 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US).
- (74) **Agent:** ISENBARGER, Thomas A.; CASIMIR JONES SC, 2275 DEMING WAY, STE 310, Middleton, Wisconsin 53562 (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, DJ, 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, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, ST, 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:

— with international search report (Art. 21(3))

[Continued on next page]

(54) Title: DETECTION OF NUCLEIC ACIDS



(57) **Abstract:** Provided herein is technology relating to detecting and identifying nucleic acids and particularly, but not exclusively, to compositions, methods, kits, and systems for detecting, identifying, and quantifying target nucleic acids with high confidence at single-molecule resolution.



— *with sequence listing part of description (Rule 5.2(a))*

DETECTION OF NUCLEIC ACIDS

This application claims priority to United States provisional patent application serial number 62/293,589, filed February 10, 2016, which is incorporated herein by reference in its entirety.

5

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with government support under grant GM062357 awarded by the U.S. National Institutes of Health, and under grant W911NF-12-1-0420 awarded by the U.S. Navy, Office of Naval Research. The government has certain rights in the invention.

10

FIELD

Provided herein is technology relating to detecting and identifying nucleic acids and particularly, but not exclusively, to compositions, methods, kits, and systems for detecting, identifying, and quantifying target nucleic acids with high confidence at single-molecule resolution.

15

BACKGROUND

Early detection is critical to the effective treatment of many diseases, especially cancer. Research related to identifying detectable biomarkers associated with early-stage disease has indicated that nucleic acids provide highly specific biomarkers of cancer and other maladies. For example, cancer cell-derived double-stranded DNA (dsDNA) together with secondary structured long non-coding RNA (lncRNA) have recently emerged as sensitive and specific biomarkers of cancer and other diseases in crude human biofluids such as blood, urine, and sputum.

20
25

However, despite their promise as diagnostic biomarkers, the sensitive and specific detection of nucleic acid biomarkers has proven challenging. In particular, existing techniques for detecting nucleic acids utilize probes that form a thermodynamically stable complex with the target molecule and are thus limited to weak and often unreliable thermodynamic discrimination against background signal, spurious targets, or closely related mutant nucleic acids. In addition, the presence of a complementary DNA or RNA strand in the sample severely limits accessibility of the target sequence.

30

Thus, a sensitive and specific assay for the amplification-free detection of nucleic acids in minimally treated native biofluids is needed to provide a rapid and reliable identification and/or quantification of nucleic acid biomarkers.

5

SUMMARY

Accordingly, provided herein is a technology for the specific and ultrasensitive detection and counting of single nucleic acid (e.g., dsDNA, lncRNA, methylated DNA, etc.) target molecules based on the transient binding of short labeled probes to a target nucleic acid. In some embodiments, the target nucleic acid is detected by a kinetic “fingerprint” signal produced by the probe-target interaction. This Single-Molecule Recognition with Equilibrium Poisson Sampling (SiMREPS) technology provides for the sensitive detection of both single-stranded nucleic acids (see, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety) and double-stranded nucleic acids. In some embodiments, detection of double-stranded nucleic acids comprises use of dCas9-guided capture and DNA melting. In some embodiments, this technology comprises the capture of unlabeled targets on a glass or fused silica surface using a guideRNA (gRNA)-loaded, catalytically inactive (“dead”) dCas9 enzyme or enzymes that bind one or more segments of the target nucleic acid with high specificity. In some embodiments (e.g., for detecting lncRNA), the technology further comprises use of a protospacer adjacent motif (PAM) oligonucleotide to provide for dCas9 targeting of nucleic acid targets (e.g., lncRNA).

In some embodiments, the technology comprises use of an intramolecular SiMREPS probe, e.g., a capture probe and a query probe that are linked to provide an intramolecular probing mechanism (see, e.g., Fig. 5 a and b; see *infra*).

More broadly, in some embodiments, one or more complexes (e.g., one or more dCas9/gRNA complexes) recognize specific segments (e.g., nucleotide sequences) in a target nucleic acid to immobilize the target nucleic acid to a surface. Furthermore, embodiments provide that these same or other complexes and/or complementary oligonucleotides melt double-stranded regions of the target nucleic acid to provide access for the binding of a labeled query probe (e.g., for binding of the query probe to a second segment, e.g., a query region). Surface capture (and, in some embodiments, melting of one or more double-stranded regions) is followed by observation of the repeated, transient binding of a short fluorescently labeled DNA query probe to the second segment (e.g., query region) of the target nucleic acid that has been made accessible by dCas9-mediated melting.

Furthermore, in addition to using dCas9/gRNA for immobilization and/or exposure of target sequences (e.g., one or more query regions), embodiments also provide a technology in which a labeled (e.g., fluorescently labeled) dCas9/gRNA complex provides detection of the target nucleic acid. In particular, the dCas9/gRNA complex provides a query probe for SiMREPS, since the dwell time of dCas9/gRNA on a DNA sequence is sensitive to the number of base pairs formed between the gRNA and the target DNA sequence. Embodiments of the technology provide that the number of base pairs formed between the gRNA and the target DNA is tuned to promote rapid dissociation from mutant sequences but slow dissociation from wild-type sequences or vice-versa. Furthermore, engineered dCas9 proteins provide the appropriate kinetics and sequence specificity (see, e.g., Kleinstiver et al. (2016) "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects" *Nature* 529: 490-495; Slaymaker et al. (2015) "Rationally engineered Cas9 nucleases with improved specificity" *Science* 351: 84-8) describing engineered Cas9 proteins that interact more weakly with the DNA backbone than native Cas9, thus reducing off-target effects.

In particular embodiments described herein, the SiMREPS probes repeatedly bind to a target sequence (e.g., query region) specifically made accessible by binding of dCas9/gRNA to one or more nucleic acid region(s) adjacent to the target region (e.g., "adjacent regions"). The repeated binding of the query probes to the query region provides a unique, continuous kinetic "fingerprint", providing a large number of independent measurements for each observed target molecule. This repeated kinetic sampling affords two main advantages: (1) arbitrarily high discrimination against background signals with increased sampling time, essentially eliminating false positive signals; and (2) exquisite sensitivity to subtle differences in the identity of the molecular target, allowing for discrimination of slightly different biomarkers (e.g., differing by only one DNA/RNA base of a disease-related mutation, a methylation pattern or other chemical marks) with very high confidence. In addition, as a single-molecule technique, this approach detects a small amount of target molecule (such as derived from a single cell) in the presence of a large excess of a closely related, but spurious (e.g., mutant) target. In contrast to techniques requiring PCR amplification (the current standard for low-abundance targets), this technique utilizes no target amplification and hence requires pre-treatment of biological samples only with, e.g., dCas9/gRNA at ambient temperature prior to target detection, avoiding the introduction of sampling bias. In addition, as a direct detection technique, it does not lose any chemical marks on the target as amplification-based detection approaches often do. The technique finds use, for

example, in the diagnosis of cancer from circulating tumor DNA and lncRNA in human blood serum. The technology finds use also in drug and antibiotic resistance gene detection in pathogens and human tumor cells.

Accordingly, provided herein are embodiments of a complex for providing a detectable fingerprint of a double-stranded target nucleic acid, the complex comprising a double-stranded target nucleic acid (e.g., a DNA, an RNA, a DNA/RNA hybrid) comprising a first region adjacent to a second region; a melting component (e.g., an immobilized melting component) interacting with the first region to form a thermodynamically stable complex and provide the second region in a single-stranded form; and a query probe that binds repeatedly to the second region to provide a detectable fingerprint associated with the double-stranded target nucleic acid. In some embodiments, the double-stranded nucleic acid is a single-stranded nucleic acid that comprises a region having a double-stranded secondary structure. In some embodiments, the melting component comprises a dCas9. In some embodiments, the melting component comprises a single-stranded binding protein. In some embodiments, the melting component is a protein and in some embodiments the melting component is a nucleic acid. Embodiments comprise use of a protein that binds to double-stranded nucleic acids (e.g., double-stranded DNA, double-stranded RNA, a double-stranded DNA/RNA hybrid) and/or a melting component (e.g., a protein or a nucleic acid) to dissociate a double-stranded nucleic acid (e.g., a region of a nucleic acid) to provide a single-stranded nucleic acid. In particular embodiments, the melting component comprises a dCas9/gRNA complex comprising a gRNA hybridized to the first region. In some embodiments, the melting component comprises a PAMmer, e.g., provided in trans to the target nucleic acid.

In some embodiments, the query probe hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 0.1 min^{-1} . In some embodiments, the query probe hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 1 min^{-1} . In some embodiments, the query probe is a fluorescently labeled nucleic acid that hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 0.1 min^{-1} . In some embodiments, the query probe is a fluorescently labeled nucleic acid that hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 1 min^{-1} .

In some embodiments, the melting component comprises a dCas9 that is immobilized to a substrate.

In some embodiments, data are analyzed, e.g., in some embodiments the fingerprint is detectable by a pattern recognition analysis.

5 Additional embodiments comprise a second melting component interacting with a third region of the target nucleic acid adjacent to the second region of the target nucleic acid. In some embodiments, the second melting component comprises a dCas9. In some
10 embodiments, the second melting component comprises a single-stranded binding protein. In some embodiments, the second melting component is a protein and in some embodiments the second melting component is a nucleic acid. In particular
15 embodiments, the second melting component comprises a dCas9/gRNA complex comprising a gRNA hybridized to the third region. In some embodiments, the melting component comprises a PAMmer, e.g., provided in trans to the target nucleic acid. In
some embodiments, the first and second melting components bind approximately 5 to 15
nucleotides apart on the target nucleic acid, e.g., to provide access to the query region by
a query probe.

The technology finds use in the detection, identification, and/or quantification of nucleic acids, e.g., in some embodiments, the target nucleic acid comprises a mutation, a single nucleotide polymorphism, or a modified base.

20 Additional embodiments provide a method for providing a detectable fingerprint of a double-stranded target nucleic acid in a sample, the method comprising
immobilizing a double-stranded target nucleic acid to a discrete region of a solid support, said double-stranded target nucleic acid comprising a first region adjacent to a second
region and said discrete region of said solid support comprising an immobilized melting
25 component interacting with the first region; providing a query probe that binds repeatedly to the second region to provide a detectable fingerprint; and associating the
detectable fingerprint with the double-stranded nucleic acid to identify the double-
stranded nucleic acid. Some embodiments comprise analyzing data using pattern
recognition or a similar analysis (e.g., machine learning, neural network, supervised
30 and/or unsupervised learning, etc.) to produce or identify the detectable fingerprint of
the double stranded nucleic acid.

In some embodiments of methods, the melting component comprises a dCas9. In
some embodiments of methods, the melting component comprises a single-stranded
binding protein. In some embodiments of methods, the melting component is a protein
35 and in some embodiments the melting component is a nucleic acid. In particular method

embodiments, the melting component comprises a dCas9/gRNA complex comprising a gRNA hybridized to the first region. In some embodiments of methods, the melting component comprises a PAMmer, e.g., provided in trans to the target nucleic acid.

Additional embodiments comprise providing a second melting component that
5 interacts with a third region of the target nucleic acid, said third target region adjacent to the second region. For example, some embodiments comprise providing a dCas9/gRNA complex comprising a gRNA complementary to a third region of the target nucleic acid adjacent to the second region of the target nucleic acid. Related
10 embodiments comprise providing conditions sufficient for the melting component to provide the second region in a second stranded form, e.g., buffered pH conditions, temperature control, solution components (e.g., salts, counterions, cofactors, etc.), etc.

Some embodiments comprise detecting repeated binding of the query probe to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 0.1 min^{-1} . Some embodiments comprise
15 detecting repeated binding of the query probe to the second region with a kinetic rate constant k_{off} that is greater than 1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 1 min^{-1} . Some embodiments comprise detecting repeated binding of a fluorescently labeled nucleic acid to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 0.1 min^{-1} . Some
20 embodiments comprise detecting repeated binding of a fluorescently labeled nucleic acid to the second region with a kinetic rate constant k_{off} that is greater than 1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 1 min^{-1} .

Some embodiments comprise calculating an amount or concentration of the double-stranded target nucleic acid in the sample from the detectable fingerprint.

25 Embodiments of the technology provide a system for the detection of a double-stranded nucleic acid. In some embodiments, the system comprises a solid support comprising an immobilized melting component, a detectably labeled query probe that binds repeated to the double-stranded nucleic acid, a fluorescence detector, and a software component configured to perform pattern recognition analysis of query probe
30 binding data. Embodiment of systems comprise compositions described herein and/or comprise components (e.g., a computer, processor, etc.) to perform methods as described herein.

The technology finds use in embodiments of a method for calculating a predictor that a subject has or is at risk of having a cancer. For example, embodiments of said
35 methods comprise determining the presence of a microRNA biomarker, determining the

presence of a mutation, and determining the presence of a modified base in genomic DNA. In some embodiments, the predictor is a value calculated from variables associated with the presence of a microRNA biomarker, the presence of a mutation, and the presence of a modified base in genomic DNA.

Related embodiments provide a complex for detecting a target nucleic acid, the complex comprising a target nucleic acid comprising a first region adjacent to a second region; a detectably labeled capture probe hybridized to the first region; a query probe labeled with a quencher or fluorescent acceptor compatible with the label of the capture probe, wherein the query probe hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 0.1 min^{-1} .

Additional embodiments will be apparent to persons skilled in the relevant art based on the teachings contained herein.

15 BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present technology will become better understood with regard to the following drawings:

Figure 1 a is a schematic drawing of an embodiment of the nucleic acid detection technology provided herein.

Figure 1 b shows exemplary SiMREPS data of fluorescently labeled query probes transiently associating non-specifically to a slide surface.

Figure 1 c shows exemplary SiMREPS data of fluorescently labeled query probes transiently binding to a target nucleic acid. The kinetic fingerprints of 1c and 1b are different, thus providing examples of the different kinetic signatures for specific and non-specific binding of query probes.

Figure 1 d is a series of histograms indicating the number of query probes counted to have a given number of intensity transitions (N_{b+d}) in the absence (thick gray bars) or presence (thin lines) of 1 pM target nucleic acid (e.g., a miR-141 microRNA). The four histograms plot data acquired with acquisition times of 1, 2, 5, and 10 minutes.

Figure 1 e shows plots of standard curves from SiMREPS assays of five miRNAs, yielding R² values > 0.99. The SiMREPS technology provides high-confidence detection of nucleic acids.

Figure 2 a is a plot showing that the fluorescent query probe for let-7a exhibits long lifetimes of binding to let-7a ($\tau_{\text{on}} = 23.3 \pm 8.3$ s) but much more transient binding to

let-7c ($\tau_{on} = 4.7 \pm 3.0$ s) due to a single mismatch in the let-7c sequence relative to let-7a (underlined "G").

Figure 2 b is a dwell time analysis showing the high-confidence single-copy-level discrimination between let-7a (closed circles) and let-7c (open circles).

5 Figure 2 c is a receiver operating characteristic (ROC) plot constructed by varying the τ_{on} threshold for discriminating between let-7a and let-7c.

Figure 2 d is a N_{b+d} histogram for the detection of let-7 in crude HeLa cell extract in the presence or absence of the miRCURY let-7 inhibitor. The N_{b+d} histogram for endogenous hsa-let-7a showed a well-defined peak (thin line) that vanished in the
10 presence of a let-7 inhibitor designed to bind and sequester let-7 family members (thick grey bars).

Figure 2 e shows the dwell times for molecules detected in crude HeLa extract using the fluorescent and capture probes for let-7a. The filled and open circles represent two clusters of target molecules classified by k-means clustering of τ_{on} values, consistent
15 with the expected τ_{on} distributions for single-nucleotide mutants hsa-let-7a and hsa-let-7c.

Figure 2 f shows the quantification of synthetic miR-141 spiked into human serum.

Figure 3 is a schematic diagram showing an embodiment of the technology
20 comprising a dCas9/gRNA. Genomic target DNA is briefly pre-treated with dCas9/gRNA. During this time, the guide RNA (gRNA) of the dCas9/gRNA complex hybridizes to the target nucleic acid (e.g., a genomic DNA) at a region adjacent to a query region (e.g., complementary to a query probe). The dCas9/gRNA melts a specific DNA sequence (e.g., the query region) in the target nucleic acid. After capture of the
25 biotinylated dCas9 onto a slide surface (e.g., by biotin-avidin interaction), SiMREPS is used to detect binding of the query probe to the now accessible query region in the target nucleic acid, which is adjacent to the site of target nucleic acid hybridized to the gRNA.

Figure 4 is a schematic diagram showing an embodiment of the technology comprising hybridization of two flanking dCas9/gRNA complexes to a target nucleic
30 acid. In some embodiments, hybridizing two dCas9/gRNA complexes to flank the query region further improves accessibility of the target nucleic acid (e.g., the query region) to the query probe for SiMREPS-based detection, and, in some embodiments, increases specificity. In the embodiment shown in the figure, one dCas9 is biotinylated for capture onto the surface and the other dCas9 is not biotinylated, though in some embodiments
35 the second dCas9 is modified, e.g., biotinylated.

Figure 5 a is a schematic showing an embodiment of intramolecular SiMREPS probing in which the query and capture probes are linked on a contiguous oligonucleotide.

5 Figure 5 b is a schematic showing an embodiment of intramolecular SiMREPS probing in which the non-contiguous query and capture probes are co-localized by an address oligonucleotide.

It is to be understood that the figures are not necessarily drawn to scale, nor are the objects in the figures necessarily drawn to scale in relationship to one another. The figures are depictions that are intended to bring clarity and understanding to various
10 embodiments of apparatuses, systems, and methods disclosed herein. Wherever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts. Moreover, it should be appreciated that the drawings are not intended to limit the scope of the present teachings in any way.

15 DETAILED DESCRIPTION

Provided herein is technology relating to detecting and identifying nucleic acids and particularly, but not exclusively, to compositions, methods, kits, and systems for detecting, identifying, and quantifying target nucleic acids with high confidence at single-molecule resolution

20 In this detailed description of the various embodiments, for purposes of explanation, numerous specific details are set forth to provide a thorough understanding of the embodiments disclosed. One skilled in the art will appreciate, however, that these various embodiments may be practiced with or without these specific details. In other instances, structures and devices are shown in block diagram form. Furthermore, one
25 skilled in the art can readily appreciate that the specific sequences in which methods are presented and performed are illustrative and it is contemplated that the sequences can be varied and still remain within the spirit and scope of the various embodiments disclosed herein.

All literature and similar materials cited in this application, including but not
30 limited to, patents, patent applications, articles, books, treatises, and internet web pages are expressly incorporated by reference in their entirety for any purpose. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which the various
35 embodiments described herein belongs. When definitions of terms in incorporated references appear to differ from the definitions provided in the present teachings, the

definition provided in the present teachings shall control. The section headings used herein are for organizational purposes only and are not to be construed as limiting the described subject matter in any way.

5 Definitions

To facilitate an understanding of the present technology, a number of terms and phrases are defined below. Additional definitions are set forth throughout the detailed description.

Throughout the specification and claims, the following terms take the meanings explicitly associated herein, unless the context clearly dictates otherwise. The phrase “in one embodiment” as used herein does not necessarily refer to the same embodiment, though it may. Furthermore, the phrase “in another embodiment” as used herein does not necessarily refer to a different embodiment, although it may. Thus, as described below, various embodiments of the invention may be readily combined, without departing from the scope or spirit of the invention.

In addition, as used herein, the term “or” is an inclusive “or” operator and is equivalent to the term “and/or” unless the context clearly dictates otherwise. The term “based on” is not exclusive and allows for being based on additional factors not described, unless the context clearly dictates otherwise. In addition, throughout the specification, the meaning of “a”, “an”, and “the” include plural references. The meaning of “in” includes “in” and “on.”

As used herein, a “nucleic acid” or a “nucleic acid sequence” refers to a polymer or oligomer of pyrimidine and/or purine bases, preferably cytosine, thymine, and uracil, and adenine and guanine, respectively (See Albert L. Lehninger, Principles of Biochemistry, at 793-800 (Worth Pub. 1982)). The present technology contemplates any deoxyribonucleotide, ribonucleotide, or peptide nucleic acid component, and any chemical variants thereof, such as methylated, hydroxymethylated, or glycosylated forms of these bases, and the like. The polymers or oligomers may be heterogenous or homogenous in composition, and may be isolated from naturally occurring sources or may be artificially or synthetically produced. In addition, the nucleic acids may be DNA or RNA, or a mixture thereof, and may exist permanently or transitionally in single-stranded or double-stranded form, including homoduplex, heteroduplex, and hybrid states. In some embodiments, a nucleic acid or nucleic acid sequence comprises other kinds of nucleic acid structures such as, for instance, a DNA/RNA helix, peptide nucleic acid (PNA), morpholino, locked nucleic acid (LNA), and/or a ribozyme. Hence, the term

“nucleic acid” or “nucleic acid sequence” may also encompass a chain comprising non-natural nucleotides, modified nucleotides, and/or non-nucleotide building blocks that can exhibit the same function as natural nucleotides (e.g., “nucleotide analogs”); further, the term “nucleic acid sequence” as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin, which may be single or double-stranded, and represent the sense or antisense strand.

The term “nucleotide analog” as used herein refers to modified or non-naturally occurring nucleotides including but not limited to analogs that have altered stacking interactions such as 7-deaza purines (i.e., 7-deaza-dATP and 7-deaza-dGTP); base analogs with alternative hydrogen bonding configurations (e.g., such as Iso-C and Iso-G and other non-standard base pairs described in U.S. Pat. No. 6,001,983 to S. Benner and herein incorporated by reference); non-hydrogen bonding analogs (e.g., non-polar, aromatic nucleoside analogs such as 2,4-difluorotoluene, described by B. A. Schweitzer and E. T. Kool, J. Org. Chem., 1994, 59, 7238-7242, B. A. Schweitzer and E. T. Kool, J. Am. Chem. Soc., 1995, 117, 1863-1872; each of which is herein incorporated by reference); “universal” bases such as 5-nitroindole and 3-nitropyrrole; and universal purines and pyrimidines (such as “K” and “P” nucleotides, respectively; P. Kong, et al., Nucleic Acids Res., 1989, 17, 10373-10383, P. Kong et al., Nucleic Acids Res., 1992, 20, 5149-5152). Nucleotide analogs include nucleotides having modification on the sugar moiety, such as dideoxy nucleotides and 2'-O-methyl nucleotides. Nucleotide analogs include modified forms of deoxyribonucleotides as well as ribonucleotides.

“Peptide nucleic acid” means a DNA mimic that incorporates a peptide-like polyamide backbone.

As used herein, the term “% sequence identity” refers to the percentage of nucleotides or nucleotide analogs in a nucleic acid sequence that is identical with the corresponding nucleotides in a reference sequence after aligning the two sequences and introducing gaps, if necessary, to achieve the maximum percent identity. Hence, in case a nucleic acid according to the technology is longer than a reference sequence, additional nucleotides in the nucleic acid, that do not align with the reference sequence, are not taken into account for determining sequence identity. Methods and computer programs for alignment are well known in the art, including blastn, Align 2, and FASTA.

The term “homology” and “homologous” refers to a degree of identity. There may be partial homology or complete homology. A partially homologous sequence is one that is less than 100% identical to another sequence.

The term “sequence variation” as used herein refers to differences in nucleic acid sequence between two nucleic acids. For example, a wild-type structural gene and a mutant form of this wild-type structural gene may vary in sequence by the presence of single base substitutions and/or deletions or insertions of one or more nucleotides. These two forms of the structural gene are said to vary in sequence from one another. A second mutant form of the structural gene may exist. This second mutant form is said to vary in sequence from both the wild-type gene and the first mutant form of the gene.

As used herein, the terms “complementary” or “complementarity” are used in reference to polynucleotides (e.g., a sequence of nucleotides such as an oligonucleotide or a target nucleic acid) related by the base-pairing rules. For example, for the sequence “5'-A-G-T-3'” is complementary to the sequence “3'-T-C-A-5'.” Complementarity may be “partial,” in which only some of the nucleic acids’ bases are matched according to the base pairing rules. Or, there may be “complete” or “total” complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids. Either term may also be used in reference to individual nucleotides, especially within the context of polynucleotides. For example, a particular nucleotide within an oligonucleotide may be noted for its complementarity, or lack thereof, to a nucleotide within another nucleic acid strand, in contrast or comparison to the complementarity between the rest of the oligonucleotide and the nucleic acid strand.

In some contexts, the term “complementarity” and related terms (e.g., “complementary”, “complement”) refers to the nucleotides of a nucleic acid sequence that can bind to another nucleic acid sequence through hydrogen bonds, e.g., nucleotides that are capable of base pairing, e.g., by Watson-Crick base pairing or other base pairing. Nucleotides that can form base pairs, e.g., that are complementary to one another, are the pairs: cytosine and guanine, thymine and adenine, adenine and uracil, and guanine and uracil. The percentage complementarity need not be calculated over the entire length of a nucleic acid sequence. The percentage of complementarity may be limited to a specific region of which the nucleic acid sequences that are base-paired, e.g., starting from a first base-paired nucleotide and ending at a last base-paired nucleotide. The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in “antiparallel association.” Certain bases not

commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine and 7-deazaguanine.

Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can

5 determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs.

Thus, in some embodiments, "complementary" refers to a first nucleobase sequence that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99%
10 identical to the complement of a second nucleobase sequence over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, or more nucleobases, or that the two sequences hybridize under stringent hybridization conditions. "Fully complementary" means each nucleobase of a first nucleic acid is capable of pairing with each nucleobase at a corresponding position
15 in a second nucleic acid. For example, in certain embodiments, an oligonucleotide wherein each nucleobase has complementarity to a nucleic acid has a nucleobase sequence that is identical to the complement of the nucleic acid over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, or more nucleobases.

20 "Mismatch" means a nucleobase of a first nucleic acid that is not capable of pairing with a nucleobase at a corresponding position of a second nucleic acid.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is influenced by such factors as the
25 degree of complementarity between the nucleic acids, stringency of the conditions involved, and the T_m of the formed hybrid. "Hybridization" methods involve the annealing of one nucleic acid to another, complementary nucleic acid, i.e., a nucleic acid having a complementary nucleotide sequence. The ability of two polymers of nucleic acid containing complementary sequences to find each other and anneal through base
30 pairing interaction is a well-recognized phenomenon. The initial observations of the "hybridization" process by Marmur and Lane, Proc. Natl. Acad. Sci. USA 46:453 (1960) and Doty et al., Proc. Natl. Acad. Sci. USA 46:461 (1960) have been followed by the refinement of this process into an essential tool of modern biology.

As used herein, the term " T_m " is used in reference to the "melting temperature."
35 The melting temperature is the temperature at which a population of double-stranded

nucleic acid molecules becomes half dissociated into single strands. Several equations for calculating the T_m of nucleic acids are well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41 * (\% \text{ G+C})$, when a nucleic acid is in aqueous solution at 1 M NaCl (see e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985). Other references (e.g., Allawi and SantaLucia, Biochemistry 36: 10581-94 (1997) include more sophisticated computations which account for structural, environmental, and sequence characteristics to calculate T_m . For example, in some embodiments these computations provide an improved estimate of T_m for short nucleic acid probes and targets (e.g., as used in the examples).

As used herein, the term “melting” when used in reference to a nucleic acid refers to the dissociation of a double-stranded nucleic acid or region of a nucleic acid into a single-stranded nucleic acid or region of a nucleic acid.

As used herein, the term “melting component” refers to a substance, molecule (e.g., a biomolecule), or a complex of more than one molecule (e.g., a complex of more than one biomolecule) that interacts with a nucleic acid and melts it, e.g., dissociates double-stranded regions (e.g., secondary structure of a single-stranded nucleic acid, a duplex structure of DNA or of a RNA/DNA hybrid) to provide single-stranded regions, e.g., to provide access to query regions for binding of a query probe. In exemplary embodiments, a melting component is a dCas9/gRNA complex (e.g., in some embodiments, comprising a biotinylated dCas9 and in some embodiments comprising a non-biotinylated dCas9). The technology is not limited, however, to melting components that comprise dCas9. The technology comprises use of any entity that provides access to a query region by a query probe and allows SiMREPS assay of a target nucleic acid.

As used herein, a “double-stranded nucleic acid” may be a portion of a nucleic acid, a region of a longer nucleic acid, or an entire nucleic acid. A “double-stranded nucleic acid” may be, e.g., without limitation, a double-stranded DNA, a double-stranded RNA, a double-stranded DNA/RNA hybrid, etc. A single-stranded nucleic acid having secondary structure (e.g., base-paired secondary structure) and/or higher order structure comprises a “double-stranded nucleic acid”. For example, triplex structures are considered to be “double-stranded”. In some embodiments, any base-paired nucleic acid is a “double-stranded nucleic acid”

As used herein, a “non-coding RNA” or “ncRNA” is a functional RNA molecule that is not translated into a protein. Less-frequently used synonyms are non-protein-coding RNA (npcRNA), non-messenger RNA (nmRNA), small non-messenger RNA

(snmRNA), and functional RNA (fRNA). The term small RNA (sRNA) is often used for bacterial ncRNAs. The DNA sequence from which a non-coding RNA is transcribed as the end product is often called an RNA gene or a non-coding RNA gene. Non-coding RNA genes include highly abundant and functionally important RNAs such as transfer RNA (tRNA) and ribosomal RNA (rRNA), as well as RNAs such as snoRNAs, microRNAs, siRNAs, and piRNAs. The number of ncRNAs encoded within the human genome is unknown, however recent transcriptomic and bioinformatic studies suggest the existence of thousands of ncRNAs. Since most of the newly identified ncRNAs have not been validated for their function, it is possible that many are non-functional.

As used herein, the term “long non-coding RNA” or “lncRNA” or “long ncRNA” refers to a non-protein coding RNA longer than approximately 200 nucleotides. As used herein, the term is used to distinguish lncRNAs from small regulatory RNAs such as microRNAs (miRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), and other short RNAs.

As used herein, the term “miRNA” refers to microRNA. As used herein, the term “miRNA target sequence” refers to a miRNA that is to be detected (e.g., in the presence of other nucleic acids). In some embodiments, a miRNA target sequence is a variant of a miRNA.

The term “siRNAs” refers to short interfering RNAs. In some embodiments, siRNAs comprise a duplex, or double-stranded region, where each strand of the double-stranded region is about 18 to 25 nucleotides long; the double-stranded region can be as short as 16, and as long as 29, base pairs long, where the length is determined by the antisense strand. Often siRNAs contain from about two to four unpaired nucleotides at the 3' end of each strand. SiRNAs appear to function as key intermediates in triggering RNA interference in invertebrates and in vertebrates, and in triggering sequence-specific RNA degradation during posttranscriptional gene silencing in plants. At least one strand of the duplex or double-stranded region of a siRNA is substantially homologous to or substantially complementary to a target RNA molecule. The strand complementary to a target RNA molecule is the “antisense” strand; the strand homologous to the target RNA molecule is the “sense” strand and is also complementary to the siRNA antisense strand. One strand of the double-stranded region need not be the exact length of the opposite strand; thus, one strand may have at least one fewer nucleotides than the opposite complementary strand, resulting in a “bubble” or at least one unmatched base in the opposite strand. One strand of the double-stranded region

need not be exactly complementary to the opposite strand; thus, the strand, preferably the sense strand, may have at least one mismatched base pair.

siRNAs may also contain additional sequences; non-limiting examples of such sequences include linking sequences, or loops, which connect the two strands of the duplex region. This form of siRNAs may be referred to “si-like RNA”, “short hairpin siRNA” where the short refers to the duplex region of the siRNA, or “hairpin siRNA”. Additional non-limiting examples of additional sequences present in siRNAs include stem and other folded structures. The additional sequences may or may not have known functions; non-limiting examples of such functions include increasing stability of an siRNA molecule, or providing a cellular destination signal.

“Pre-miRNA” or “pre-miR” means a non-coding RNA having a hairpin structure, which is the product of cleavage of a pri-miR by the double-stranded RNA-specific ribonuclease known as Drosha.

“Stem-loop sequence” means an RNA having a hairpin structure and containing a mature miRNA sequence. Pre-miRNA sequences and stem-loop sequences may overlap. Examples of stem-loop sequences are found in the miRNA database known as miRBase (available at the worldwide web at microma.sanger.ac.uk).

“Pri-miRNA” or “pri-miR” means a non-coding RNA having a hairpin structure that is a substrate for the double-stranded RNA-specific ribonuclease Drosha.

“miRNA precursor” means a transcript that originates from a genomic DNA and that comprises a non-coding, structured RNA comprising one or more miRNA sequences. For example, in certain embodiments a miRNA precursor is a pre-miRNA. In certain embodiments, a miRNA precursor is a pri-miRNA.

The term “gene” refers to a DNA sequence that comprises control and coding sequences necessary for the production of an RNA having a non-coding function (e.g., a ribosomal or transfer RNA), a polypeptide or a precursor. The RNA or polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or function is retained.

The term “wild-type” refers to a gene or a gene product that has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designated the “normal” or “wild-type” form of the gene. In contrast, the term “modified,” “mutant,” or “polymorphic” refers to a gene or gene product that displays modifications in sequence and or functional properties (i.e., altered characteristics) when compared to the wild-type gene or gene product. It is noted that

naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term “oligonucleotide” as used herein is defined as a molecule comprising two or more deoxyribonucleotides or ribonucleotides, preferably at least 5 nucleotides, more preferably at least about 10 to 15 nucleotides and more preferably at least about 15 to 30 nucleotides. The exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, PCR, or a combination thereof.

Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage, an end of an oligonucleotide is referred to as the “5' end” if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the “3' end” if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. A first region along a nucleic acid strand is said to be upstream of another region if the 3' end of the first region is before the 5' end of the second region when moving along a strand of nucleic acid in a 5' to 3' direction.

When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the “upstream” oligonucleotide and the latter the “downstream” oligonucleotide. Similarly, when two overlapping oligonucleotides are hybridized to the same linear complementary nucleic acid sequence, with the first oligonucleotide positioned such that its 5' end is upstream of the 5' end of the second oligonucleotide, and the 3' end of the first oligonucleotide is upstream of the 3' end of the second oligonucleotide, the first oligonucleotide may be called the “upstream” oligonucleotide and the second oligonucleotide may be called the “downstream” oligonucleotide.

As used herein, the terms “subject” and “patient” refer to any organisms including plants, microorganisms, and animals (e.g., mammals such as dogs, cats, livestock, and humans).

The term “sample” in the present specification and claims is used in its broadest sense. On the one hand it is meant to include a specimen or culture (e.g., microbiological

cultures). On the other hand, it is meant to include both biological and environmental samples. A sample may include a specimen of synthetic origin.

As used herein, a “biological sample” refers to a sample of biological tissue or fluid. For instance, a biological sample may be a sample obtained from an animal (including a human); a fluid, solid, or tissue sample; as well as liquid and solid food and feed products and ingredients such as dairy items, vegetables, meat and meat by-products, and waste. Biological samples may be obtained from all of the various families of domestic animals, as well as feral or wild animals, including, but not limited to, such animals as ungulates, bear, fish, lagomorphs, rodents, etc. Examples of biological samples include sections of tissues, blood, blood fractions, plasma, serum, urine, or samples from other peripheral sources or cell cultures, cell colonies, single cells, or a collection of single cells. Furthermore, a biological sample includes pools or mixtures of the above mentioned samples. A biological sample may be provided by removing a sample of cells from a subject, but can also be provided by using a previously isolated sample. For example, a tissue sample can be removed from a subject suspected of having a disease by conventional biopsy techniques. In some embodiments, a blood sample is taken from a subject. A biological sample from a patient means a sample from a subject suspected to be affected by a disease.

Environmental samples include environmental material such as surface matter, soil, water, and industrial samples, as well as samples obtained from food and dairy processing instruments, apparatus, equipment, utensils, disposable and non-disposable items. These examples are not to be construed as limiting the sample types applicable to the present invention.

The term “label” as used herein refers to any atom or molecule that can be used to provide a detectable (preferably quantifiable) effect, and that can be attached to a nucleic acid or protein. Labels include, but are not limited to, dyes (e.g., fluorescent dyes or moieties); radiolabels such as ^{32}P ; binding moieties such as biotin; haptens such as digoxigenin; luminogenic, phosphorescent, or fluorogenic moieties; mass tags; and fluorescent dyes alone or in combination with moieties that can suppress or shift emission spectra by fluorescence resonance energy transfer (FRET). Labels may provide signals detectable by fluorescence, radioactivity, colorimetry, gravimetry, X-ray diffraction or absorption, magnetism, enzymatic activity, characteristics of mass or behavior affected by mass (e.g., MALDI time-of-flight mass spectrometry; fluorescence polarization), and the like. A label may be a charged moiety (positive or negative charge)

or, alternatively, may be charge neutral. Labels can include or consist of nucleic acid or protein sequence, so long as the sequence comprising the label is detectable.

“Support” or “solid support”, as used herein, refers to a matrix on or in which nucleic acid molecules, microparticles, and the like may be immobilized, e.g., to which they may be covalently or noncovalently attached or in or on which they may be partially or completely embedded so that they are largely or entirely prevented from diffusing freely or moving with respect to one another.

As used herein, “moiety” refers to one of two or more parts into which something may be divided, such as, for example, the various parts of an oligonucleotide, a molecule, a chemical group, a domain, a probe, etc.

As used herein, a “query probe” or “reader probe” is any entity (e.g., molecule, biomolecule, etc.) that recognizes a nucleic acid (e.g., binds to a nucleic acid, e.g., binds specifically to a nucleic acid). In exemplary embodiments, the query probe is a protein that recognizes a nucleic acid (e.g., a nucleic acid binding protein, an antibody, antibody fragment, a transcription factor, or any other protein that binds to a particular sequence in a nucleic acid). In some other exemplary embodiments, the query probe is a nucleic acid (e.g., a DNA, an RNA, a nucleic acid comprising DNA and RNA, a nucleic acid comprising modified bases and/or modified linkages between bases; e.g., a nucleic acid as described hereinabove, a nucleic acid aptamer or any other nucleic acid that binds to a particular sequence in a nucleic acid). In some embodiments, the query probe is labeled, e.g., with a detectable label such as, e.g., a fluorescent moiety as described herein. In some embodiments, the query probe comprises more than one type of molecule (e.g., more than one of a protein, a nucleic acid, a chemical linker or a chemical moiety).

As used herein, a “capture probe” is any entity (e.g., molecule, biomolecule, etc.) that recognizes a nucleic acid (e.g., binds to a nucleic acid, e.g., binds specifically to a nucleic acid). In exemplary embodiments, the capture probe is a protein that recognizes a nucleic acid (e.g., a nucleic acid binding protein, an antibody, a fragment of an antibody, a transcription factor, or any other protein that binds to a particular sequence in a nucleic acid). In some other exemplary embodiments, a capture probe is a nucleic acid (e.g., a DNA, an RNA, a nucleic acid comprising DNA and RNA, a nucleic acid comprising modified bases and/or modified linkages between bases; e.g., a nucleic acid as described hereinabove). In some embodiments, a capture probe is labeled, e.g., with a detectable label such as, e.g., a fluorescent moiety as described herein. In some

embodiments, the capture probe comprises more than one type of molecule (e.g., more than one of a protein, a nucleic acid, a chemical linker or a chemical moiety).

Description

5 Provided herein are embodiments of a technique for the specific and ultrasensitive detection of single nucleic acids. As previously described (see, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety), SiMREPS uses total internal reflection fluorescence (TIRF) microscopy, single-molecule visualization, and kinetic analysis of binding and release of fluorescently labeled probes to target
10 molecules (see, e.g., Fig. 1 a). Target molecules are quantified by simple, amplification-free, direct counting upon kinetic fingerprint identification that provides for exquisite discrimination between single nucleotide variants, as demonstrated previously for the detection of microRNA (see, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety). The technology provided herein provides for the detection of
15 additional forms of nucleic acids, e.g., DNA, mutant DNA, methylated DNA, e.g., in an abundant wild-type background.

Existing techniques for nucleic acid detection utilize probes that form a thermodynamically stable complex with the target molecule, and are thus limited to weak and often unreliable thermodynamic discrimination against background signal or
20 spurious targets. In contrast, the technology described herein utilizes probes that repeatedly bind to the target molecule at the query region and related methods to record the large number of independent binding events that occur for each observed target molecule. This repeated kinetic sampling provides a unique kinetic “fingerprint” for the target and provides for a highly specific and sensitive detection of nucleic acids. In some
25 embodiments, the technology provides for the discrimination of two nucleic acid molecules that differ by as few as one nucleotide. In some embodiments, the technology provides for the discrimination of two nucleic acid molecules when one of the two nucleic acid molecules is present in a large excess (e.g., 10×; 100×; 1000×; 10,000×; or 1,000,000× or more in excess). See, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated
30 herein by reference in its entirety.

In some embodiments, a labeled nucleic acid is detected, e.g., using an instrument to detect a signal produced by the label. For instance, some embodiments comprise use of a detectably labeled (e.g., fluorescently labeled) query probe and a detector of fluorescence emission such a fluorescent microscopy technique. In some
35 embodiments, the technology finds use as a diagnostic tool for identifying mutant or

aberrantly expressed nucleic acid targets in biological samples. See, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety.

In some embodiments, this approach involves the capture of unlabeled nucleic acids by a dCas9/gRNA complex linked to a solid support (e.g., glass or fused silica) and melting of a query region, followed by observation of the repeated, transient binding of a short detectably labeled (e.g., fluorescently labeled) nucleic acid (e.g., DNA) query probe to the query region.

In some embodiments, the dCas9/gRNA complex is attached or fixed to a solid support. In some embodiments, the dCas9/gRNA complex comprises a moiety that provides for the immobilization of the dCas9/gRNA complex to a solid support by interaction of the moiety with a second moiety attached to the solid support. The dCas9/gRNA complex may be fixed directly or indirectly to a solid support.

Any of a variety of materials may be used as a support for the dCas9/gRNA complex, e.g., matrices or particles made of nitrocellulose, nylon, glass, polyacrylate, mixed polymers, polystyrene, silane polypropylene, and magnetically attractable materials. A planar surface is a preferred support for imaging by microscopy as described herein. A dCas9/gRNA complex may be immobilized by linking it directly to the solid support, e.g., by using any of a variety of covalent linkages, chelation, or ionic interaction, or may be immobilized by linking it indirectly via one or more linkers joined to the support. In some embodiments, the linker is a nucleic acid; in some embodiments, the linker is a nucleic acid comprising one or more nucleotides that is/are not intended to hybridize (e.g., that do not hybridize) to the target nucleic acid capture region but that are intended to act as a spacer between the dCas9/gRNA complex and its solid support.

In some embodiments, the dCas9/gRNA complex comprises a biotin group (e.g., the dCas9/gRNA complex is biotinylated) and the solid support comprises a streptavidin group (e.g., attached to the solid support by a linker moiety, e.g., a polyethylene glycol (PEG) linker). The specific interaction of the biotin and streptavidin thus immobilizes the capture probe to the solid support (Figs. 1a, 3, and 4).

Various other chemical methods can be employed for the immobilization of a dCas9/gRNA complex to a solid support. An example of such a method is to use a combination of a maleimide group and a thiol (–SH) group. In this method, a thiol (–SH) group is bonded to a dCas9/gRNA complex, and the solid support comprises a maleimide group. Accordingly, the thiol group of the dCas9/gRNA complex reacts with the maleimide group on the solid support to form a covalent bond, whereby the dCas9/gRNA

complex is immobilized. Introduction of the maleimide group can utilize a process of firstly allowing a reaction between a glass substrate and an aminosilane coupling agent and then introducing the maleimide group onto the glass substrate by a reaction of the amino group with an EMCS reagent (N-(6-maleimidocaproyloxy)succinimide, available from Dojindo). Introduction of the thiol group to a DNA can be carried out using 5'-Thiol-Modifier C6 (available from Glen Research) when the DNA is synthesized by an automatic DNA synthesizer.

Instead of the above-described combination of a thiol group and a maleimide group, a combination of, e.g., an epoxy group (on the solid support) and an amino group (dCas9/gRNA complex), is used in some embodiments as a combination of functional groups for immobilization. Surface treatments using various kinds of silane coupling agents are also effective. Other techniques for the attachment of proteins to solid supports and solid surfaces are known in the art.

15 Poisson processes

Embodiments of the technology are related to single-molecule recognition by recording the characteristic kinetics of a probe (e.g., a query probe) binding to a target (e.g., a query region). In particular embodiments, this process is a Poisson process. A Poisson process is a continuous-time stochastic process that counts the number of events and the time that events (e.g., transient binding of a detectably labeled (e.g., fluorescent) query probe to an immobilized target) occur in a given time interval. The time interval between each pair of consecutive events has an exponential distribution and each interval is assumed to be independent of other intervals. The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of the events occurring in the given time interval if these events occur with a known average rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area, or volume.

A Poisson distribution is a special case of the general binomial distribution where the number of trials n is large, the probability of success p is small, and the product $np = \lambda$ is moderate. In a Poisson process, the probability that a number of events N is j at any arbitrary time t follows the Poisson probability distribution $P_j(t)$:

$$P_j(t) = \frac{e^{-\lambda t} (\lambda t)^j}{j!}, j = 0, 1, 2, \dots \quad (1).$$

That is, the number N of events that occur up to time t has a Poisson distribution with parameter λt . Statistical and mathematical methods relevant to Poisson processes and Poisson distributions are known in the art. See, e.g., “Stochastic Processes (i): Poisson Processes and Markov Chains” in *Statistics for Biology and Health – Statistical Methods in Bioinformatics* (Ewans and Grant, eds.), Springer (New York, 2001), page 129 et seq., incorporated herein by reference in its entirety. Software packages such as Matlab and R may be used to perform mathematical and statistical methods associated with Poisson processes, probabilities, and distributions.

Kinetics of detection

Particular embodiments of the technology are related to detecting a nucleic acid by analyzing the kinetics of the interaction of a query probe with a query region of a target nucleic acid to be detected. For the interaction of a query probe Q (e.g., at an equilibrium concentration $[Q]$) with a target nucleic acid T (e.g., at an equilibrium concentration $[T]$), the kinetic rate constant k_{on} describes the time-dependent formation of the complex QT comprising the query probe Q hybridized to the query region of the target nucleic acid T. In particular embodiments, while the formation of the QT complex is associated with a second order rate constant that is dependent on the concentration of query probe and has units of $M^{-1}min^{-1}$ (or the like), the formation of the QT complex is sufficiently described by a k_{on} that is a pseudo-first order rate constant associated with the formation of the QT complex. Thus, as used herein, k_{on} is an apparent (“pseudo”) first-order rate constant.

Likewise, the kinetic rate constant k_{off} describes the time-dependent dissociation of the complex QT into the query probe Q and the target nucleic acid T. Kinetic rates are typically provided herein in units of min^{-1} or s^{-1} . The “dwell time” of the query probe Q in the bound state (τ_{on}) is the time interval (e.g., length of time) that the probe Q is hybridized to the query region of the target nucleic acid T during each instance of query probe Q binding to the query region of the target nucleic acid T to form the QT complex. The “dwell time” of the query probe Q in the unbound state (τ_{off}) is the time interval (e.g., length of time) that the probe Q is not hybridized to the query region of the target nucleic acid T between each instance of query probe Q binding to the query region of the target nucleic acid T to form the QT complex (e.g., the time the query probe Q is dissociated from the target nucleic acid T between successive binding events of the

query probe Q to the target nucleic acid T). Dwell times may be provided as averages or weighted averages integrating over numerous binding and non-binding events.

Further, in some embodiments, the repeated, stochastic binding of query probes (e.g., detectably labeled query probes (e.g., fluorescent probes), e.g., nucleic acid probes such as DNA or RNA probes) to immobilized targets is modeled as a Poisson process occurring with constant probability per unit time and in which the standard deviation in the number of binding and dissociation events per unit time (N_{b+d}) increases as $(N_{b+d})^{1/2}$. Thus, the statistical noise becomes a smaller fraction of N_{b+d} as the observation time is increased. Accordingly, the observation is lengthened as needed in some embodiments to achieve discrimination between target and off-target binding. And, as the acquisition time is increased, the signal and background peaks in the N_{b+d} histogram become increasingly separated and the width of the signal distribution increases as the square root of N_{b+d} , consistent with kinetic Monte Carlo simulations. An acquisition time of approximately 10 minutes (e.g., approximately 1 to 100 minutes, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 minutes) yields sufficient (e.g., complete) separation of the signal from background distributions of N_{b+d} , providing for substantially background-free quantification of the target. See, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety.

Further, in some embodiments the probe length is chosen to provide sufficient separation of signal and background peaks on convenient experimental time scales. In particular, the kinetics of query probe exchange are related to the number of complementary bases between the query probe and target nucleic acid. For instance, in some embodiments, the interaction of a short DNA query probe with its complement increases as an approximately exponential function of the number of base pairs formed, while the rate constant of binding is affected only weakly for interactions comprising at least 6 to 7 base pairs. Thus, varying query probe length provides for tuning the kinetic behavior to improve discrimination of query probe binding events to the target from background binding. In particular, a query (e.g., fluorescent) probe length of 9 nt to 10 nt (providing theoretical T_m values of 17.5°C to 25°C) yields rapid target binding that is distinguished from background signal, as displayed in histograms of intensity transitions per candidate molecule in the presence and absence of target. See, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety. Further, in some embodiments the kinetics of binding and dissociation are more closely correlated to probe length than to the melting temperature of the duplex. While some embodiments comprise use of a probe having a length of 9 to 10 nt, the technology is not

limited by this length. Indeed, use of probes longer or shorter than 9 to 10 nt is contemplated by the technology, e.g., as discussed throughout.

Although the disclosure herein refers to certain illustrated embodiments, it is to be understood that these embodiments are presented by way of example and not by way of limitation.

Detection of double-stranded nucleic acids

Embodiments of the technology provide for the detection of double-stranded nucleic acids. Some embodiments provide compositions, reaction mixtures, and complexes comprising a plurality of molecules for detecting one or more nucleic acids. Some embodiments of compositions, reaction mixtures, and complexes comprise a nucleic acid (e.g., a target nucleic acid) that is to be detected, identified, quantified, and/or characterized; a solid substrate comprising a dCas9/gRNA complex linked to a solid surface that binds one or more regions of the target with high specificity; and a detectably labeled (e.g., fluorescent) query probe. Some embodiments further comprise a protospacer adjacent motif (PAM) DNA oligonucleotide.

The SiMREPS technology exploits the direct binding of a short (6-12-nucleotide) fluorescently labeled DNA probe to an unlabeled nucleic acid target (e.g., miRNA) immobilized on a glass surface (Fig. 1 a) (see, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety). Using TIRF microscopy (Walter et al (2008) "Do-it-yourself guide: how to use the modern single-molecule toolkit" *Nat Methods* 5: 475-89), both non-specific surface binding (Fig. 1 b) and specific binding to the immobilized target (Fig. 1 c) are detected. However, equilibrium binding of the probe to target yields a distinctive kinetic signature, or fingerprint, that can achieve ultra-high discrimination against background binding (compare Fig. 1 b with Fig. 1 c). Since the transient binding of probes to an immobilized target resembles a Poisson process, the standard deviation in the number of binding and dissociation events (N_{b+d}) increases as $\sqrt{N_{b+d}}$. As experimental acquisition time is increased, the signal and background peaks in histograms of N_{b+d} are progressively better-resolved (Fig. 1 d) (see, e.g., Johnson-Bucket et al. (2015) "Kinetic fingerprinting to identify and count single nucleic acids" *Nat Biotechnol* 33: 730-2), allowing for arbitrarily high discrimination between target and off-target binding.

SiMREPS finds use in quantifying RNA (see, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety). In particular, previous experiments quantified four human miRNAs that are dysregulated in cancer and other

diseases. Discrimination (e.g., specificity = 1) was achieved for all target-probe pairs and standard curves showed linear dependence on target concentration (Fig. 1 e). A single fluorescent probe discriminates with highest specificity between two microRNAs differing by a single nucleotide (Fig. 2 a-c). SiMREPS detects targets in complex biological matrices. (Fig. 2 d, e). Also the prostate cancer biomarker hsa-miR-1418 was detected in a serum sample after spiking-in varying target concentrations. The measured concentration was strongly correlated with the nominal spiked-in concentration (Fig. 2 f, $R > 0.999$, slope = 1.07).

One challenge in applying the SiMREPS technology to the detection of double-stranded nucleic acids is that the transient binding of a query probe to the target nucleic acid (e.g., at the query region) competes with association of the complementary strand to the query region (e.g., dsDNA) at the target locus. Accordingly, provided herein is a technology in which SiMREPS finds use in detecting double-stranded DNA. To overcome this challenge, some embodiments of the technology comprise use of a catalytically inactive ("dead") dCas9 enzyme loaded with a specific guide-RNA (gRNA) to melt dsDNA structure locally in a sequence-specific fashion, providing access for the SiMREPS probe (Fig. 3). Related embodiments comprise use of a protein that binds to double-stranded nucleic acids (e.g., double-stranded DNA, double-stranded RNA, a double-stranded DNA/RNA hybrid) and/or a melting component (e.g., a protein or a nucleic acid) to dissociate a double-stranded nucleic acid (e.g., a region of a nucleic acid) to provide a single-stranded nucleic acid.

dCas9/gRNA complexes

The technology comprises use of a sequence-specific nucleic acid binding component (e.g., molecule, biomolecule, or complex of one or more molecules and/or biomolecules) to immobilize nucleic acids and/or convert double-stranded nucleic acids (e.g., regions of double-stranded nucleic acids) to single-stranded regions. In exemplary embodiments, the sequence-specific nucleic acid binding component comprises an enzymatically inactive, or "dead", Cas9 protein ("dCas9") and a guide RNA ("gRNA"). While nucleic acid-binding molecules such as the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) (CRISPR/Cas) system have been used extensively for genome editing in cells of various types and species, recombinant and engineered nucleic acid-binding proteins find use in the present technology to melt double-stranded nucleic acids and provide single-stranded nucleic acids for probe binding.

The Cas9 protein was discovered as a component of the bacterial adaptive immune system (see, e.g., Barrangou et al. (2007) "CRISPR provides acquired resistance against viruses in prokaryotes" *Science* 315: 1709–1712). Cas9 is an RNA-guided endonuclease that targets and destroys foreign DNA in bacteria using RNA:DNA base-pairing between the gRNA and foreign DNA to provide sequence specificity. Recently, Cas9/gRNA complexes have found use in genome editing (see, e.g., Doudna et al. (2014) "The new frontier of genome engineering with CRISPR-Cas9" *Science* 346: 6213).

Accordingly, some Cas9/RNA complexes comprise two RNA molecules: (1) a CRISPR RNA (crRNA), possessing a nucleotide sequence complementary to the target nucleotide sequence; and (2) a trans-activating crRNA (tracrRNA). In this mode, Cas9 functions as an RNA-guided nuclease that uses both the crRNA and tracrRNA to recognize and cleave a target sequence. Recently, a single chimeric guide RNA (sgRNA) mimicking the structure of the annealed crRNA/tracrRNA has become more widely used than crRNA/tracrRNA because the gRNA approach provides a simplified system with only two components (e.g., the Cas9 and the sgRNA). Thus, sequence-specific binding to a nucleic acid can be guided by a natural dual-RNA complex (e.g., comprising a crRNA, a tracrRNA, and Cas9) or a chimeric single-guide RNA (e.g., a sgRNA and Cas9). (see, e.g., Jinek et al. (2012) "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity" *Science* 337:816–821).

As used herein, the targeting region of a crRNA (2-RNA system) or a sgRNA (single guide system) is referred to as the "guide RNA" (gRNA). In some embodiments, the gRNA comprises, consists of, or essentially consists of 10 to 50 bases, e.g., 15 to 40 bases, e.g., 15 to 30 bases, e.g., 15 to 25 bases (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 bases). Methods are known in the art for determining the length of the gRNA that provides the most efficient target recognition for a Cas9. See, e.g., Lee et al. (2016) "The *Neisseria meningitidis* CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells" *Molecular Therapy*, 19 January 2016; doi:10.1038/mt.

Accordingly, in some embodiments the gRNA is a short synthetic RNA comprising a "scaffold" sequence for Cas9-binding and a user-defined approximately 20-nucleotide "targeting" sequence that is complementary to the nucleic acid target.

In some embodiments, DNA targeting specificity is determined by two factors: 1) a DNA sequence matching the gRNA targeting sequence and a protospacer adjacent motif (PAM) directly downstream of the target sequence. Some Cas9/gRNA complexes

recognize a DNA sequence comprising a protospacer adjacent motif (PAM) sequence and the adjacent approximately 20 bases complementary to the gRNA. Canonical PAM sequences are NGG or NAG for Cas9 from *Streptococcus pyogenes* and NNNNGATT for the Cas9 from *Neisseria meningitidis*. Following DNA recognition by hybridization of
5 the gRNA to the DNA target sequence, Cas9 cleaves the DNA sequence via an intrinsic nuclease activity. For genome editing and other purposes, the CRISPR/Cas system from *S. pyogenes* has been used most often. Using this system, one can target a given target nucleic acid (e.g., for editing or other manipulation) by designing a gRNA having nucleotide sequence complementary to an approximately 20-base DNA sequence 5'-
10 adjacent to the PAM. Methods are known in the art for determining the PAM sequence that provides the most efficient target recognition for a Cas9. See, e.g., Zhang et al. (2013) "Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*" *Molecular Cell* 50: 488-503; Lee et al., supra.

The present technology comprises use of a catalytically inactive form of Cas9
15 ("dead Cas9" or "dCas9"), in which point mutations are introduced that disable the nuclease activity. In some embodiments, the dCas9 protein is from *S. pyogenes*. In some embodiments, the dCas9 protein comprises mutations at, e.g., D10, E762, H983, and/or D986; and at H840 and/or N863, e.g., at D10 and H840, e.g., D10A or D10N and H840A or H840N or H840Y. In some embodiments, the dCas9 is provided as a fusion protein
20 comprising a functional domain for attaching the dCas9 to a solid surface (e.g., an epitope tag, linker peptide, etc.)

The dCas9/gRNA complex binds to a target nucleic acid with a sequence specificity provided by the gRNA, but does not cleave the nucleic acid. In this form, the dCas9/gRNA "melts" the target sequence to provide single-stranded regions of the target
25 nucleic acid in a sequence-specific manner (see, e.g., Qi et al. (2013) "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression" *Cell* 152(5): 1173-83).

Furthermore, while the Cas9/gRNA system and dCas9/gRNA system initially targeted sequences adjacent to a PAM, the dCas9/gRNA system as used herein has been
30 engineered to target any nucleotide sequence for binding. Also, Cas9 and dCas9 orthologs encoded by compact genes (e.g., Cas9 from *Staphylococcus aureus*) are known (see, e.g., Ran et al. (2015) "In vivo genome editing using *Staphylococcus aureus* Cas9" *Nature* 520: 186-191), which improves the cloning and manipulation of the Cas9 components in vitro.

A number of bacteria express Cas9 protein variants. The Cas9 from *Streptococcus pyogenes* is presently the most commonly used; some of the other Cas9 proteins have high levels of sequence identity with the *S. pyogenes* Cas9 and use the same guide RNAs. Others are more diverse, use different gRNAs, and recognize
 5 different PAM sequences as well (the 2-5 nucleotide sequence specified by the protein which is adjacent to the sequence specified by the RNA). Chylinski et al. classified Cas9 proteins from a large group of bacteria (RNA Biology 10:5, 1-12; 2013), and a large number of Cas9 proteins are listed in supplementary FIG. 1 and supplementary table 1 thereof, which are incorporated by reference herein. Additional Cas9 proteins are
 10 described in Esvelt et al., Nat Methods. 2013 November; 10(11):1116-21 and Fonfara et al., "Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems." Nucleic Acids Res. 2013 Nov. 22. [Epub ahead of print] doi:10.1093/nar/gkt1074.

Cas9, and thus dCas9, molecules of a variety of species find use in the technology
 15 described herein. While the *S. pyogenes* and *S. thermophilus* Cas9 molecules are widely used, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed herein find use in embodiments of the technology. Accordingly, the technology provides for the replacement of *S. pyogenes* and *S. thermophilus* Cas9 and dCas9 molecules with Cas9 and dCas9 molecules from the other species can replace them, e.g.:
 20

	GenBank Acc No.	Bacterium
	303229466	<i>Veillonella atypica</i> ACS-134-V-Col7a
	34762592	<i>Fusobacterium nucleatum</i> subsp. <i>vincentii</i>
25	374307738	<i>Filifactor alocis</i> ATCC 35896
	320528778	<i>Solobacterium moorei</i> F0204
	291520705	<i>Coprococcus catus</i> GD-7
	42525843	<i>Treponema denticola</i> ATCC 35405
	304438954	<i>Peptoniphilus duerdenii</i> ATCC BAA-1640
30	224543312	<i>Catenibacterium mitsuokai</i> DSM 15897
	24379809	<i>Streptococcus mutans</i> UA159
	15675041	<i>Streptococcus pyogenes</i> SF370
	16801805	<i>Listeria innocua</i> Clip11262
	116628213	<i>Streptococcus thermophilus</i> LMD-9
35	323463801	<i>Staphylococcus pseudintermedius</i> ED99
	352684361	<i>Acidaminococcus intestini</i> RyC-MR95
	302336020	<i>Olsenella uli</i> DSM 7084
	366983953	<i>Oenococcus kitaharae</i> DSM 17330
	310286728	<i>Bifidobacterium bifidum</i> S17
40	258509199	<i>Lactobacillus rhamnosus</i> GG
	300361537	<i>Lactobacillus gasseri</i> JV-V03
	169823755	<i>Fingoldia magna</i> ATCC 29328
	47458868	<i>Mycoplasma mobile</i> 163K
	284931710	<i>Mycoplasma gallisepticum</i> str. F
45	363542550	<i>Mycoplasma ovipneumoniae</i> SC01
	384393286	<i>Mycoplasma canis</i> PG 14
	71894592	<i>Mycoplasma synoviae</i> 53
	238924075	<i>Eubacterium rectale</i> ATCC 33656

	116627542	<i>Streptococcus thermophilus</i> LMD-9
	315149830	<i>Enterococcus faecalis</i> TX0012
	315659848	<i>Staphylococcus lugdunensis</i> M23590
5	160915782	<i>Eubacterium dolichum</i> DSM 3991
	336393381	<i>Lactobacillus coryniformis</i> subsp. <i>torquens</i>
	310780384	<i>Llyobacter polytropus</i> DSM 2926
	325677756	<i>Ruminococcus albus</i> 8
	187736489	<i>Akkermansia muciniphila</i> ATCC BAA-835
	117929158	<i>Acidothermus cellulolyticus</i> 11B
10	189440764	<i>Bifidobacterium longum</i> DJO10A
	283456135	<i>Bifidobacterium dentium</i> Bd1
	38232678	<i>Corynebacterium diphtheriae</i> NCTC 13129
	187250660	<i>Elusimicrobium minutum</i> Pei191
	319957206	<i>Nitratifractor salsuginis</i> DSM 16511
15	325972003	<i>Sphaerochaeta globus</i> str. Buddy
	261414553	<i>Fibrobacter succinogenes</i> subsp. <i>succinogenes</i>
	60683389	<i>Bacteroides fragilis</i> NCTC 9343
	256819408	<i>Capnocytophaga ochracea</i> DSM 7271
	90425961	<i>Rhodopseudomonas palustris</i> BisB18
20	373501184	<i>Prevotella micans</i> F0438
	294674019	<i>Prevotella ruminicola</i> 23
	365959402	<i>Flavobacterium columnare</i> ATCC 49512
	312879015	<i>Aminomonas paucivorans</i> DSM 12260
	83591793	<i>Rhodospirillum rubrum</i> ATCC 11170
25	294086111	<i>Candidatus Puniceispirillum marinum</i> IMCC1322
	121608211	<i>Verminephrobacter eiseniae</i> EF01-2
	344171927	<i>Ralstonia syzygii</i> R24
	159042956	<i>Dinoroseobacter shibae</i> DFL 12
30	288957741	<i>Azospirillum</i> sp. B510
	92109262	<i>Nitrobacter hamburgensis</i> X14
	148255343	<i>Bradyrhizobium</i> sp. BTAi1
	34557790	<i>Wolinella succinogenes</i> DSM 1740
	218563121	<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>
	291276265	<i>Helicobacter mustelae</i> 12198
35	229113166	<i>Bacillus cereus</i> Rock1-15
	222109285	<i>Acidovorax ebreus</i> TPSY
	189485225	uncultured Termite group 1
	182624245	<i>Clostridium perfringens</i> D str.
	220930482	<i>Clostridium cellulolyticum</i> H10
40	154250555	<i>Parvibaculum lavamentivorans</i> DS-1
	257413184	<i>Roseburia intestinalis</i> L1-82
	218767588	<i>Neisseria meningitidis</i> Z2491
	15602992	<i>Pasteurella multocida</i> subsp. <i>multocida</i>
	319941583	<i>Sutterella wadsworthensis</i> 3 1
45	254447899	gamma proteobacterium HTCC5015
	54296138	<i>Legionella pneumophila</i> str. Paris
	331001027	<i>Parasutterella excrementihominis</i> YIT 11859
	34557932	<i>Wolinella succinogenes</i> DSM 1740
50	118497352	<i>Francisella novicida</i> U112

The technology described herein encompasses the use of a dCas9 derived from any Cas9 protein (e.g., as listed above) and their corresponding guide RNAs or other guide RNAs that are compatible. The Cas9 from *Streptococcus thermophilus* LMD-9 CRISPR1 system has been shown to function in human cells (see, e.g., Cong et al. (2013) Science 339: 819). Additionally, Jinek showed in vitro that Cas9 orthologs from *S. thermophilus* and *L. innocua*, can be guided by a dual *S. pyogenes* gRNA to cleave target plasmid DNA.

In some embodiments, the present technology comprises the Cas9 protein from *S. pyogenes*, either as encoded in bacteria or codon-optimized for expression in mammalian cells, containing mutations at D10, E762, H983, or D986 and H840 or N863, e.g., D10A/D10N and H840A/H840N/H840Y, to render the nuclease portion of the protein catalytically inactive; substitutions at these positions are, in some embodiments, alanine (Nishimasu (2014) Cell 156: 935-949) or, in some embodiments, other residues, e.g., glutamine, asparagine, tyrosine, serine, or aspartate, e.g., E762Q, H983N, H983Y, D986N, N863D, N863S, or N863H. The sequence of one *S. pyogenes* dCas9 protein that finds use in the technology provided herein is described in US20160010076, which is incorporated herein by reference in its entirety.

For example, in some embodiments, the dCas9 used herein is at least about 50% identical to the sequence of *S. pyogenes* Cas9, e.g., at least 50% identical to the following sequence of dCas9 comprising the D10A and H840A substitutions (SEQ ID NO: 1).

15	Met	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Ala	Ile	Gly	Thr	Asn	Ser	Val	1	5	10	15
	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro	Ser	Lys	Lys	Phe	20	25	30	
	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	Lys	Asn	Leu	Ile	35	40	45	
20	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala	Thr	Arg	Leu	50	55	60	
	Lys	Arg	Thr	Ala	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg	Ile	Cys		65	70	75	80
25	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp	Ser	85	90	95	
	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys	100	105	110	
	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr	115	120	125	
30	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp	130	135	140	
	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His	145	150	155	160
35	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro	165	170	175	
	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr	180	185	190	
	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala	195	200	205	
40	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	210	215	220	
	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn	225	230	235	240
45	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	245	250	255	
	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	260	265	270	
50	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	275	280	285	

	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp
	290						295					300				
	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser
	305					310					315					320
5	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys
					325					330					335	
	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe
				340					345					350		
10	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser
			355					360					365			
	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp
	370						375					380				
	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg
	385					390					395					400
15	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu
					405					410					415	
	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe
				420					425					430		
20	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile
		435						440					445			
	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp
		450					455					460				
	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu
	465					470					475					480
25	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr
					485					490					495	
	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser
				500					505					510		
30	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys
		515						520					525			
	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln
		530					535					540				
	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr
	545					550					555					560
35	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp
					565					570					575	
	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly
				580					585					590		
40	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp
			595					600					605			
	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr
		610					615						620			
	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala
	625					630					635					640
45	His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr
					645					650					655	
	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp
				660					665					670		
50	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe
			675					680					685			
	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe
		690					695						700			
	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu
	705					710					715					720
55	His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly
					725					730					735	
	Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly
				740					745					750		
60	Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln
			755					760					765			
	Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile

		770				775				780							
		Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro
		785					790					795					800
5		Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu
						805					810					815	
		Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg
						820				825					830		
		Leu	Ser	Asp	Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys
						835			840						845		
10		Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg
						850			855					860			
		Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys
		865				870						875					880
15		Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys
						885					890					895	
		Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp
						900				905					910		
		Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr
						915			920						925		
20		Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp
						930			935				940				
		Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser
		945				950					955						960
25		Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg
						965					970					975	
		Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val
						980				985					990		
		Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe
						995			1000					1005			
30		Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala	
						1010			1015					1020			
		Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe	
						1025			1030					1035			
35		Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala	
						1040			1045					1050			
		Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu	
						1055			1060					1065			
		Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val	
						1070			1075					1080			
40		Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr	
						1085			1090					1095			
		Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys	
						1100			1105					1110			
45		Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro	
						1115			1120					1125			
		Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val	
						1130			1135					1140			
		Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys	
						1145			1150					1155			
50		Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser	
						1160			1165					1170			
		Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys	
						1175			1180					1185			
55		Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	
						1190			1195					1200			
		Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly	
						1205			1210					1215			
		Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val	
						1220			1225					1230			
60		Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser	
						1235			1240					1245			

	Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys
	1250						1255					1260			
	His	Tyr	Leu	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser	Lys
	1265						1270					1275			
5	Arg	Val	Ile	Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser	Ala
	1280						1285					1290			
	Tyr	Asn	Lys	His	Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala	Glu	Asn
	1295						1300					1305			
10	Ile	Ile	His	Leu	Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala	Ala
	1310						1315					1320			
	Phe	Lys	Tyr	Phe	Asp	Thr	Thr	Ile	Asp	Arg	Lys	Arg	Tyr	Thr	Ser
	1325						1330					1335			
	Thr	Lys	Glu	Val	Leu	Asp	Ala	Thr	Leu	Ile	His	Gln	Ser	Ile	Thr
	1340						1345					1350			
15	Gly	Leu	Tyr	Glu	Thr	Arg	Ile	Asp	Leu	Ser	Gln	Leu	Gly	Gly	Asp
	1355						1360					1365			

In some embodiments, the technology comprises use of a nucleotide sequence that is approximately 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to a nucleotide sequence that encodes a protein described by SEQ ID NO: 1.

In some embodiments, the dCas9 used herein is at least about 50% identical to the sequence of the catalytically inactive *S. pyogenes* Cas9, i.e., at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ ID NO:1, wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

In some embodiments, any differences from SEQ ID NO:1 are in non-conserved regions, as identified by sequence alignment of sequences set forth in Chylinski et al., RNA Biology 10:5, 1-12; 2013 (e.g., in supplementary FIG. 1 and supplementary table 1 thereof); Esvelt et al., Nat Methods. 2013 November; 10(11):1116-21 and Fonfara et al., Nucl. Acids Res. (2014) 42 (4): 2577-2590. [Epub ahead of print 2013 Nov. 22] doi:10.1093/nar/gkt1074, and wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

To determine the percent identity of two sequences, the sequences are aligned for optimal comparison purposes (gaps are introduced in one or both of a first and a second amino acid or nucleic acid sequence as required for optimal alignment, and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 50% (in some embodiments, about 50%, 55%, 60%, 65%, 70%, 75%, 85%, 90%, 95%, or 100% of the length of the reference sequence) is aligned. The nucleotides or residues at corresponding positions are then compared. When a position in the first sequence is occupied by the same nucleotide or residue as the corresponding position in the second

sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

5 The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For purposes of the present application, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software
10 package, using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

 Accordingly, provided herein is a dCas9/gRNA-based approach for detecting double-stranded nucleic acids (e.g., double-stranded genomic DNA) using SiMREPS probes. Embodiments of the technology comprise capturing unlabeled genomic DNA
15 targets on a glass or fused silica surface using a gRNA-loaded, enzymatically dead dCas9 enzyme or enzymes that bind one or more segments of the target with high specificity and stability (e.g., forming approximately 20 base pairs at the site complementary to the gRNA; Fig. 3). Surface capture of nucleic acid targets by biotinylated dCas9 is followed by observing repeated, transient binding of the SiMREPS
20 probe to a second segment of the target that has been made accessible by dCas9/gRNA-mediated melting.

 Thus, in some embodiments a biotinylated dCas9 comprising a guide RNA (gRNA) comprising an appropriate sequence captures a target nucleic acid. In addition, dCas9, with the help of the gRNA, melts the DNA, which produces a complementary
25 non-template DNA strand that is single-stranded and accessible to the query probe. In some embodiments, the technology comprises use of a second non-biotinylated dCas9 comprising a second gRNA that binds to the target nucleic acid (Fig. 4). The second dCas9/gRNA binds to the target nucleic acid at a distance from the first dCas9/gRNA that is approximately the size of the query probe, e.g., 5 to 30 nucleotides. That is, the
30 biotinylated dCas9/gRNA (capture dCas9/gRNA) and second dCas9/gRNA bind to regions adjacent to the region of the target nucleic acid to which the query probe binds. The two dCas9/gRNA complexes melt the region of nucleic acid between them to provide a single-stranded query region accessible for query probe binding. Accordingly, the spacing between the two dCas9/gRNA complexes has is appropriate for the binding of a
35 query probe between them.

Methods

In some embodiments, the technology provides a method for detecting a double-stranded nucleic acid. For example, in some embodiments a nucleic acid (e.g., a genomic DNA) target DNA is briefly pre-treated with dCas9/gRNA (e.g., at or near ambient ("room") temperature). Next, the guide RNA (gRNA) of the dCas9/gRNA complex hybridizes to the target nucleic acid (e.g., a genomic DNA) at a region adjacent to a query region (e.g., a region of the target nucleic acid that is complementary to a query probe). The dCas9/gRNA melts a specific DNA sequence (e.g., the query region) in the target nucleic acid. Methods comprise capturing the dCas9 (e.g., a biotinylated dCas9) onto a slide surface (e.g., by biotin-avidin interaction). Following capture and immobilization of the dCas9/gRNA-target nucleic acid to the surface, SiMREPS is used to detect binding of a query probe to the query region in the target nucleic acid, which is adjacent to the site of target nucleic acid hybridized to the gRNA. See, e.g., Fig. 3.

In some embodiments, two dCas9/gRNA complexes are used to make the query region of a target nucleic acid accessible to a query probe. In some embodiments, hybridizing two dCas9/gRNA complexes to flank the query region improves accessibility of the target nucleic acid (e.g., the query region) to the query probe for SiMREPS-based detection. One dCas9 is biotinylated for capture onto the surface; the other (e.g., second) dCas9 is optionally biotinylated (in preferred embodiments, the second dCas9 is not biotinylated). See, e.g., Fig. 4. The space between the regions bound by the two dCas9/gRNA complexes bound to the target nucleic acid provides appropriate space for binding of the query probe to the query region of the target nucleic acid.

In some embodiments, the detectable (e.g., fluorescent) query probe produces a fluorescence emission signal when it is close to the surface of the solid support (e.g., within about 100 nm of the surface of the solid support). When unbound, query probes quickly diffuse and thus are not individually detected; accordingly, when in the unbound state, the query probes produce a low level of diffuse background fluorescence. Consequently, in some embodiments detection of bound query probes comprises use of total internal reflection fluorescence microscopy (TIRF), HiLo microscopy (see, e.g., US20090084980, EP2300983 B1, WO2014018584 A1, WO2014018584 A1, incorporated herein by reference), confocal scanning microscopy, or other technologies comprising illumination schemes that illuminate (e.g., excite) only those query probe molecules near or on the surface of the solid support. Thus, in some embodiments, only query probes that are bound to an immobilized target near or on the surface produce a point-like

emission signal (e.g., a “spot”) that can be confirmed as originating from a single molecule.

In general terms, the observation comprises monitoring fluorescence emission at a number of discrete locations on the solid support where the target nucleic acids are immobilized (e.g., by being specifically bound to the dCas9/gRNA attached to the surface), e.g., at a number of fluorescent “spots” that blink, e.g., that can be in “on” and “off” states. The presence of fluorescence emission (spot is “on”) and absence of fluorescence emission (spot is “off”) at each discrete location (e.g., at each “spot” on the solid support) are recorded. Each spot “blinks” – e.g., a spot alternates between “on” and “off” states, respectively, as a query probe binds to the immobilized target nucleic acid at that spot and as the query probe dissociates from the immobilized target nucleic acid at that spot.

The data collected provide for the determination of the number of times a query probe binds to each immobilized target (e.g., the number of times each spot blinks “on”) and a measurement of the amount of time a query probe remains bound (e.g., the length of time a spot remains “on” before turning “off”).

In some embodiments, the query probe comprises a fluorescent label having an emission wavelength. Detection of fluorescence emission at the emission wavelength of the fluorescent label indicates that the query probe is bound to an immobilized target nucleic acid. Binding of the query probe to the target nucleic acid is a “binding event”. In some embodiments of the technology, a binding event has a fluorescence emission having a measured intensity greater than a defined threshold. For example, in some embodiments a binding event has a fluorescence intensity that is above the background fluorescence intensity (e.g., the fluorescence intensity observed in the absence of a target nucleic acid). In some embodiments, a binding event has a fluorescence intensity that is at least 1, 2, 3, 4 or more standard deviations above the background fluorescence intensity (e.g., the fluorescence intensity observed in the absence of a target nucleic acid). In some embodiments, a binding event has a fluorescence intensity that is at least 2 standard deviations above the background fluorescence intensity (e.g., the fluorescence intensity observed in the absence of a target nucleic acid). In some embodiments, a binding event has a fluorescence intensity that is at least 1.5, 2, 3, 4, or 5 times the background fluorescence intensity (e.g., the mean fluorescence intensity observed in the absence of a target nucleic acid).

Accordingly, in some embodiments detecting fluorescence at the emission wavelength of the fluorescent probe that has an intensity above the defined threshold

(e.g., at least 2 standard deviations greater than background intensity) indicates that a binding event has occurred (e.g., at a discrete location on the solid support where a target nucleic acid is immobilized). Also, in some embodiments detecting fluorescence at the emission wavelength of the fluorescent probe that has an intensity above the defined threshold (e.g., at least 2 standard deviations greater than background intensity) indicates that a binding event has started. Accordingly, in some embodiments detecting an absence of fluorescence at the emission wavelength of the fluorescent probe that has an intensity above the defined threshold (e.g., at least 2 standard deviations greater than background intensity) indicates that a binding event has ended (e.g., the query probe has dissociated from the target nucleic acid). The length of time between when the binding event started and when the binding event ended (e.g., the length of time that fluorescence at the emission wavelength of the fluorescent probe having an intensity above the defined threshold (e.g., at least 2 standard deviations greater than background intensity) is detected) is the dwell time of the binding event. A “transition” refers to the binding and dissociation of a query probe to the target nucleic acid (e.g., an on/off event).

Methods according to the technology comprise counting the number of query probe binding events that occur at each discrete location on the solid support during a defined time interval that is the “acquisition time” (e.g., a time interval that is tens to hundreds to thousands of seconds, e.g., 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 seconds; e.g., 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or 0 minutes; e.g., 1, 1.5, 2, 2.5, or 3 hours). In some embodiments, the acquisition time is approximately 1 to 10 seconds to 1 to 10 minutes (e.g., approximately 1 to 100 seconds, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 seconds, e.g., 1 to 100 minutes, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 minutes).

Further, the length of time the query probe remains bound to the target nucleic acid during a binding event is the “dwell time” of the binding event. The number of binding events detected during the acquisition time and/or the lengths of the dwell times recorded for the binding events is/are characteristic of a query probe binding to a target nucleic acid and thus provide an indication that the target nucleic acid is immobilized at said discrete location and thus that the target nucleic acid is present in the sample.

Binding of the query probe to the immobilized target nucleic acid and/or and dissociation of the query probe from the immobilized target nucleic acid is/are monitored (e.g., using a light source to excite the fluorescent probe and detecting fluorescence

emission from a bound query probe, e.g., using a fluorescence microscope) and/or recorded during a defined time interval (e.g., during the acquisition time). The number of times the query probe binds to the nucleic acid during the acquisition time and/or the length of time the query probe remains bound to the nucleic acid during each binding event and the length of time the query probe remains unbound to the nucleic acid between each binding event (e.g., the “dwell times” in the bound and unbound states, respectively) are determined, e.g., by the use of a computer and software (e.g., to analyze the data using a hidden Markov model and Poisson statistics).

In some embodiments, control samples are measured (e.g., in absence of target). Fluorescence detected in a control sample is “background fluorescence” or “background (fluorescence) intensity” or “baseline”.

In some embodiments, data comprising measurements of fluorescence intensity at the emission wavelength of the query probe are recorded as a function of time. In some embodiments, the number of binding events and the dwell times of binding events (e.g. for each immobilized nucleic acid) are determined from the data (e.g., by determining the number of times and the lengths of time the fluorescence intensity is above a threshold background fluorescence intensity). In some embodiments, transitions (e.g., binding and dissociation of a query probe) are counted for each discrete location on the solid support where a target nucleic acid is immobilized. In some embodiments, a threshold number of transitions is used to discriminate the presence of a target nucleic acid at a discrete location on the solid support from background signal, non-target nucleic acid, and/or spurious binding of the query probe. In some embodiments, a number of transitions greater than 10 recorded during the acquisition time indicates the presence of a target nucleic acid at the discrete location on the solid support.

In some embodiments, a distribution of the number of transitions for each immobilized target is determined – e.g., the number of transitions is counted for each immobilized nucleic acid target observed. In some embodiments a histogram is produced. In some embodiments, characteristic parameters of the distribution are determined, e.g., the mean, median, peak, shape, etc. of the distribution are determined. In some embodiments, data and/or parameters (e.g., fluorescence data (e.g., fluorescence data in the time domain), kinetic data, characteristic parameters of the distribution, etc.) are analyzed by algorithms that recognize patterns and regularities in data, e.g., using artificial intelligence, pattern recognition, machine learning, statistical inference, neural nets, etc. In some embodiments, the analysis comprises use of a frequentist analysis and in some embodiments the analysis comprises use of a bayesian analysis. In

some embodiments, pattern recognition systems are trained using known “training” data (e.g., using supervised learning) and in some embodiments algorithms are used to discover previously unknown patterns (e.g., unsupervised learning). See, e.g., Duda, et al. (2001) *Pattern classification* (2nd edition), Wiley, New York; Bishop (2006) *Pattern*
5 *Recognition and Machine Learning*, Springer.

Pattern recognition (e.g., using training sets, supervised learning, unsupervised learning, and analysis of unknown samples) associates identified patterns with nucleic acids such that particular patterns provide a “fingerprint” of particular nucleic acids that find use in detection, quantification, and identification of nucleic acids.

10 In some embodiments, the distribution produced from a target nucleic acid is significantly different than a distribution produced from a non-target nucleic acid or the distribution produced in the absence of a target nucleic acid. In some embodiments, a mean number of transitions is determined for the plurality of immobilized target nucleic acids. In some embodiments, the mean number of transitions observed for a sample
15 comprising a target nucleic acid is approximately linearly related as a function of time and has a positive slope (e.g., the mean number of transitions increases approximately linearly as a function of time).

In some embodiments, the data are treated using statistics (e.g., Poisson statistics) to determine the probability of a transition occurring as a function of time at
20 each discrete location on the solid support. In some particular embodiments, a relatively constant probability of a transition event occurring as a function of time at a discrete location on the solid support indicates the presence of a target nucleic acid at said discrete location on the solid support. In some embodiments, a correlation coefficient relating event number and elapsed time is calculated from the probability of a transition
25 event occurring as a function of time at a discrete location on the solid support. In some embodiments, a correlation coefficient relating event number and elapsed time greater than 0.95 when calculated from the probability of a transition event occurring as a function of time at a discrete location on the solid support indicates the presence of a target nucleic acid at said discrete location on the solid support.

30 In some embodiments, dwell times of bound query probe (τ_{on}) and unbound query probe (τ_{off}) are used to identify the presence of a target nucleic acid in a sample and/or to distinguish a sample comprising a target nucleic acid from a sample comprising a non-target nucleic acid and/or not comprising the target nucleic acid. For example, the τ_{on} for a target nucleic acid is greater than the τ_{on} for a non-target nucleic acid; and, the τ_{off} for
35 a target nucleic acid is smaller than the τ_{off} for a non-target nucleic acid. In some

embodiments, measuring τ_{on} and τ_{off} for a negative control and for a sample indicates the presence or absence of the target nucleic acid in the sample. In some embodiments, a plurality of τ_{on} and τ_{off} values is determined for each of a plurality of spots imaged on a solid support, e.g., for a control (e.g., positive and/or negative control) and a sample
5 suspected of comprising a target nucleic acid. In some embodiments, a mean τ_{on} and/or τ_{off} is determined for each of a plurality of spots imaged on a solid support, e.g., for a control (e.g., positive and/or negative control) and a sample suspected of comprising a target nucleic acid. In some embodiments, a plot of τ_{on} versus τ_{off} (e.g., mean τ_{on} and τ_{off} , time-averaged τ_{on} and τ_{off} , etc.) for all imaged spots indicates the presence or absence of
10 the target nucleic acid in the sample.

Applications

The technology finds use in the detection of nucleic acids, e.g., single-stranded and double-stranded nucleic acids. Accordingly, the technology provides for the detection of
15 various forms of DNA (e.g., DNA comprising modified bases, e.g., methylated DNA, unmethylated DNA) and RNA (e.g., lncRNA, miRNA, etc.), e.g., to provide multi-nucleic acid detection on a common platform. Accordingly, the technology finds use in exemplary applications such as, e.g., detecting one or more of microRNAs, RNAs, mutant DNA alleles, wild-type DNA alleles, and locus-specific methylated DNAs, all on
20 the SiMREPS platform. Detection of these types of nucleic acids occurs, in some embodiments, in the same sample.

For example, the technology finds use in detecting methylated DNA. Methylated DNA is a marker for many states of health and disease, including, for example, identifying patients at higher risk of colorectal cancer based on presence of specific
25 methylated loci. Methylated DNA also provides the basis of a diagnostic test for the early detection of colorectal cancer as well as pre-cancerous adenomas and dysplastic lesions. Detection of methylated DNA at specific loci is currently performed by sodium bisulfite treatment of DNA, which deaminates unmethylated cytosines to produce uracil in the DNA, followed by PCR using distinct primer sets that selectively amplify
30 methylated DNA fragments (e.g., 5-mC or 5-hmC, which are protected from conversion by Na bisulfite) or the unmethylated fragments (where primers are designed to bind to the anticipated converted sequence containing uracils in place of cytosines). Another approach is to perform next generation sequencing of bisulfite treated and/or untreated DNA to infer methylated bases.

Although bisulfite conversion followed by PCR or next generation sequencing are commonly used approaches, they suffer from significant limitations. Surprisingly, some of these limitations of bisulfite based technologies provide advantages for detection of methylated DNA using the SiMREPS technology provided herein. First, the bisulfite treatment not only deaminates cytosines but also randomly fragments the DNA into shorter fragments, creating challenges for designing PCR primers, especially as the sequence length gets very short. In addition, following bisulfite conversion, the regions of DNA that have undergone deamination at cytosines are no longer complementary and therefore not double-stranded, which makes PCR somewhat more challenging as only one strand can be amplified with any given PCR primer set. A SiMREPS approach, on the other hand, fundamentally benefits from the conversion of double-stranded nucleic acids to single-stranded nucleic acids, as well as from fragmentation of nucleic acids to short fragments.

In some embodiments, the presence or absence of nucleic acid modifications are detected (e.g., modified bases, nucleotide analogs, etc.). For example, in some embodiments, epigenetic modifications of nucleic acids that influence gene expression are detected. In some embodiments, methylation of DNA is detected. In some embodiments, the technology finds use in detecting (e.g., identifying the presence or absence of) nucleotide analogs, nucleotide bases – or nucleosides and/or nucleotides comprising bases – other than adenine, thymine, guanosine, cytosine, and uracil. For example, in some embodiments the technology finds use in detecting, identifying, and/or quantifying a nucleotide, nucleoside, and/or a base including but not limited to, e.g., 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC), N(6)-methyladenosine (m(6)A), pseudouridine (Ψ), dihydrouridine (D), inosine (I), 7-methylguanosine (m7G), hypoxanthine, xanthine, 2,6-diaminopurine, and 6,8-diaminopurine.

In an exemplary embodiment, the technology comprises detecting methylated DNA by SiMREPS comprises analyzing samples that are treated with bisulfite and samples that have not been treated with bisulfite. Embodiments provide using SiMREPS query probes that distinguish sequences expected from conversion of unmethylated cytosines to uracil from sequences expected if cytosine(s) at a given locus are not converted to uracil (due to methylation). In some embodiments, both query probes are provided in the sample chamber at the same time and each probe comprises a different fluorophore. In some embodiments, bisulfite reagent is provided in real-time while imaging a SiMREPS experiment with both query probes present (e.g., one probe

that binds to the methylated sequence and a second probe that binds to the uracil-converted sequence, each with a separate fluorophore), where DNA fragments blinking in one color would shift to blinking in the other color based on conversion. Such an assay provides greater accuracy and precision in the measurements than comparing a

5 bisulfite-treated aliquot of sample to an untreated aliquot, which is required for current PCR or next generation sequencing-based approaches. In some embodiments, the technology is a multiplexed analysis using microfluidics and/or multi-spectral imaging, for example. In addition to bisulfate modification, the technology comprises use of other reagents that convert additional chemical markers on DNA or RNA into modifications

10 that are easily detected by SiMREPS.

In some embodiments, the technology finds use in detecting a combination of mutant DNA, methylated DNA, and microRNA biomarkers on the same platform. For example, such a technology finds use in the detection of colorectal cancer and advanced adenoma, e.g., by analyzing a stool sample. The technology provides for the detection of

15 mutant DNA (e.g., detecting 1 mutant molecule in a background of 1,000,000 wild-type molecules; e.g., detecting KRAS mutant DNA in a background of wild-type DNA). Moreover, the technology also provides a technology for measuring all three types of markers (methylated DNA, miRNA, and/or mutant DNA). The technology finds use in analyzing nucleic acids in a buffer solution; the technology finds use in analyzing nucleic

20 acids in a matrix extracted from stool. An average stool weighs 200 g and comprises approximately 10 million diploid-genome equivalents of human DNA. A sensitivity of 1:1,000,000 provides detection of as few as 10 mutant molecules in DNA extracted from a typical whole stool. This is 10,000-fold more sensitive than current clinically-used KRAS assays and 100-fold more sensitive than best-performing research-grade methods.

25 See, e.g., Domagała et al. (2012) "KRAS mutation testing in colorectal cancer as an example of the pathologist's role in personalized targeted therapy: a practical approach" *Pol J Pathol* 63(3): 145-64; Gerecke et al. (2013) "Ultrasensitive detection of unknown colon cancer-initiating mutations using the example of the Adenomatous polyposis coli gene" *Cancer Prev Res (Phila)*. 6(9): 898-907. This dramatic advance in performance is

30 provided by the arbitrary specificity for allele discrimination that is inherent to the kinetic fingerprinting of the SiMREPS technology.

Detecting biomarkers for cancer

Colonoscopy is the dominant screening approach for colorectal cancer (CRC) in the U.S.,

35 despite its invasiveness, high cost, low patient compliance, and risk of complications.

New diagnostics, such as stool-based colorectal cancer screening, are limited by low sensitivity for detecting advanced adenomas (AA), removal of which prevents CRC. The currently best-performing stool-based test analyzes stool DNA (mutant DNA and methylation) and occult blood, but is technically complex, expensive (\$500/test), and
5 challenged by limited ability to detect rare mutant DNA alleles in a high background of wild-type DNA.

The detection technology described herein provides a new technology for low-cost, rapid measurement of rare mutant DNA alleles in stool with exquisite analytic specificity, with concurrent measurement of occult blood (e.g., using a microRNA marker
10 of occult blood) and methylated DNA on a single platform. The technology provides increased sensitivity for detecting advanced adenoma at a >10-fold lower cost than the current state-of-the-art.

The technology provides quantification of rare mutant DNA alleles with orders of magnitude higher specificity than current methods, leading to significantly sensitized
15 AA detection. In some embodiments, adenoma-defining mutations such as in the APC gene are detected.

The technology provides for the measurement of multiple stool biomarker types, all on a single platform. In some embodiments, in addition to mutant DNA, markers detected on the platform include microRNA (see, e.g., U.S. Pat. App. Ser. No. 14/589,467,
20 incorporated herein by reference in its entirety), a stool occult blood marker (e.g., a microRNA marker of stool occult blood), and methylated DNA.

Detection of RNA

In some embodiments, the nucleic acid to be detected, characterized, quantified, and/or
25 identified (e.g., the target nucleic acid) is a RNA (e.g., a lncRNA, e.g., a non-protein coding RNA longer than approximately 200 nucleotides). Embodiments provide that the secondary structure of a nucleic acid (e.g., a RNA, e.g., a lncRNA) is melted to provide access to query regions by query probes for SiMREPS detection. For example, in some embodiments the technology comprises use of a dCas9/gRNA that recognizes and melts
30 a secondary structured target in an RNA using an auxiliary PAM oligonucleotide (e.g., a PAMmer). In some embodiments the technology comprises use of two dCas9/gRNA complexes that recognize and melt a secondary structured target in an RNA using an auxiliary PAM oligonucleotide (e.g., a PAMmer).

In some embodiments, the dCas9 binds to single-stranded RNA targets matching
35 the gRNA sequence when the PAM is presented in trans as a separate DNA

oligonucleotide (a “PAM-presenting oligonucleotide” or “PAMmer”). Accordingly, in some embodiments PAMmers provide for the site-specific binding of dCas9/gRNA to single-stranded RNA targets (e.g., lncRNA). Furthermore, the technology provides for the use of PAMmers to direct dCas9 to bind to specific RNA targets and melt secondary

5 structure to make query regions available for the binding of query probes in a SiMREPs assay. See, e.g., O’Connell et al. 2014 “Programmable RNA recognition and cleavage by CRISPR/Cas9” *Nature* 516: 263-6. Thus, embodiments provide compositions comprising a dCas9, a gRNA (e.g., a dCas9/gRNA complex), and a PAMmer.

10 **Detection of microRNA**

In some embodiments, the nucleic acid to be detected, characterized, quantified, and/or identified (e.g., the target nucleic acid) is a microRNA. microRNAs (miRNA or μ RNA) are single-stranded RNA molecules of approximately 21 to 23 nucleotides in length that regulate gene expression. miRNAs are encoded by genes from whose DNA they are

15 transcribed, but miRNAs are not translated into protein (see, e.g., Carrington et al, 2003, which is hereby incorporated by reference). The genes encoding miRNAs are much longer than the processed mature miRNA molecule. See, e.g., U.S. Pat. App. Ser. No. 14/589,467, incorporated herein by reference in its entirety.

20 **Detection of genomic aberrations**

In some embodiments, the SiMREPS technology provides for identifying and/or counting genomic aberrations (e.g., other than simple point mutations) in a DNA sample based on detecting unpaired regions after comparative genome hybridization. In some

25 embodiments, SiMREPS finds use to determine the overall degree of genomic instability in a sample, e.g., as evidenced by presence of deletions and insertions compared to a reference, normal DNA sample. Exemplary embodiments comprise providing a normal DNA (for example, wild-type DNA from normal blood cells of a patient with a solid tumor like lung cancer) and mix it with matched tumor DNA (or even circulating cell-free DNA), with fragmentation of the DNA so that it is present as fragments and is

30 immobilized onto a surface, potentially through end modification (e.g., biotinylation) or other approaches. Most of the DNA will form hybridized DNA segments, but areas of deletion or insertion are present as duplexes where one strand bulges out. In some embodiments, the normal DNA and the tumor DNA are provided in ratios appropriate for efficient detection. Embodiments provide a SiMREPS-based affinity reagent that is

35 used to detect these unmatched regions, and counting them provides a measure of

genomic instability, which finds use, e.g., in some embodiments as a biomarker of cancer risk. In some embodiments, the technology finds use in pre-natal screening for chromosomal abnormalities. In some embodiments, the affinity reagent is a single-stranded DNA binding proteins, a Holliday junction recombinases modified not to cleave
5 nucleic acid, e.g., for identification of balanced chromosomal translocations, etc.

In some embodiments, the technology finds use in the detection of microsatellite repeat aberrations in patients with microsatellite-unstable colorectal cancer, for example. Accordingly, embodiments comprise providing a panel of SiMREPS probes corresponding to different microsatellite loci that show differential kinetic binding
10 properties depending on whether the microsatellite repeats have expanded or not in a sample. In some embodiments, the technology comprises a comparative DNA hybridization approach in which hybridization of an expanded microsatellite repeat sequence to one without expansion generates an unpaired segment that is detected using a SiMREPS-based approach, using a DNA probe or other effective affinity
15 SiMREPS reader reagent.

Bifunctional affinity reagents

In some embodiments, the technology comprises detection of proteins and other analytes with SiMREPS by incorporation of a bifunctional affinity reagent that binds to the
20 target analyte and comprises a nucleic acid that can be counted using a SiMREPS reader probe. As an example, some embodiments comprise the use of an antibody linked to a short DNA oligonucleotide, such that if the target protein analyte were immobilized onto the surface of a slide (for example by simple drying onto the surface, or other nonspecific or specific capture methods), the binding of the antibody to the target
25 protein analyte is measured using a SiMREPS-based reading of the conjugated DNA. This would allow multiple antibodies to be multiplexed and distinguished by different DNA sequence “barcodes” linked to the various antibodies. Embodiments provide that the samples are a large variety of types including even cells or cell lysates. Affinity reagents include DNA or RNA binding proteins, aptamers, antibodies and antibody
30 fragments, linked to a DNA barcode.

Non-nucleic acid SiMREPS probes

Embodiments provide probes that are not nucleic acids. For example, embodiments provide an antibody or other affinity reagent that have a binding interaction with a
35 target analyte that has a stability amenable for SiMREPS, e.g., a transient associated

that provides a “blinking” signal and, in some embodiments, a kinetic binding fingerprint. In some embodiments, the non-nucleic acid query probe is engineered to weaken its binding relative to the non-engineered version, e.g., to provide a binding and/or association that is less thermodynamically stable. In addition to antibodies, 5 embodiments comprise target analytes and query probes that are proteins, e.g., where one binding partner is the affinity reagent and the other would be the target analyte being measured. Embodiments comprise the use of aptamers binding any ligand, lectins binding glycosylated proteins, proteins or other molecules binding lipids, etc. The technology comprises the use of any binding pair with transient binding behavior 10 suitable for detection on the SiMREPS platform, e.g., that produces a kinetic fingerprint.

Intramolecular SiMREPS probing

In some embodiments, the technology provides a capture probe and a query probe that 15 are linked to provide an intramolecular probing mechanism (see, e.g., Fig. 5 a and b). The probe is asymmetric, so that when the target nucleic acid binds (e.g., with thermodynamic stability, e.g., irreversibly) to the capture sequence, the target nucleic acid undergoes transient binding with the query probe. The transient binding and dissociation of the query probe yields a time-dependent change in donor fluorophore 20 intensity or FRET whose kinetics are sensitive to the sequence of the target nucleic acid. In some embodiments, an address strand binds the Query/Capture complex to the surface, to provide rigidity that exerts control over the transient binding kinetics, and to provide a means to immobilize many different Query/Capture sequences to different regions of the imaging surface (e.g., as in a DNA microarray).

25 In some embodiments, intramolecular SiMREPS probing provides faster acquisition. In particular, binding of the query probe is rapid because the binding is an intramolecular hybridization reaction after the target nucleic acid binds to the capture probe. Furthermore, in some embodiments, imaging times are reduced compared to the other embodiments of the SiMREPS technology. For instance, the same number of 30 binding and dissociation events occur in 1 to 10 seconds in some embodiments of intramolecular SiMREPS experiments as occur in 10 minutes in other embodiments of the SiMREPS technology. The intramolecular SiMREPS technology provides for the parallelization of experiments through spatial segregation. In some embodiments, the intramolecular SiMREPS technology reduces the concentrations of query probe that 35 provide efficient detection. The intramolecular SiMREPS technology provides a platform

that, in some embodiments, comprises many different Capture/Query probes immobilized within different regions of the imaging surface in a manner specified by the Address strand. As an example, one might use a standard microarray chip containing thousands of distinct sequences; these sequences could serve as the Address strands for
5 immobilization of SiMREPS Capture/Query probes, permitting SiMREPS assays of thousands of target sequences (microRNAs, lncRNAs, DNA converted to single-stranded form) on a single chip.

Notably, in some embodiments, the Address strands are not related in sequence to any of the targets, since interaction occurs indirectly through the query and capture
10 probes. Indeed, the Address strands are not required to be related to the targets. Furthermore, embodiments provide that the query and capture probes comprise affinity reagents other than DNA sequences, such as the dCas9/gRNA complexes discussed elsewhere in this application.

Embodiments provide control of exposure of the fluorophores to excitation
15 sources, e.g., to reduce or minimize photobleaching prior to analysis. In some embodiments, kinetic signatures provide a correction mechanism to identify and correct false positive detections resulting from, e.g., deposit of a Capture/Query probe on the wrong part of the imaging surface (outside of its Address region). Embodiments also provide a technology in which false positives are minimized or reduced by splitting the
20 Query and Capture probes into two non-contiguous probes that co-localize upon binding to the Address sequence (see, e.g., Fig. 5 b).

Fluorescent moieties

In some embodiments, a nucleic acid comprises a fluorescent moiety (e.g., a fluorogenic
25 dye, also referred to as a “fluorophore” or a “fluor”). A wide variety of fluorescent moieties is known in the art and methods are known for linking a fluorescent moiety to a nucleotide prior to incorporation of the nucleotide into an oligonucleotide and for adding a fluorescent moiety to an oligonucleotide after synthesis of the oligonucleotide.

Examples of compounds that may be used as the fluorescent moiety include but
30 are not limited to xanthene, anthracene, cyanine, porphyrin, and coumarin dyes. Examples of xanthene dyes that find use with the present technology include but are not limited to fluorescein, 6-carboxyfluorescein (6-FAM), 5-carboxyfluorescein (5-FAM), 5- or 6-carboxy-4, 7, 2', 7'-tetrachlorofluorescein (TET), 5- or 6-carboxy-4'5'2'4'5'7' hexachlorofluorescein (HEX), 5' or 6'-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein
35 (JOE), 5-carboxy-2',4',5',7'-tetrachlorofluorescein (ZOE), rhodol, rhodamine,

tetramethylrhodamine (TAMRA), 4,7-dichlorotetramethyl rhodamine (DTAMRA), rhodamine X (ROX), and Texas Red. Examples of cyanine dyes that may find use with the present invention include but are not limited to Cy 3, Cy 3B, Cy 3.5, Cy 5, Cy 5.5, Cy 7, and Cy 7.5. Other fluorescent moieties and/or dyes that find use with the present
 5 technology include but are not limited to energy transfer dyes, composite dyes, and other aromatic compounds that give fluorescent signals. In some embodiments, the fluorescent moiety comprises a quantum dot.

In some embodiments, the fluorescent moiety comprises a fluorescent protein (e.g., a green fluorescent protein (GFP), a modified derivative of GFP (e.g., a GFP
 10 comprising S65T, an enhanced GFP (e.g., comprising F64L)), or others known in the art such as, e.g., blue fluorescent protein (e.g., EBFP, EBFP2, Azurite, mKalama1), cyan fluorescent protein (e.g., ECFP, Cerulean, CyPet, mTurquoise2), and yellow fluorescent protein derivatives (e.g., YFP, Citrine, Venus, YPet). Embodiments provide that the fluorescent protein may be covalently or noncovalently bonded to one or more query
 15 and/or capture probes.

Fluorescent dyes include, without limitation, d-Rhodamine acceptor dyes including Cy 5, dichloro[R110], dichloro[R6G], dichloro[TAMRA], dichloro[ROX] or the like, fluorescein donor dyes including fluorescein, 6-FAM, 5-FAM, or the like; Acridine including Acridine orange, Acridine yellow, Proflavin, pH 7, or the like; Aromatic
 20 Hydrocarbons including 2-Methylbenzoxazole, Ethyl p-dimethylaminobenzoate, Phenol, Pyrrole, benzene, toluene, or the like; Arylmethine Dyes including Auramine O, Crystal violet, Crystal violet, glycerol, Malachite Green or the like; Coumarin dyes including 7-Methoxycoumarin-4-acetic acid, Coumarin 1, Coumarin 30, Coumarin 314, Coumarin 343, Coumarin 6 or the like; Cyanine Dyes including 1,1'-diethyl-2,2'-cyanine iodide,
 25 Cryptocyanine, Indocarbocyanine (C3) dye, Indodicarbocyanine (C5) dye, Indotricarbocyanine (C7) dye, Oxacarbocyanine (C3) dye, Oxadicarbocyanine (C5) dye, Oxatricarbocyanine (C7) dye, Pinacyanol iodide, Stains all, Thiocarbocyanine (C3) dye, ethanol, Thiocarbocyanine (C3) dye, n-propanol, Thiadicarbocyanine (C5) dye, Thiatricarbocyanine (C7) dye, or the like; Dipyrin dyes including N,N'-Difluoroboryl-
 30 1,9-dimethyl-5-(4-iodophenyl)-dipyrin, N,N'-Difluoroboryl-1,9-dimethyl-5-[(4-(2-trimethylsilyl)ethynyl)], N,N'-Difluoroboryl-1,9-dimethyl-5-phenyldipyrin, or the like; Merocyanines including 4-(dicyanomethylene)-2-methyl-6-(p-dimethylaminostyryl)-4H-pyran (DCM), acetonitrile, 4-(dicyanomethylene)-2-methyl-6-(p-dimethylaminostyryl)-4H-pyran (DCM), methanol, 4-Dimethylamino-4'-nitrostillbene, Merocyanine 540, or the
 35 like; Miscellaneous Dyes including 4',6'-Diamidino-2-phenylindole (DAPI),

dimethylsulfoxide, 7-Benzylamino-4-nitrobenz-2-oxa-1,3-diazole, Dansyl glycine, Dansyl glycine, dioxane, Hoechst 33258, DMF, Hoechst 33258, Lucifer yellow CH, Piroxicam, Quinine sulfate, Quinine sulfate, Squarylium dye III, or the like; Oligophenylenes including 2,5-Diphenyloxazole (PPO), Biphenyl, POPOP, p-Quaterphenyl, p-Terphenyl, or the like; Oxazines including Cresyl violet perchlorate, Nile Blue, methanol, Nile Red, ethanol, Oxazine 1, Oxazine 170, or the like; Polycyclic Aromatic Hydrocarbons including 9,10-Bis(phenylethynyl)anthracene, 9,10-Diphenylanthracene, Anthracene, Naphthalene, Perylene, Pyrene, or the like; polyene/polyyne including 1,2-diphenylacetylene, 1,4-diphenylbutadiene, 1,4-diphenylbutadiyne, 1,6-diphenylhexatriene, Beta-carotene, Stilbene, or the like; Redox-active Chromophores including Anthraquinone, Azobenzene, Benzoquinone, Ferrocene, Riboflavin, Tris(2,2'-bipyridyl)pruthenium(II), Tetrapyrrole, Bilirubin, Chlorophyll a, diethyl ether, Chlorophyll a, methanol, Chlorophyll b, Diprotonated-tetraphenylporphyrin, Hematin, Magnesium octaethylporphyrin, Magnesium octaethylporphyrin (MgOEP), Magnesium phthalocyanine (MgPc), PrOH, Magnesium phthalocyanine (MgPc), pyridine, Magnesium tetramesitylporphyrin (MgTMP), Magnesium tetraphenylporphyrin (MgTPP), Octaethylporphyrin, Phthalocyanine (Pc), Porphin, ROX, TAMRA, Tetra-t-butylazaporphine, Tetra-t-butylphthalocyanine, Tetrakis(2,6-dichlorophenyl)porphyrin, Tetrakis(o-aminophenyl)porphyrin, Tetramesitylporphyrin (TMP), Tetraphenylporphyrin (TPP), Vitamin B12, Zinc octaethylporphyrin (ZnOEP), Zinc phthalocyanine (ZnPc), pyridine, Zinc tetramesitylporphyrin (ZnTMP), Zinc tetramesitylporphyrin radical cation, Zinc tetraphenylporphyrin (ZnTPP), or the like; Xanthenes including Eosin Y, Fluorescein, basic ethanol, Fluorescein, ethanol, Rhodamine 123, Rhodamine 6G, Rhodamine B, Rose bengal, Sulforhodamine 101, or the like; or mixtures or combination thereof or synthetic derivatives thereof.

Several classes of fluorogenic dyes and specific compounds are known that are appropriate for particular embodiments of the technology: xanthene derivatives such as fluorescein, rhodamine, Oregon green, eosin, and Texas red; cyanine derivatives such as cyanine, indocarbocyanine, oxacarbocyanine, thiocarbocyanine, and merocyanine; naphthalene derivatives (dansyl and prodan derivatives); coumarin derivatives; oxadiazole derivatives such as pyridyloxazole, nitrobenzoxadiazole, and benzoxadiazole; pyrene derivatives such as cascade blue; oxazine derivatives such as Nile red, Nile blue, cresyl violet, and oxazine 170; acridine derivatives such as proflavin, acridine orange, and acridine yellow; arylmethine derivatives such as auramine, crystal violet, and malachite green; and tetrapyrrole derivatives such as porphin, phtalocyanine, bilirubin.

In some embodiments the fluorescent moiety a dye that is xanthene, fluorescein, rhodamine, BODIPY, cyanine, coumarin, pyrene, phthalocyanine, phycobiliprotein, ALEXA FLUOR® 350, ALEXA FLUOR® 405, ALEXA FLUOR® 430, ALEXA FLUOR® 488, ALEXA FLUOR® 514, ALEXA FLUOR® 532, ALEXA FLUOR® 546, ALEXA
5 FLUOR® 555, ALEXA FLUOR® 568, ALEXA FLUOR® 568, ALEXA FLUOR® 594, ALEXA FLUOR® 610, ALEXA FLUOR® 633, ALEXA FLUOR® 647, ALEXA FLUOR® 660, ALEXA FLUOR® 680, ALEXA FLUOR® 700, ALEXA FLUOR® 750, or a squaraine dye. In some embodiments, the label is a fluorescently detectable moiety as described in, e.g., Haugland (September 2005) MOLECULAR PROBES HANDBOOK OF
10 FLUORESCENT PROBES AND RESEARCH CHEMICALS (10th ed.), which is herein incorporated by reference in its entirety.

In some embodiments the label (e.g., a fluorescently detectable label) is one available from ATTO-TEC GmbH (Am Eichenhang 50, 57076 Siegen, Germany), e.g., as described in U.S. Pat. Appl. Pub. Nos. 20110223677, 20110190486, 20110172420,
15 20060179585, and 20030003486; and in U.S. Pat. No. 7,935,822, all of which are incorporated herein by reference (e.g., ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633,
20 ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO740).

One of ordinary skill in the art will recognize that dyes having emission maxima outside these ranges may be used as well. In some cases, dyes ranging between 500 nm to 700 nm have the advantage of being in the visible spectrum and can be detected using
25 existing photomultiplier tubes. In some embodiments, the broad range of available dyes allows selection of dye sets that have emission wavelengths that are spread across the detection range. Detection systems capable of distinguishing many dyes are known in the art.

30 Samples

In some embodiments, nucleic acids (e.g., DNA or RNA) 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 material (e.g., cellular material (live or dead), extracellular material, viral material,
35 environmental samples (e.g., metagenomic samples), synthetic material (e.g., amplicons

such as provided by PCR or other amplification technologies)), obtained from an animal, plant, bacterium, archaeon, fungus, or any other organism. Biological samples for use in the present technology include viral particles or preparations thereof. Nucleic acid molecules can be obtained directly from an organism or from a biological sample
5 obtained from an organism, e.g., from blood, urine, cerebrospinal fluid, seminal fluid, saliva, sputum, stool, hair, sweat, tears, skin, and tissue. Exemplary samples include, but are not limited to, whole blood, lymphatic fluid, serum, plasma, buccal cells, sweat, tears, saliva, sputum, hair, skin, biopsy, cerebrospinal fluid (CSF), amniotic fluid, seminal fluid, vaginal excretions, serous fluid, synovial fluid, pericardial fluid,
10 peritoneal fluid, pleural fluid, transudates, exudates, cystic fluid, bile, urine, gastric fluids, intestinal fluids, fecal samples, and swabs, aspirates (e.g., bone marrow, fine needle, etc.), washes (e.g., oral, nasopharyngeal, bronchial, bronchialalveolar, optic, rectal, intestinal, vaginal, epidermal, etc.), and/or other specimens.

Any tissue or body fluid specimen may be used as a source for nucleic acid for use
15 in the technology, including forensic specimens, archived specimens, preserved specimens, and/or specimens stored for long periods of time, e.g., fresh-frozen, methanol/acetic acid fixed, or formalin-fixed paraffin embedded (FFPE) specimens and samples. Nucleic acid template 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
20 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. A sample may also be isolated DNA from a non-cellular origin, e.g. amplified/isolated DNA that has been stored in a freezer.

Nucleic acid molecules can be obtained, e.g., by extraction from a biological
25 sample, e.g., by a variety of techniques such as those described by Maniatis, et al. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y. (see, e.g., pp. 280–281).

In some embodiments, the technology provides for the size selection of nucleic acids, e.g., to remove very short fragments or very long fragments.

30 In some embodiments, the technology is used to identify a nucleic acid in situ. In particular, embodiments of the technology provide for the identification of a nucleic acid directly in a tissue, cell, etc. (e.g., after permeabilizing the tissue, cell, etc.) without extracting the nucleic acid from the tissue, cell, etc. In some embodiments of the technology related to in situ detection, the technology is applied in vivo, ex vivo, and/or
35 in vitro. In some embodiments, the sample is a crude sample, a minimally treated cell

lysates, or a biofluid lysate. In some embodiments, the nucleic acid is detected in a crude lysates without nucleic acid purification.

Kits

5 Some embodiments are related to kits for the detection of a nucleic acid. For instance, in some embodiments are provided a kit comprising a solid support (e.g., a microscope slide, a bead, a coverslip, an avidin (e.g., streptavidin)-conjugated microscope slide or coverslip, a solid support comprising a zero mode waveguide array, or the like), a dCas9/gRNA (e.g., comprising a biotinylated dCas9), and a query probe as described
10 herein. Some embodiments further provide a non-biotinylated dCas9/gRNA.

Some embodiments further provide software on a computer-readable format or downloadable from the internet for the collection and analysis of query probe binding events and dwell times as described herein. In some embodiments, kits for multiplex detection comprise two or more query probes each comprising a sequence
15 complementary to distinct query regions of one or more target nucleic acids and each comprising a different fluorescent moiety. In some embodiments, query probes are complementary to query regions of one or more nucleic acid targets. Some embodiments of kits comprise one or more positive controls and/or one or more negative controls. Some embodiments comprise a series of controls having known concentrations, e.g., to produce
20 a standard curve of concentrations.

Systems

Some embodiments of the technology provide systems for the detection and quantification of a target nucleic acid. Systems according to the technology comprise,
25 e.g., a solid support (e.g., a microscope slide, a coverslip, an avidin (e.g., streptavidin)-conjugated microscope slide or coverslip, a solid support comprising a zero mode waveguide array, or the like), a dCas9/gRNA (e.g., comprising a biotinylated dCas9), and a query probe as described herein. Some embodiments further provide a non-biotinylated dCas9/gRNA.

30 Some embodiments further comprise a fluorescence microscope comprising an illumination configuration to excite bound query probes (e.g., a prism-type total internal reflection fluorescence (TIRF) microscope, an objective-type TIRF microscope, a near-TIRF or HiLo microscope, a confocal laser scanning microscope, a zero-mode waveguide, and/or an illumination configuration capable of parallel monitoring of a large area of the
35 slide or coverslip ($> 100 \mu\text{m}^2$) while restricting illumination to a small region of space

near the surface). Some embodiments comprise a fluorescence detector, e.g., a detector comprising an intensified charge coupled device (ICCD), an electron-multiplying charge coupled device (EM-CCD), a complementary metal-oxide-semiconductor (CMOS), a photomultiplier tube (PMT), an avalanche photodiode (APD), and/or another detector
5 capable of detecting fluorescence emission from single chromophores. Some embodiments comprise a computer and software encoding instructions for the computer to perform.

Some embodiments comprise optics, such as lenses, mirrors, dichroic mirrors, optical filters, etc., e.g., to detect fluorescence selectively within a specific range of
10 wavelengths or multiple ranges of wavelengths.

For example, in some embodiments, computer-based analysis software is used to translate the raw data generated by the detection assay (e.g., the presence, absence, or amount of one or more nucleic acids (e.g., one or more biomarkers) into data of predictive value for a clinician. The clinician can access the predictive data using any
15 suitable means.

For instance, some embodiments comprise a computer system upon which embodiments of the present technology may be implemented. In various embodiments, a computer system includes a bus or other communication mechanism for communicating information and a processor coupled with the bus for processing information. In various
20 embodiments, the computer system includes a memory, which can be a random access memory (RAM) or other dynamic storage device, coupled to the bus, and instructions to be executed by the processor. Memory also can be used for storing temporary variables or other intermediate information during execution of instructions to be executed by the processor. In various embodiments, the computer system can further include a read only
25 memory (ROM) or other static storage device coupled to the bus for storing static information and instructions for the processor. A storage device, such as a magnetic disk or optical disk, can be provided and coupled to the bus for storing information and instructions.

In various embodiments, the computer system is coupled via the bus to a display,
30 such as a cathode ray tube (CRT) or a liquid crystal display (LCD), for displaying information to a computer user. An input device, including alphanumeric and other keys, can be coupled to the bus for communicating information and command selections to the processor. Another type of user input device is a cursor control, such as a mouse, a trackball, or cursor direction keys for communicating direction information and
35 command selections to the processor and for controlling cursor movement on the display.

This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allows the device to specify positions in a plane.

A computer system can perform embodiments of the present technology.

Consistent with certain implementations of the present technology, results can be
5 provided by the computer system in response to the processor executing one or more
sequences of one or more instructions contained in the memory. Such instructions can be
read into the memory from another computer-readable medium, such as a storage
device. Execution of the sequences of instructions contained in the memory can cause
the processor to perform the methods described herein. Alternatively, hard-wired
10 circuitry can be used in place of or in combination with software instructions to
implement the present teachings. Thus, implementations of the present technology are
not limited to any specific combination of hardware circuitry and software.

The term “computer-readable medium” as used herein refers to any media that
participates in providing instructions to the processor for execution. Such a medium can
15 take many forms, including but not limited to, non-volatile media, volatile media, and
transmission media. Examples of non-volatile media can include, but are not limited to,
optical or magnetic disks, such as a storage device. Examples of volatile media can
include, but are not limited to, dynamic memory. Examples of transmission media can
include, but are not limited to, coaxial cables, copper wire, and fiber optics, including the
20 wires that comprise the bus.

Common forms of computer-readable media include, for example, a floppy disk, a
flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any
other optical medium, punch cards, paper tape, any other physical medium with
patterns of holes, a RAM, PROM, and EPROM, a FLASH-EPROM, any other memory
25 chip or cartridge, or any other tangible medium from which a computer can read.

Various forms of computer readable media can be involved in carrying one or
more sequences of one or more instructions to the processor for execution. For example,
the instructions can initially be carried on the magnetic disk of a remote computer. The
remote computer can load the instructions into its dynamic memory and send the
30 instructions over a network connection (e.g., a LAN, a WAN, the internet, a telephone
line). A local computer system can receive the data and transmit it to the bus. The bus
can carry the data to the memory, from which the processor retrieves and executes the
instructions. The instructions received by the memory may optionally be stored on a
storage device either before or after execution by the processor.

In accordance with various embodiments, instructions configured to be executed by a processor to perform a method are stored on a computer-readable medium. The computer-readable medium can be a device that stores digital information. For example, a computer-readable medium includes a compact disc read-only memory (CD-ROM) as is
5 known in the art for storing software. The computer-readable medium is accessed by a processor suitable for executing instructions configured to be executed.

In accordance with such a computer system, some embodiments of the technology provided herein further comprise functionalities for collecting, storing, and/or analyzing data (e.g., presence, absence, concentration of a nucleic acid). For example, some
10 embodiments contemplate a system that comprises a processor, a memory, and/or a database for, e.g., storing and executing instructions, analyzing fluorescence, image data, performing calculations using the data, transforming the data, and storing the data. In some embodiments, an algorithm applies a statistical model (e.g., a Poisson model or hidden Markov model) to the data.

Many diagnostics involve determining the presence of, or a nucleotide sequence of, one or more nucleic acids (e.g., a nucleic acid biomarker). Thus, in some
15 embodiments, an equation comprising variables representing the presence, absence, concentration, amount, or sequence properties of multiple nucleic acids produces a value that finds use in making a diagnosis or assessing the presence or qualities of a nucleic
20 acid. As such, in some embodiments this value is presented by a device, e.g., by an indicator related to the result (e.g., an LED, an icon on a display, a sound, or the like). In some embodiments, a device stores the value, transmits the value, or uses the value for additional calculations. In some embodiments, an equation comprises variables representing the presence, absence, concentration, amount, or sequence properties of
25 one or more of a methylated locus in genomic DNA, a microRNA, a mutant gene biomarker, or a chromosomal aberration.

Thus, in some embodiments, the present technology provides the further benefit that a clinician, who is not likely to be trained in genetics or molecular biology, need not understand the raw data. The data are presented directly to the clinician in its most
30 useful form. The clinician is then able to utilize the information to optimize the care of a subject. The present invention contemplates any method capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information providers, medical personal, and/or subjects. For example, in some
35 embodiments of the present technology, a sample is obtained from a subject and submitted to a profiling service (e.g., a clinical lab at a medical facility, genomic profiling

business, etc.), located in any part of the world (e.g., in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center or subjects may collect the sample themselves and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (e.g., an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced that is specific for the diagnostic or prognostic information desired for the subject. The profile data are then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw expression data, the prepared format may represent a diagnosis or risk assessment for the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (e.g., at the point of care) or displayed to the clinician on a computer monitor. In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data are then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data are stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers. In some embodiments, the subject is able to access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data are used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition associated with the disease.

All publications and patents mentioned in the above specification are herein incorporated by reference in their entirety for all purposes. Various modifications and variations of the described compositions, methods, and uses of the technology will be apparent to those skilled in the art without departing from the scope and spirit of the

technology as described. Although the technology has been described in connection with specific exemplary embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to
5 those skilled in the art are intended to be within the scope of the following claims.

CLAIMS

WE CLAIM:

1. A complex for providing a detectable fingerprint of a double-stranded target
5 nucleic acid, the complex comprising:
 - a) a double-stranded target nucleic acid comprising a first region adjacent to a second region;
 - b) an immobilized melting component interacting with the first region to form a thermodynamically stable complex and provide the second region
10 in a single-stranded form; and
 - c) a query probe that binds repeatedly to the second region to provide a detectable fingerprint associated with the double-stranded target nucleic acid.
- 15 2. The complex of claim 1 wherein the melting component comprises a dCas9.
3. The complex of claim 1 wherein the melting component comprises a dCas9/gRNA complex comprising a gRNA hybridized to the first region.
- 20 4. The complex of claim 1 wherein the melting component comprises a PAMmer.
5. The complex according to claim 1 wherein the query probe is a fluorescently labeled nucleic acid that hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on}
25 that is greater than 0.1 min^{-1} .
6. The complex according to claim 2 wherein the dCas9 is immobilized to a substrate.
- 30 7. The complex according to claim 1 wherein the fingerprint is detectable by a pattern recognition analysis.
8. The complex according to claim 1 further comprising a second melting component interacting with a third region of the target nucleic acid adjacent to the second
35 region of the target nucleic acid.

9. The complex according to claim 1 wherein the target nucleic acid comprises a mutation, single nucleotide polymorphism, or a modified base.
- 5 10. The complex of claim 8 wherein the first and second melting components bind approximately 5 to 15 nucleotides apart on the target nucleic acid.
11. A method for providing a detectable fingerprint of a double-stranded target nucleic acid in a sample, the method comprising:
- 10 a) immobilizing a double-stranded target nucleic acid to a discrete region of a solid support, said double-stranded target nucleic acid comprising a first region adjacent to a second region and said discrete region of said solid support comprising an immobilized melting component interacting with the first region;
- 15 b) providing a query probe that binds repeatedly to the second region to provide a detectable fingerprint; and
- c) associating the detectable fingerprint with the double-stranded nucleic acid to identify the double-stranded nucleic acid.
- 20 12. The method of claim 11 comprising analyzing data using pattern recognition to produce or identify the detectable fingerprint of the double stranded nucleic acid.
13. The method of claim 11 wherein the melting component comprises a dCas9.
- 25 14. The method of claim 11 further comprising providing a second dCas9/gRNA complex comprising a gRNA complementary to a third region of the target nucleic acid adjacent to the second region of the target nucleic acid.
15. The method according to claim 11 comprising providing conditions sufficient for
- 30 the melting component to provide the second region in a single-stranded form.
16. The method of claim 11 comprising detecting repeated binding of the detectably labeled query probe to the second region of the target nucleic acid, wherein the kinetic rate constant k_{off} is greater than 1 min^{-1} and/or the kinetic rate constant k_{on} is greater than 1 min^{-1} .
- 35

17. The method of claim 16 further comprising calculating an amount or concentration of the double-stranded target nucleic acid in the sample from the detectable fingerprint.
- 5
18. A system for the detection of a double-stranded nucleic acid, the system comprising a solid support comprising an immobilized melting component, a detectably labeled query probe that binds repeatedly to the double-stranded nucleic acid, a fluorescence detector, and a software component configured to perform pattern recognition analysis of query probe binding data.
- 10
19. A method for calculating a predictor that a subject has or is at risk of having a cancer, the method comprising one or more of determining the presence of a microRNA biomarker, determining the presence of a mutation, and/or determining the presence of a modified base in genomic DNA.
- 15
20. The method of claim 19 wherein the predictor is a value calculated from variables associated with one or more of the presence of a microRNA biomarker, the presence of a mutation, and/or the presence of a modified base in genomic DNA.
- 20
21. A complex for detecting a target nucleic acid, the complex comprising:
- 1) a target nucleic acid comprising a first region adjacent to a second region;
 - 2) a detectably labeled capture probe hybridized to the first region;
 - 3) a query probe labeled with a quencher or fluorescent acceptor compatible with the label of the capture probe,
- 25
- wherein the query probe hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 0.1 min^{-1} .
- 30
22. A system for providing a detectable fingerprint of a double-stranded target nucleic acid, the system comprising:
- a) an immobilized melting component capable of interacting with a first region of the double-stranded target nucleic acid to form a thermodynamically stable complex and provide a second region of the double-stranded target nucleic acid in a single-stranded form; and
- 35

- b) a query probe that binds repeatedly to the second region to provide a detectable fingerprint associated with the double-stranded target nucleic acid.
- 5 23. The system of claim 22 wherein the melting component comprises a dCas9.
24. The system of claim 22 wherein the melting component comprises a dCas9/gRNA complex comprising a gRNA complementary to the first region.
- 10 25. The system of claim 22 wherein the melting component comprises a PAMmer.
26. The system according to claim 22 wherein the query probe is a fluorescently labeled nucleic acid that hybridizes repeatedly to the second region with a kinetic rate constant k_{off} that is greater than 0.1 min^{-1} and/or a kinetic rate constant k_{on} that is greater than 0.1 min^{-1} .
- 15 27. The system according to claim 23 wherein the dCas9 is immobilized to a substrate.
- 20 28. The system according to claim 22 wherein the fingerprint is detectable by a pattern recognition analysis.
29. The system according to claim 22 further comprising a second melting component interacting with a third region of the target nucleic acid adjacent to the second region of the target nucleic acid.
- 25 30. The system according to claim 22 wherein the target nucleic acid comprises a mutation, single nucleotide polymorphism, or a modified base.
- 30 31. The system of claim 29 wherein the first and second melting components bind approximately 5 to 15 nucleotides apart on the target nucleic acid.
32. The system of claim 22 further comprising a fluorescence detector.

33. The system of claim 22 further comprising a software component configured to perform pattern recognition analysis of query probe binding data.
34. The system of claim 22 further comprising a software component configured to calculate an amount or concentration of the double-stranded target nucleic acid in the sample from the detectable fingerprint.
35. The system of claim 22 further comprising an indicator or display for providing a result related to the presence, absence, concentration, amount, or sequence properties of the target nucleic acid.
36. A kit for providing a result related to the presence, absence, concentration, amount, or sequence properties of a target nucleic acid, the kit comprising a solid support, a dCas9/gRNA, and a query probe.
37. The kit of claim 36 further comprising software on a computer-readable format for the collection and analysis of query probe binding events and/or dwell times.
38. The kit of claim 36 further comprising one or more positive controls and/or one or more negative controls.
39. The kit of claim 36 wherein the solid support comprises a microscope slide, a bead, or a coverslip
40. The kit of claim 36 wherein the solid support comprises avidin or streptavidin
41. The kit of claim 36 wherein the solid support comprises a zero mode waveguide array.
42. Use of a system for providing a detectable fingerprint of a double-stranded target nucleic acid, the system comprising:
- a) an immobilized melting component capable of interacting with a first region of the double-stranded target nucleic acid to form a thermodynamically stable complex and provide a second region of the double-stranded target nucleic acid in a single-stranded form; and

- b) a query probe that binds repeatedly to the second region to provide a detectable fingerprint associated with the double-stranded target nucleic acid.
- 5 43. Use of a method to provide a detectable fingerprint of a double-stranded target nucleic acid in a sample, the method comprising:
- a) immobilizing a double-stranded target nucleic acid to a discrete region of a solid support, said double-stranded target nucleic acid comprising a first region adjacent to a second region and said discrete region of said solid support comprising an immobilized melting component interacting with the first region;
- 10 b) providing a query probe that binds repeatedly to the second region to provide a detectable fingerprint; and
- c) associating the detectable fingerprint with the double-stranded nucleic acid to identify the double-stranded nucleic acid.
- 15
44. Use of a complex to provide a detectable fingerprint of a double-stranded target nucleic acid in a sample, the complex comprising:
- a) a double-stranded target nucleic acid comprising a first region adjacent to a second region;
- 20 b) an immobilized melting component interacting with the first region to form a thermodynamically stable complex and provide the second region in a single-stranded form; and
- c) a query probe that binds repeatedly to the second region to provide a detectable fingerprint associated with the double-stranded target nucleic acid.
- 25
45. Use of a kit to provide a result related to the presence, absence, concentration, amount, or sequence properties of a target nucleic acid, the kit comprising a solid support, a dCas9/gRNA, and a query probe.
- 30
46. A data structure comprising a fingerprint of a double-stranded target nucleic acid in a sample produced according to a method comprising:
- a) immobilizing a double-stranded target nucleic acid to a discrete region of a solid support, said double-stranded target nucleic acid comprising a first region adjacent to a second region and said discrete region of said solid
- 35

- support comprising an immobilized melting component interacting with the first region;
- b) providing a query probe that binds repeatedly to the second region to provide a detectable fingerprint; and
- 5 c) associating the detectable fingerprint with the double-stranded nucleic acid to identify the double-stranded nucleic acid.
47. A data structure comprising a fingerprint of a double-stranded target nucleic acid in a sample produced using a system comprising:
- 10 a) an immobilized melting component capable of interacting with a first region of the double-stranded target nucleic acid to form a thermodynamically stable complex and provide a second region of the double-stranded target nucleic acid in a single-stranded form; and
- b) a query probe that binds repeatedly to the second region to provide a
- 15 detectable fingerprint associated with the double-stranded target nucleic acid.
48. A data structure comprising a fingerprint of a double-stranded target nucleic acid in a sample produced using a complex comprising:
- 20 a) a double-stranded target nucleic acid comprising a first region adjacent to a second region;
- b) an immobilized melting component interacting with the first region to form a thermodynamically stable complex and provide the second region in a single-stranded form; and
- 25

Figure 1

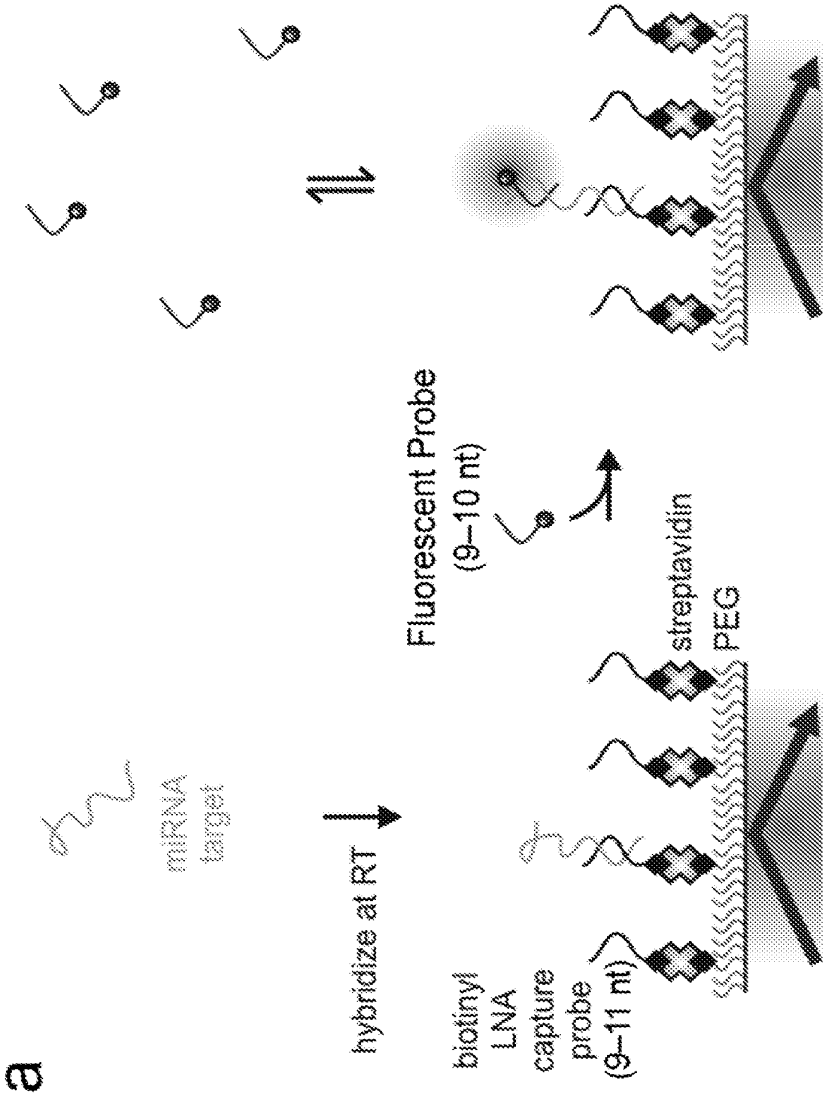


FIGURE 1 (cont.)

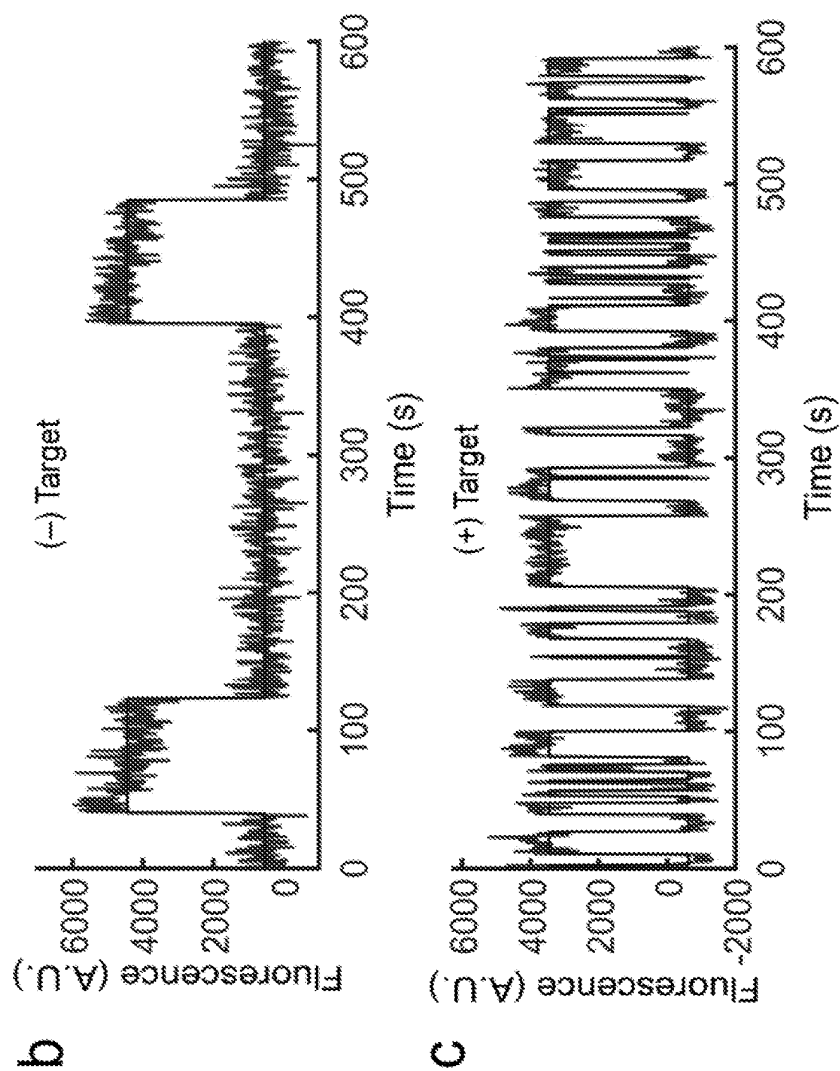


FIGURE 1 (cont.)

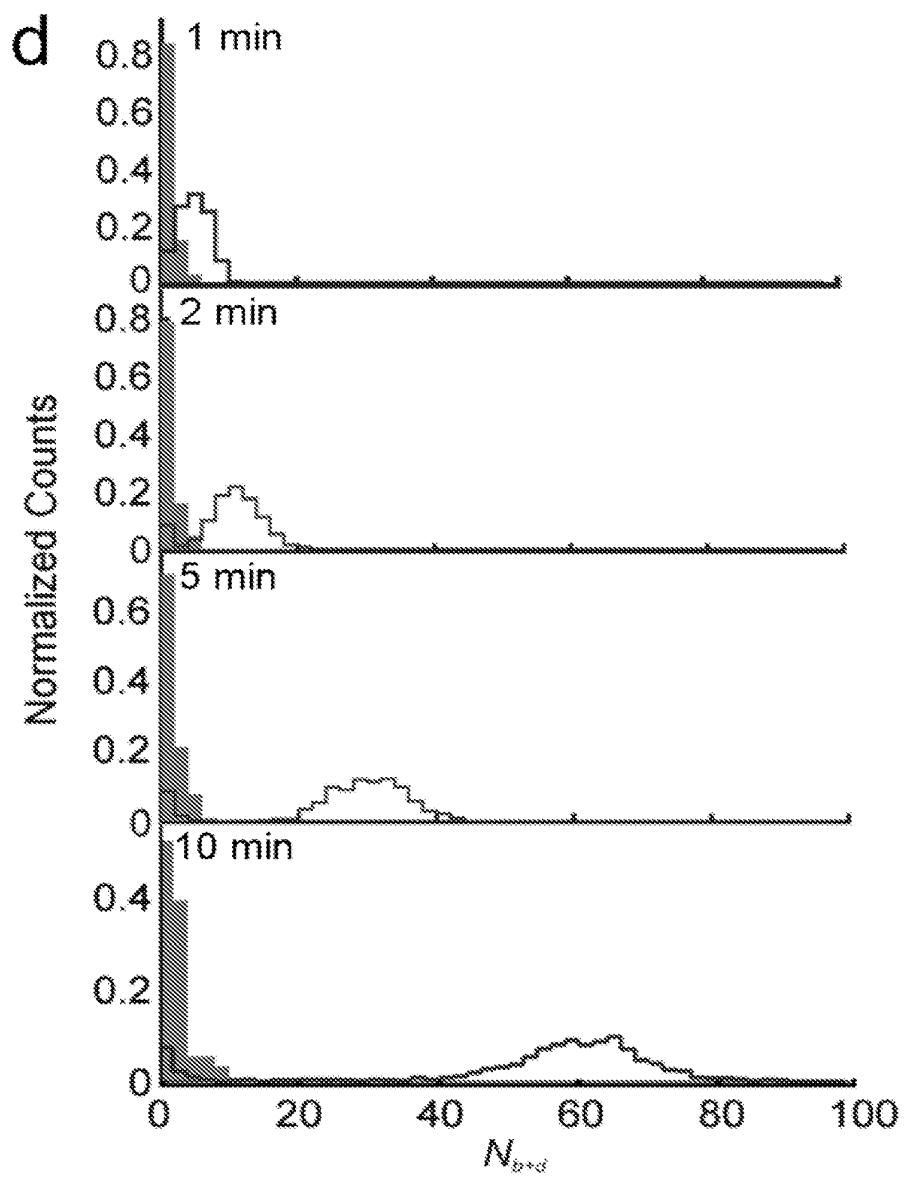


FIGURE 1 (cont.)

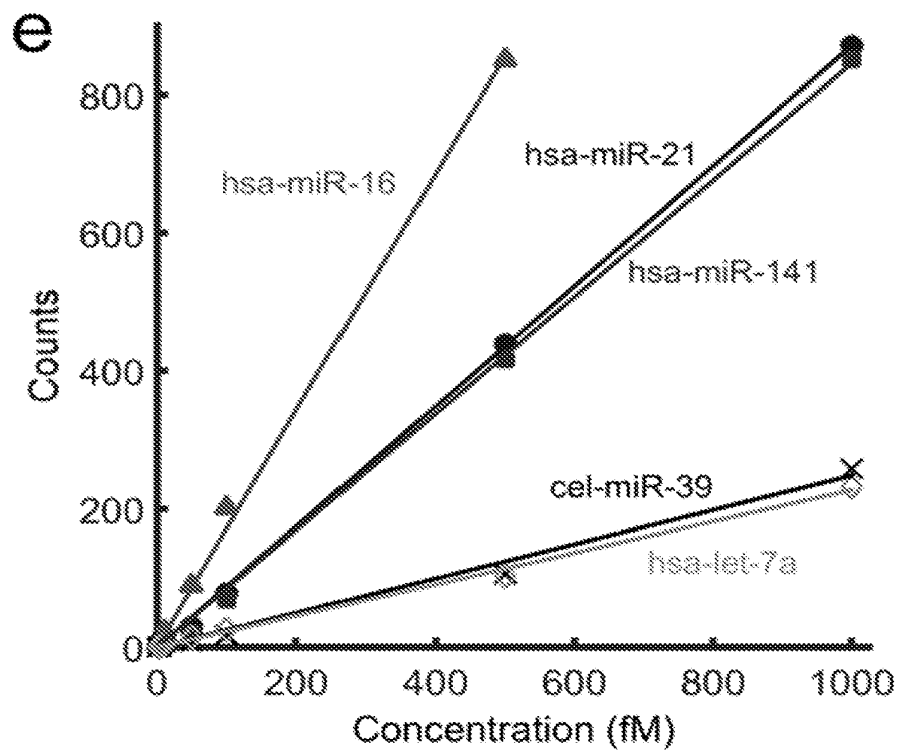


FIGURE 2

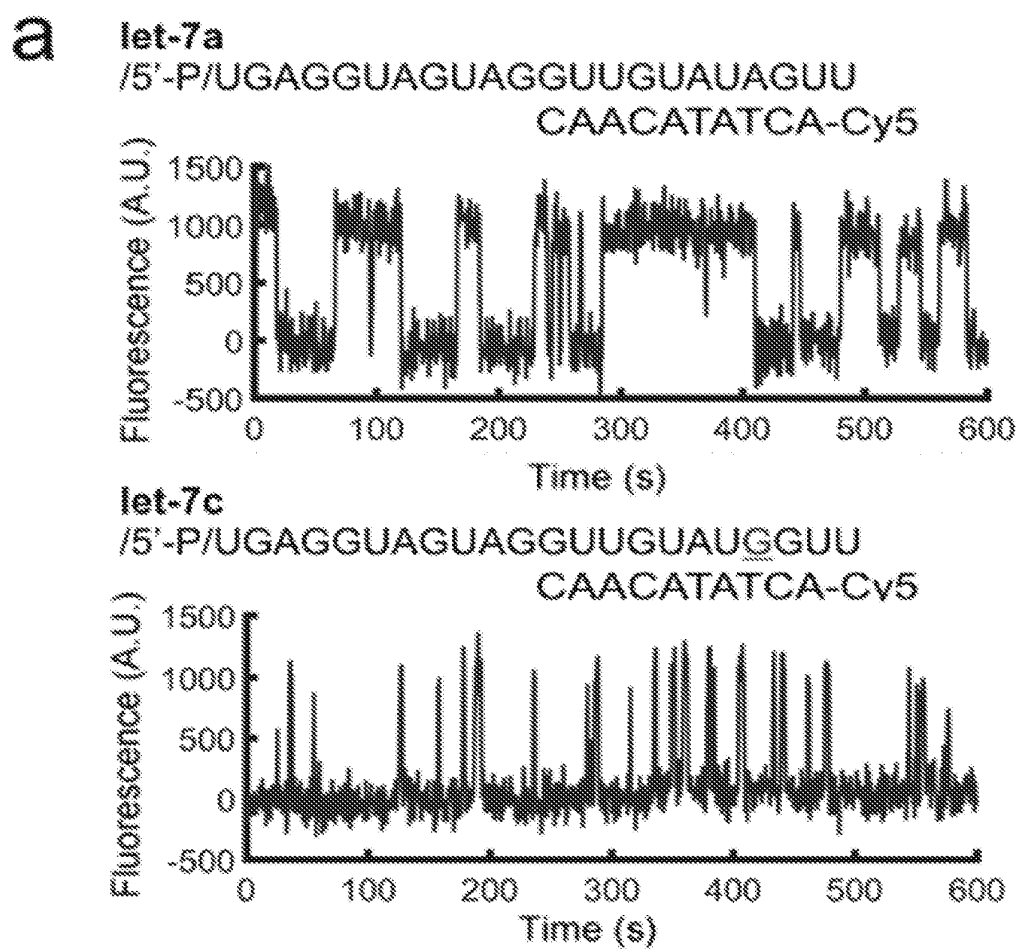


FIGURE 2 (cont.)

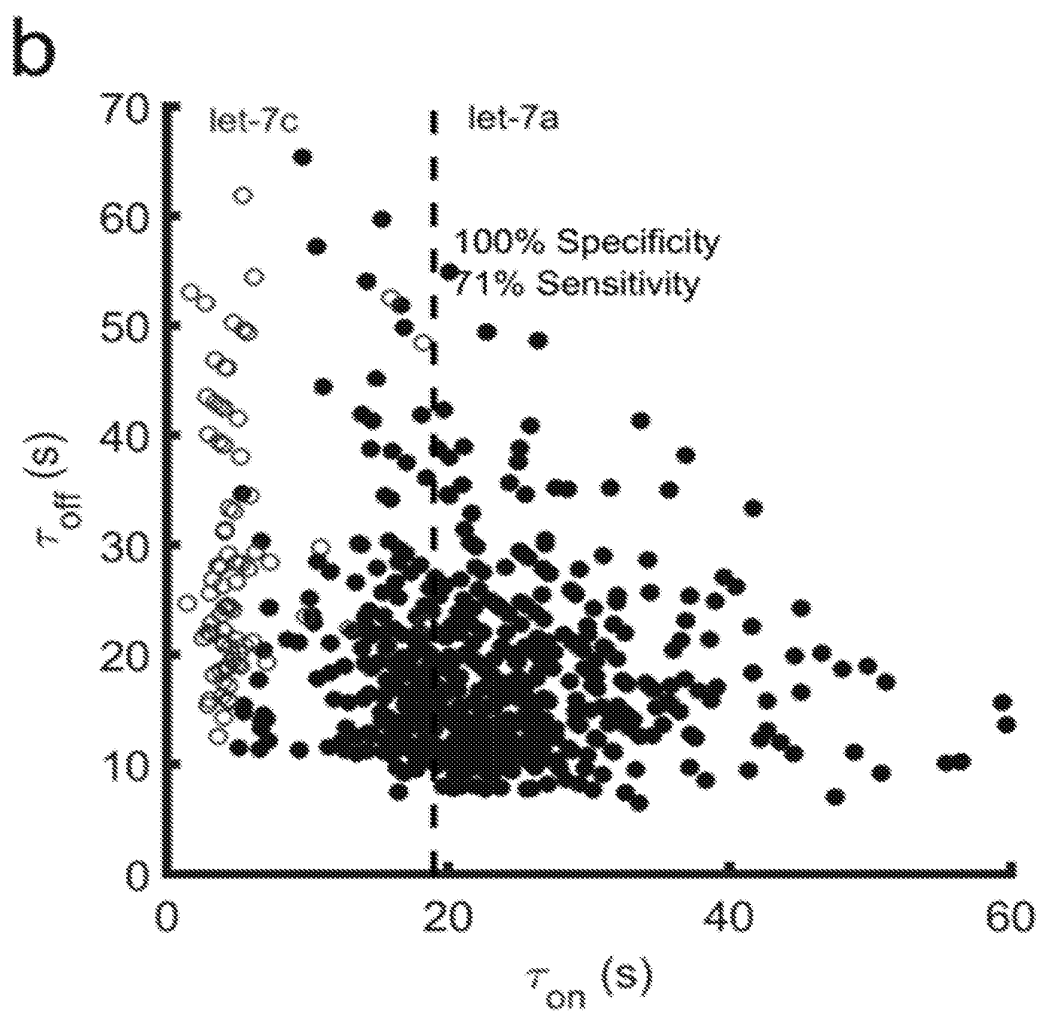


FIGURE 2 (cont.)

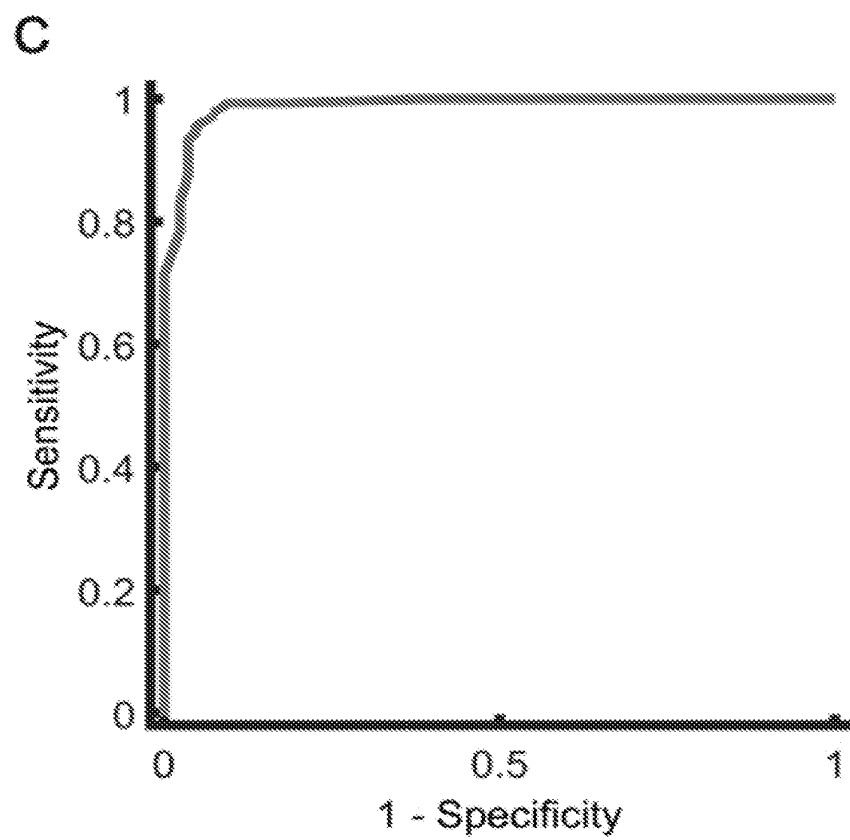


FIGURE 2 (cont.)

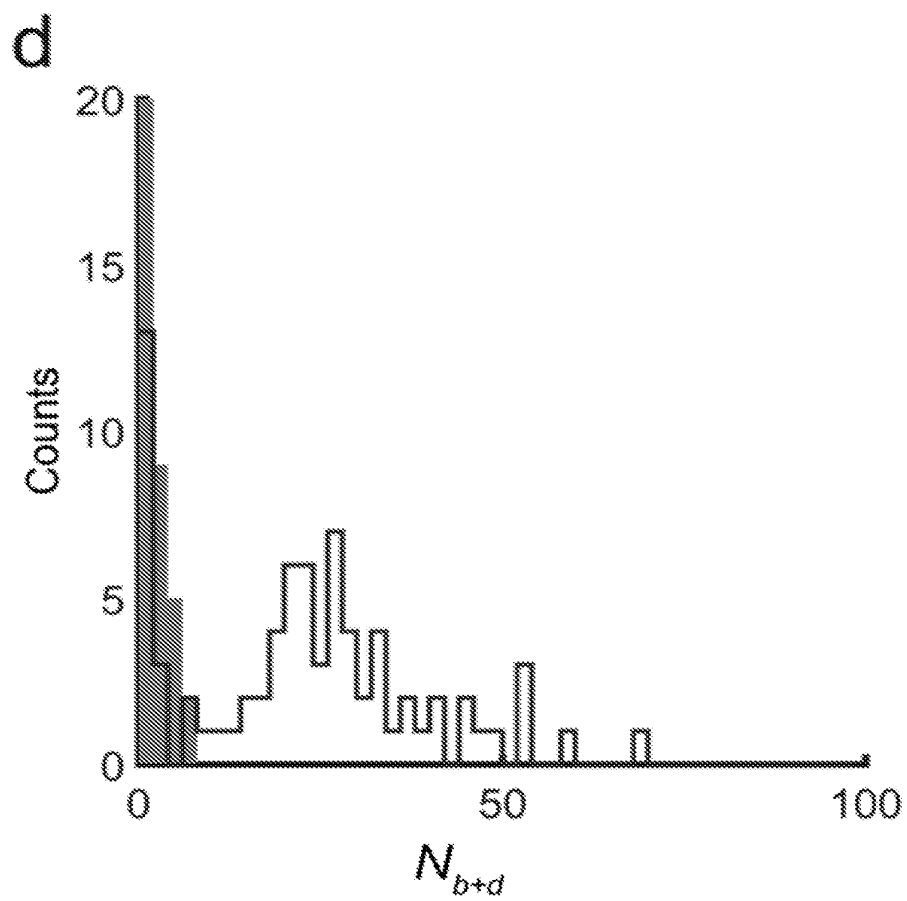


FIGURE 2 (cont.)

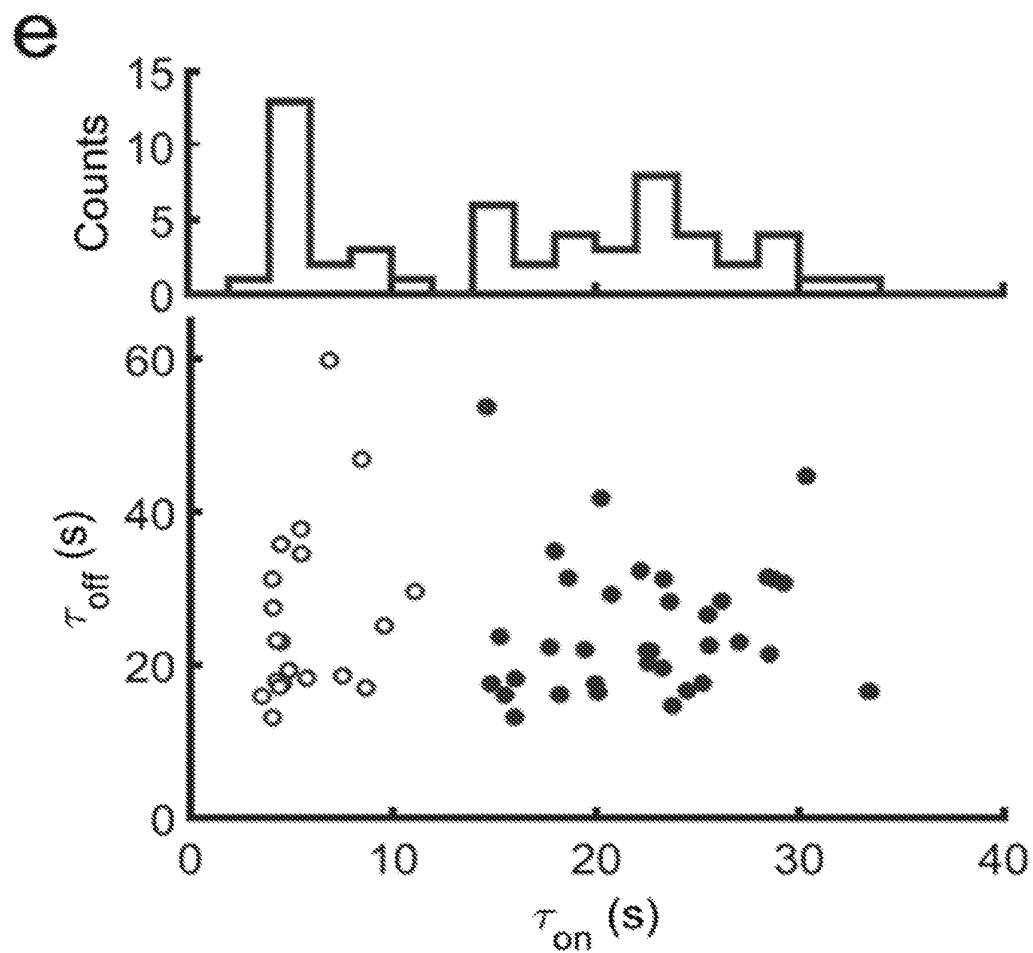


FIGURE 2 (cont.)

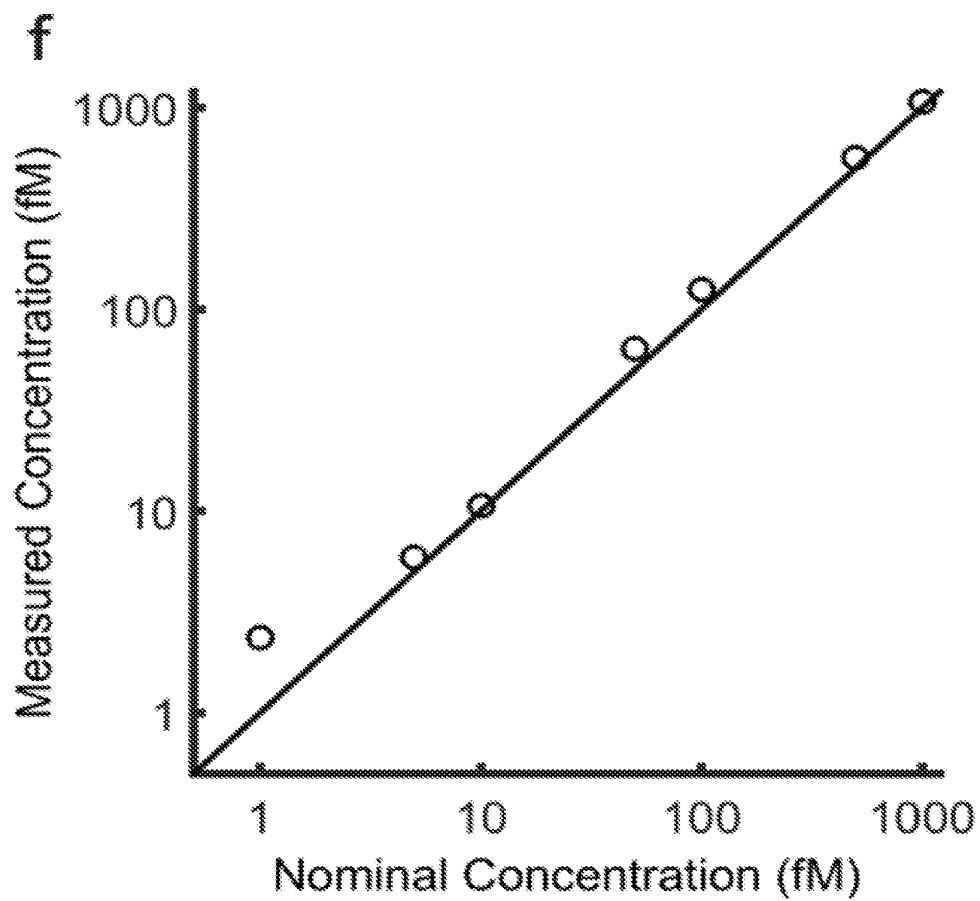


FIGURE 3

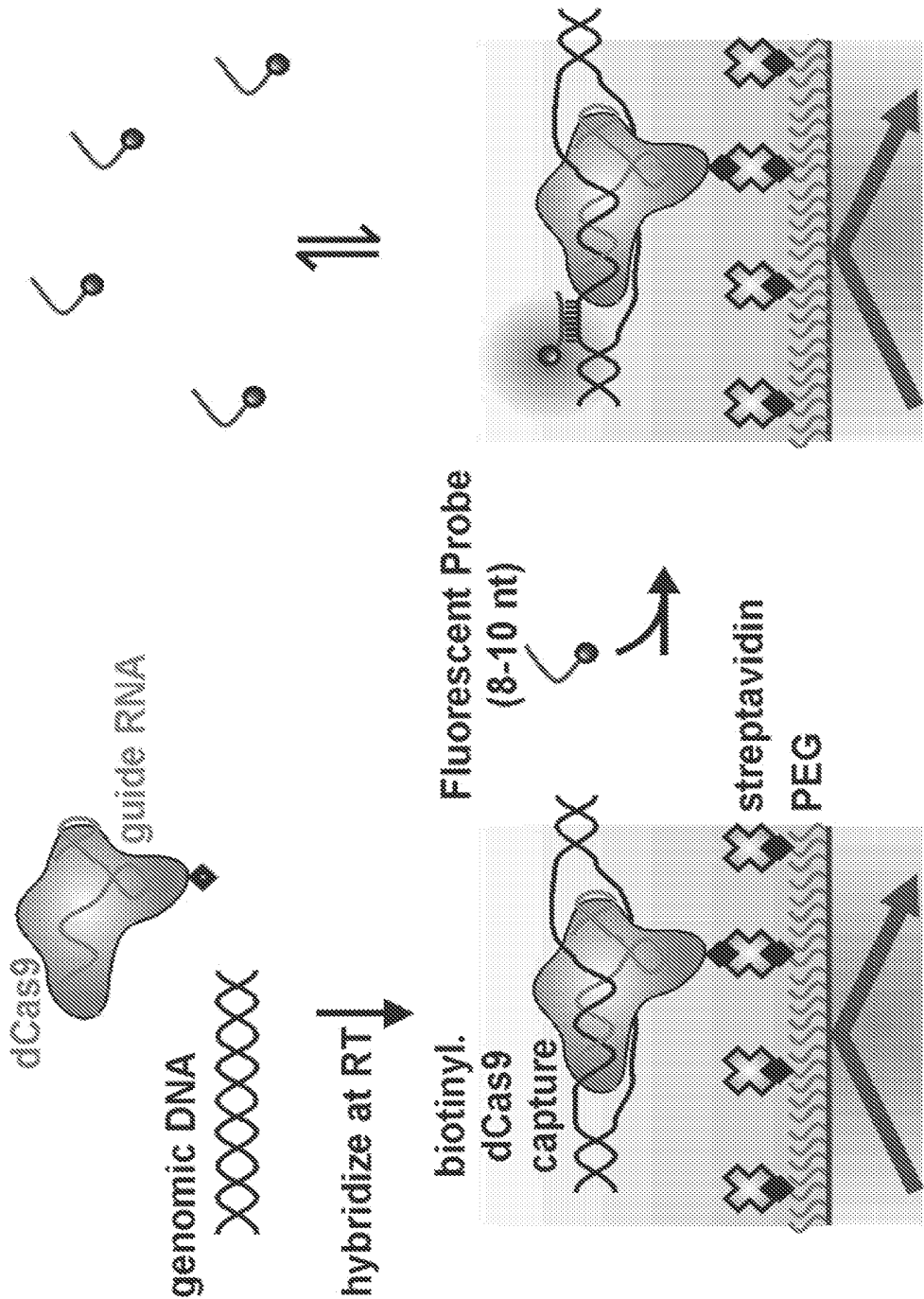


FIGURE 4

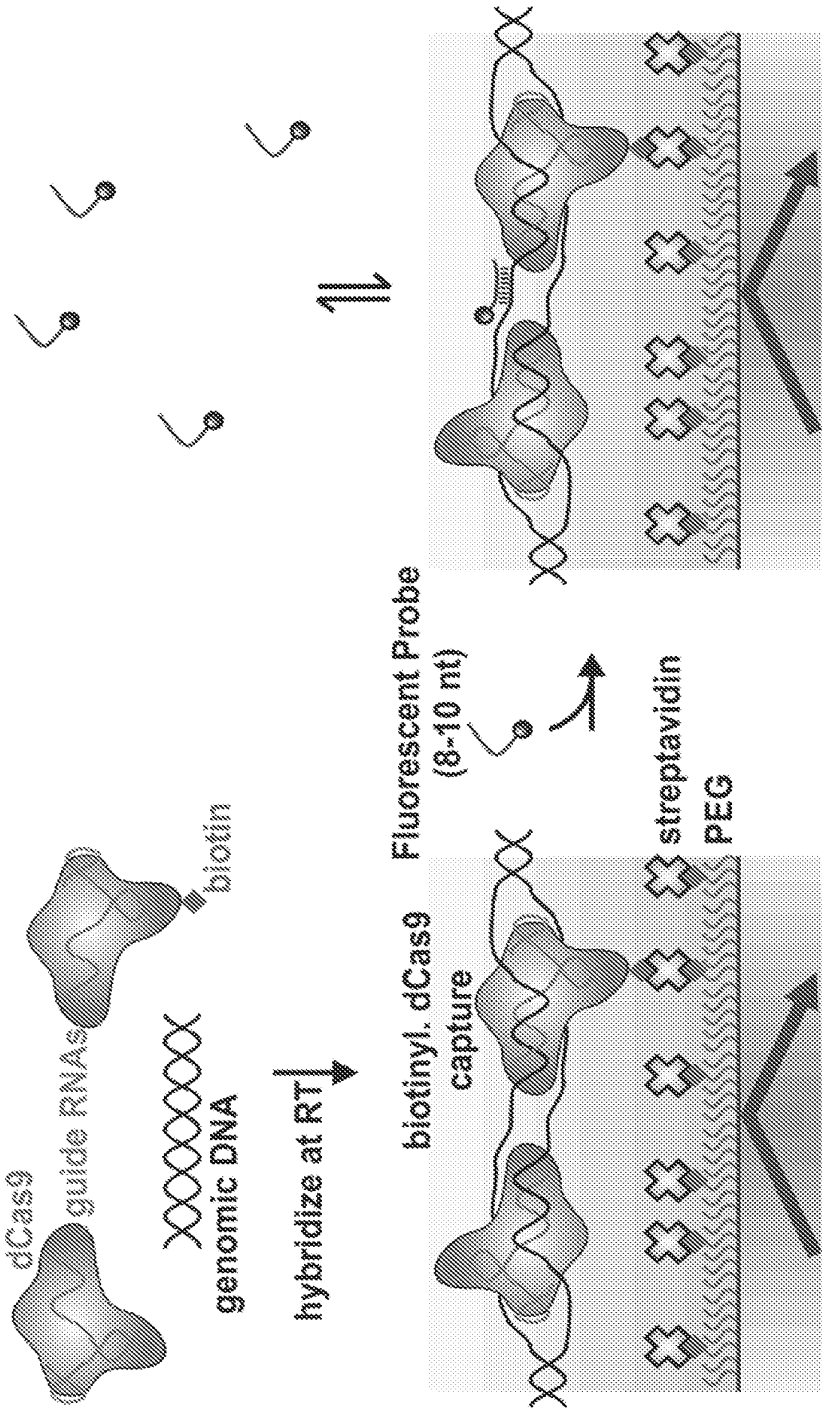


FIGURE 5

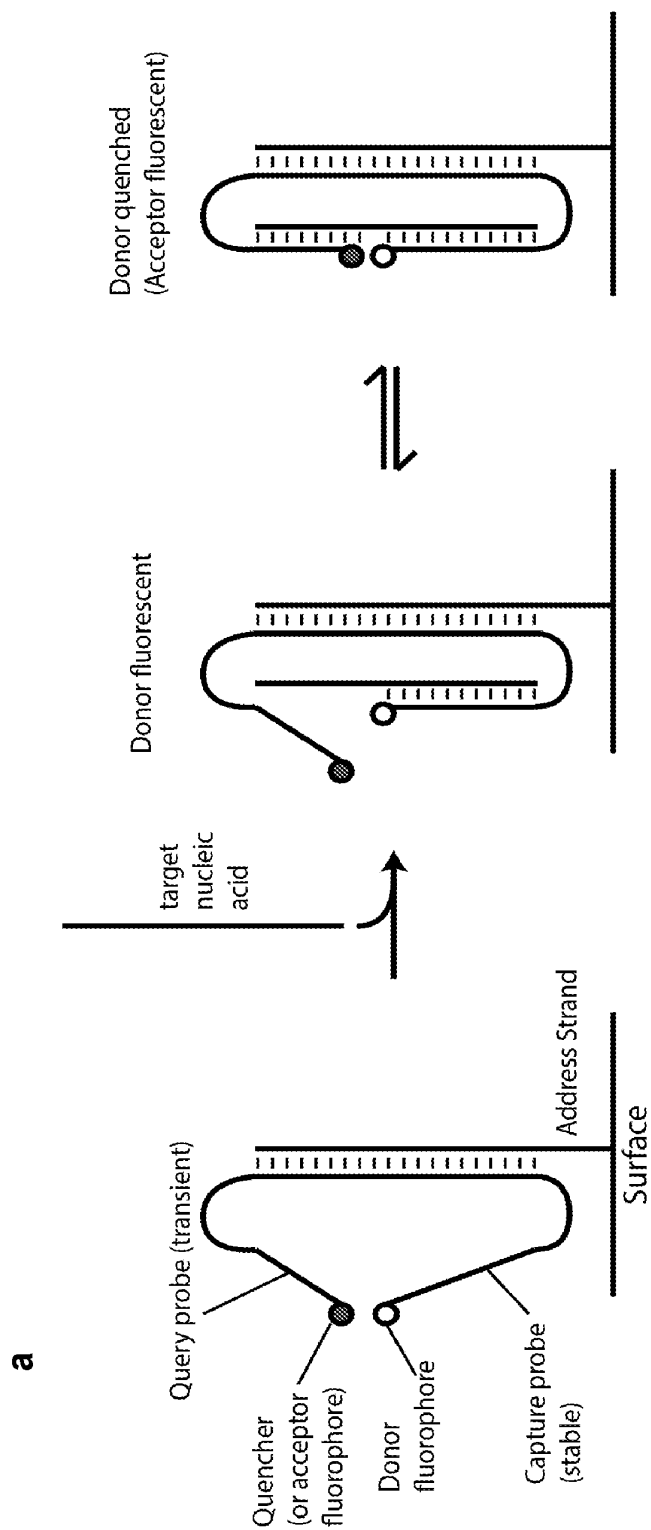
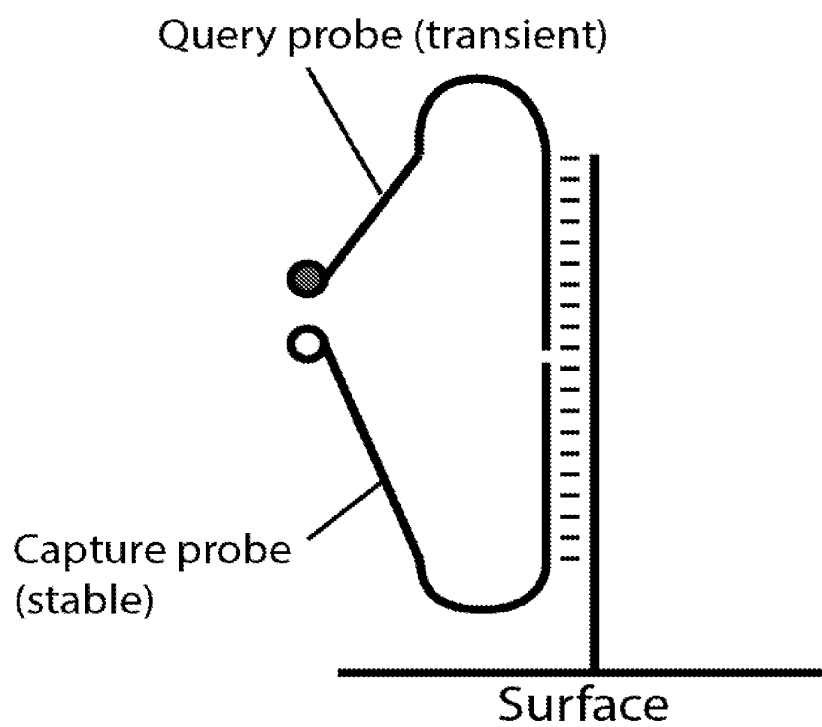


FIGURE 5 (cont.)

b

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/016977

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 9/22; C12P 19/34; C40B 30/04; C40B 40/06 (2017.01)

CPC - C12N 9/22; C12P 19/34; C12Q 1/6816; C12Q 1/6834; C12Q 1/6874; C12Q 2600/156; C40B 30/04; C40B 40/06; G01N 2500/00 (2017.02)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 435/6.11; 435/6.19; 436/501; 506/9 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOHNSON-BUCK et al. "Kinetic fingerprinting to identify and count single nucleic acids," Nat Biotechnol. 22 June 2015 (22.06.2015), Vol. 33, No. 7, Pgs. 730-732. entire document	19-21
Y		1-18, 22-48
Y	NASEF et al. "Melting temperature of surface-tethered DNA," Anal Biochem, 06 July 2010 (06.07.2010), Vol. 406, No. 1, Pgs. 34-40. entire document	1-18, 22-35, 42-44, 46-48
Y	SELLECK et al. "Biophysical characterization and direct delivery of S. Pyogenes Cas9 ribonucleoprotein complexes," Mol Ther, 27 April 2015 (27.04.2015), Vol. 23, Suppl. 1, Pg. S66. entire document	2, 3, 6, 13, 14, 23, 24, 27, 36-41, 45
Y	ANDERS et al. "Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease," Nature, 27 July 2014 (27.07.2014), Vol. 513, No. 7519, Pgs. 569-573. entire document	4, 25
A	ATHAMANOLAP et al. "Trainable high resolution melt curve machine learning classifier for large-scale reliable genotyping of sequence variants," PLoS One, 02 October 2014 (02.10.2014), Vol. 9, No. 9, e109094, Pgs. 1-10. entire document	1-48
P, X	WO 2016/172727 A1 (EDITAS MEDICINE, INC.) 27 October 2016 (27.10.2016) entire document	1-48
P, X	WO 2016/025477 A1 (THE REGENTS OF THE UNIVERSITY OF MICHIGAN) 18 February 2016 (18.02.2016) entire document	1-48

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

15 March 2017

Date of mailing of the international search report

25 APR 2017

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450
Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

IPR2024-01451

California Institute of Technology - Ex. 2003 Page 3448



(51) International Patent Classification:

C07H 21/00 (2006.01) G01N 21/64 (2006.01)
C07H 21/04 (2006.01) G01N 33/533 (2006.01)
C12Q 1/68 (2006.01) G01N 33/542 (2006.01)

(21) International Application Number:

PCT/US2017/024933

(22) International Filing Date:

30 March 2017 (30.03.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/317,151 1 April 2016 (01.04.2016) US

(71) Applicant: CHROMACODE INC. [US/US]; 2075 Corte
Del Nogal, Carlsbad, California 92011 (US).(72) Inventors: JACKY, Lucien A.E.; 166 N Waverly Street,
Orange, California 92866 (US). RAJAGOPAL, Aditya;
221 W. Sparkleberry Avenue, Orange, California 92865
(US). MENGE, Karen L.; 987 Gardenia Court, San Mar-
cos, California 92078 (US). GOSCH, Gregory; 14445
Caminito Lanza, San Diego, California 92127 (US).(74) Agent: JOHNSON, Lee A.; WILSON SONSINI
GOODRICH & ROSATI, 650 Page Mill Road, Palo Alto,
California 94304 (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, DJ, 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, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, ST, 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).

Declarations under Rule 4.17:

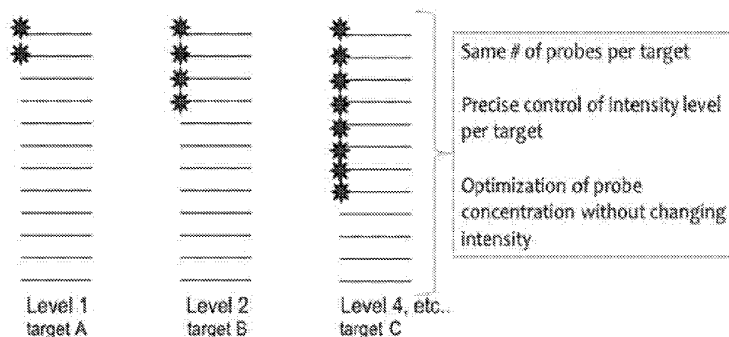
- as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))

(54) Title: COMPETITIVE PROBES FOR ENGINEERING SIGNAL GENERATION

FIG. 1



(57) Abstract: A method of identifying a first target nucleic acid comprising, providing a sample comprising the first target nucleic acid, providing a first set of paired oligonucleotides with complementarity to the first target nucleic acid, the first set of paired oligonucleotides comprising a first ratio of (a) first competitive oligonucleotides to (b) first signal oligonucleotides comprising a signal tag, wherein the competitive oligonucleotides compete with the signal oligonucleotides for binding to the first target nucleic acid, amplifying the first target nucleic acid with the polymerase chain reaction, thereby degrading the first signal oligonucleotide and permitting generation of a first signal, generating the first signal, measuring intensity of the first signal, and correlating the intensity of the first signal to the first ratio, thereby identifying the first target nucleic acid.

COMPETITIVE PROBES FOR ENGINEERING SIGNAL GENERATION**CROSS-REFERENCE**

[001] This application claims the benefit of U.S. Provisional Patent Application No. 62/317,151, filed April 1, 2016, which is hereby incorporated by reference in its entirety.

BACKGROUND

[002] Real-Time PCR is a process of monitoring a PCR reaction by recording the fluorescence generated either by an intercalating dye such as SYBR Green or a target-specific reporter probe at each cycle. This is generally performed on a Real-Time PCR instrument that executes thermal cycling of the sample to complete the PCR cycles and at a specified point in each cycle measures the fluorescence of the sample in each channel through a series of excitation/emission filter sets.

[003] In the age of whole genome sequencing and high-throughput analysis, there is a need for methods of detecting multiple qPCR targets at once. A limitation of current qPCR machines is the number of channels available for target detection. Modern instruments have between two and six channels, meaning that only two to six reporters can be detected in each experiment. Contemporary multiplexing technology provides for two to four different fluorescent reporters, and therefore two to four targets, to be detected in a single reaction. Unfortunately, accurately distinguishing each of the two to four targets from one another is thwarted by overlapping fluorescence signals of the different reporters. Therefore, a need still remains for a reliable and precise method for distinguishing multiple reporter signals during multiplex qPCR reactions.

SUMMARY

[004] The real-time polymerase chain reaction or quantitative PCR (qPCR) is a common molecular biology technique that uses fluorescent reporters to monitor the progress of the exponential amplification of target nucleic acids in a reaction. Traditionally the process is limited to differentiating the presence of one target per fluorescent emission filter (color channel) on a qPCR instrument.

[005] Recent advances made at the California Institute of Technology nanofabrication lab allow for detection of multiple analytes per emission filter using mathematical algorithms on real-time PCR amplification curve characteristics such as amplitude and curve shape (US Patent 8,838,394 issued September 16, 2014 and US Patent Application Publication Number US2015-0057178 published February 26, 2015, both of which are incorporated herein in their entirety). Targets in the same color channel are “coded” to exhibit these amplification curve characteristics. Coding includes setting the intensity level of each target in a color channel.

[006] Disclosed herein are methods by which the effective relative fluorescent intensity generated by a qPCR probe is adjusted independent of the total probe concentration for each target. This is a key improvement to the multi-analyte coding technique.

[007] Disclosed herein are methods and compositions for uniquely and unambiguously detecting multiple targets in a single color detection channel. One aspect of the disclosed methods is the ability to differentiate multiple targets by ratio-metric coding of signal intensity (brightness) of any given color for each target and mathematical strategies to circumvent degeneracy. When used within each of the available color channels on a qPCR instrument, the disclosed methods significantly increase the number of target measurements that can be interrogated.

[008] Provided herein are methods of identifying a first target nucleic acid comprising, providing a sample comprising the first target nucleic acid, providing a first set of paired oligonucleotides with complementarity to the first target nucleic acid, said first set of paired oligonucleotides comprising a first ratio of (a) first competitive oligonucleotides to (b) first signal oligonucleotides comprising a first signal tag, wherein the first competitive oligonucleotides compete with the first signal oligonucleotides for binding to the first target nucleic acid, amplifying the first target nucleic acid with the polymerase chain reaction, thereby degrading the first signal oligonucleotide and permitting generation of a first signal, generating the first signal, measuring intensity of the first signal, and correlating the intensity of the first signal to the first ratio, thereby identifying the first target nucleic acid. In some aspects, the sample comprises a second target nucleic acid. In some aspects, the method further comprises identifying the second target nucleic acid by providing a second set of paired oligonucleotides with complementarity to the second target nucleic acid, said second set of paired oligonucleotides comprising a second ratio of (a) second competitive oligonucleotides to (b) second signal oligonucleotides comprising a second signal tag, wherein the second competitive oligonucleotides compete with the second signal oligonucleotides for binding to the second target nucleic acid, amplifying the second target sequence with the polymerase chain reaction, thereby degrading the second signal oligonucleotide and permitting generation of a second signal, generating the second signal, measuring intensity of the second signal, and correlating the intensity of the second signal to the second ratio, thereby identifying the second target nucleic acid. In some aspects, the first signal oligonucleotides and the second signal oligonucleotides comprise the same signal tag. In some aspects, the first target nucleic acid and the second target nucleic acid are identified from differing signal intensities generated during the same polymerase chain reaction. In some aspects, the first signal intensity is different from the second signal intensity. In some aspects, the paired oligonucleotides comprise qPCR probes. In

some aspects, the signal oligonucleotides comprise a signal tag and a signal quenching tag. In some aspects, the signal tag comprises a fluorophore. In some aspects, the signal quenching tag blocks the generation of the signal corresponding with the signal tag until the signal quenching tag is separated from the signal tag by oligonucleotide degradation during the polymerase chain reaction, thereby permitting the signal to be generated. In some aspects, generating the signal comprises exciting the signal tag. In some aspects, exciting comprises directing at the signal tag at least one of illumination, electromagnetic radiation, and energy emitted from an energy source. In some aspects, the competitive oligonucleotides do not comprise a signal tag. In some aspects, the competitive oligonucleotides comprise at least one modified end. In some aspects, the modified end prevents enzymatic extension during amplification. In some aspects, the modified end comprises a 3' end lacking a hydroxyl group. In some aspects, the modified end comprises a 5' end lacking a phosphate group. In some aspects, the modified end comprises a 5' capped end. In some aspects, the modified end comprises a 3' capped end. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:1. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:2. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:3. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:4. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:5. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:6. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:7. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:8. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:9. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:10. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:16. In some aspects, the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:20. In some aspects, the paired oligonucleotides comprise only

signal oligonucleotides or only competitive oligonucleotides. In some aspects, the total concentration of the first set of paired oligonucleotides is the same as the total concentration of the second set of paired oligonucleotides. In some aspects, the methods further comprise identifying up to thirty target nucleic acid sequences using up to thirty sets of paired oligonucleotides, wherein each set of paired oligonucleotides possesses complementarity to one of the up to thirty target nucleic acid sequences, wherein each of the up to thirty sets of paired oligonucleotides comprises a unique ratio of (a) competitive oligonucleotides to (b) signal oligonucleotides comprising a signal tag, wherein the competitive oligonucleotides within each set compete with the signal oligonucleotides within the same set for binding to a corresponding target nucleic acid sequence, amplifying the up to thirty target nucleic acid sequences with the polymerase chain reaction, thereby degrading the signal oligonucleotides and permitting generation of up to thirty signals, generating up to thirty signals, measuring the intensities of the up to thirty signals, and correlating the intensities of the up to thirty signals to the up to thirty ratios, thereby identifying the up to thirty target nucleic acid sequences. In some aspects, the signal tag of the signal oligonucleotides within each of the up to thirty sets of paired oligonucleotides are the same. In some aspects, the up to thirty target nucleic acid sequences are identified from signal intensities generated during the same polymerase chain reaction. In some aspects, each of the up to thirty ratios differs from every other of the up to thirty ratios. In some aspects, each of the up to thirty signal intensities differs from every other of the up to thirty signal intensities. In some aspects, the polymerase chain reaction is carried out by use of a nucleic acid polymerase having 5' to 3' exonuclease activity, an exo-polymerase, or a molecular beacon. In some aspects, the up to five signals are measured in a single fluorescence channel. In some aspects, the total concentration of each of the up to five sets of paired oligonucleotides is the same. In some aspects, a total number of molecules in the reaction remains constant.

[009] Provided herein are compositions comprising a first set of paired oligonucleotides with complementarity to a first target nucleic acid, said first set of paired oligonucleotides comprising a first ratio of (a) competitive oligonucleotides to (b) signal oligonucleotides comprising a signal tag, wherein the competitive oligonucleotides are configured to compete with the signal oligonucleotides for binding to the first target nucleic acid during amplification of the first target nucleic acid with the polymerase chain reaction, wherein the signal oligonucleotides are adapted to generate a first signal having a first intensity when degraded by a nucleic acid polymerase during the polymerase chain reaction and excited by an energy source, and wherein the first intensity correlates with the first ratio.

[010] Provided herein are methods of identifying a first target molecule comprising, providing a sample comprising the first target molecule, providing a first set of paired capture molecules

that bind to the first target molecule, said first set of paired capture molecules comprising a first ratio of (a) first competitive capture molecules to (b) first signal capture molecules comprising a signal tag, wherein the competitive capture molecules compete with the signal capture molecules for binding to the first target molecule, generating the first signal, measuring intensity of the first signal, and correlating the intensity of the first signal to the first ratio, thereby identifying the first target molecule. In some aspects, the sample comprises a second target molecule. In some aspects, the methods further comprise identifying the second target molecule by providing a second set of paired capture molecules that binds to the second target molecule, said second set of paired capture molecules comprising a second ratio of (a) second competitive capture molecules to (b) second signal capture molecules comprising the signal tag, wherein the second competitive capture molecules compete with the second signal capture molecules for binding to the second target molecule, generating the second signal, measuring intensity of the second signal, and correlating the intensity of the second signal to the second ratio, thereby identifying the second target molecule. In some aspects, the first signal capture molecules and the second signal capture molecules comprise the same signal tag. In some aspects, the first target capture molecules and the second target capture molecules are identified from differing signal intensities generated. In some aspects, the first signal intensity is different from the second signal intensity. In some aspects, the signal tag comprises a fluorophore. In some aspects, generating the signal comprises exciting the signal tag. In some aspects, exciting comprises directing at the signal tag at least one of illumination, electromagnetic radiation, and energy emitted from an energy source. In some aspects, the competitive capture molecules do not comprise a signaling tag. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:1. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:2. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:3. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:4. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:5. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:6. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:7. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the

ratio of competitive capture molecules to signal capture molecules is about 1:8. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:9. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:10. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:16. In some aspects, the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:20. In some aspects, the first paired capture molecules comprise only signal capture molecules or only competitive capture molecules. In some aspects, the methods further comprise identifying up to one hundred target molecules using up to one hundred sets of paired capture molecules, wherein each set of paired capture molecules that binds to one of the up to one hundred target molecules, wherein each of the up to one hundred sets of paired capture molecules comprises a unique ratio of (a) competitive capture molecules to (b) signal capture molecules comprising a signal tag, wherein the competitive capture molecules within each set compete with the signal capture molecules within the same set for binding to a corresponding target molecule, generating up to one hundred signals, measuring the intensities of the up to one hundred signals, and correlating the intensities of the up to one hundred signals to the up to one hundred ratios, thereby identifying the up to one hundred target molecules. In some aspects, the signal tag of the signal capture molecules within each of the up to one hundred sets of capture molecules are the same. In some aspects, the up to one hundred target molecules are identified from signal intensities generated at the same time. In some aspects, each of the up to one hundred ratios differs from every other of the up to one hundred ratios. In some aspects, each of the up to one hundred signal intensities differs from every other of the up to one hundred signal intensities. In some aspects, the up to one hundred signals are measured in a single fluorescence channel. In some aspects, the target molecule is fixed to a surface. In some aspects, the capture molecule is fixed to a surface. In some aspects, the target molecule comprises a protein or a functional fragment thereof. In some aspects, the target molecule comprises an antigen. In some aspects, the target molecule comprises an aptamer. In some aspects, the target molecule comprises a nucleic acid. In some aspects, the capture molecule comprises a protein or a functional fragment thereof. In some aspects, the capture molecule comprises an antibody or a protein or a functional fragment thereof. In some aspects, the capture molecule comprises an aptamer. In some aspects, a total number of molecules in the reaction remains constant.

[011] Provided herein are compositions comprising a first set of paired capture molecules with complementarity to a first target molecule, said first set of paired capture molecules comprising

a first ratio of (a) competitive capture molecules to (b) signal capture molecules comprising a signal tag, wherein the competitive capture molecules are configured to compete with the signal capture molecules for binding to the target molecule, wherein the signal capture molecules are adapted to generate a first signal having a first intensity when excited by an energy source, and wherein the first intensity correlates with the first ratio. In some aspects, the target molecule comprises a protein or a functional fragment thereof. In some aspects, the target molecule comprises an antigen. In some aspects, the target molecule comprises an aptamer. In some aspects, the target molecule comprises a nucleic acid. In some aspects, the capture molecule comprises a protein or a functional fragment thereof. In some aspects, the capture molecule comprises an antibody or a functional fragment thereof. In some aspects, the capture molecule comprises an aptamer.

[012] Provided herein are methods of altering amplification level of a target nucleic acid comprising, providing a sample comprising the target nucleic acid, providing a set of paired oligonucleotides with complementarity to the target nucleic acid, said set of paired oligonucleotides comprising a ratio of (a) competitive oligonucleotides to (b) functional oligonucleotides, wherein the competitive oligonucleotides compete with the functional oligonucleotides for binding to the target nucleic acid and are not extendible by a nucleic acid polymerase, wherein the functional oligonucleotides are extendible by a nucleic acid polymerase as primers for amplifying the target nucleic acid by the polymerase chain reaction (PCR), and amplifying the target nucleic acid with PCR. In some aspects, the competitive oligonucleotides comprise at least one modified end. In some aspects, the modified end prevents enzymatic extension during amplification. In some aspects, the modified end comprises a 3' end lacking a hydroxyl group. In some aspects, the modified end comprises a 5' end lacking a phosphate group. In some aspects, the modified end comprises a 5' capped end. In some aspects, the modified end comprises a 3' capped end. In some aspects, a total number of molecules in the reaction remains constant.

[013] Provided herein are compositions comprising a set of paired oligonucleotides with complementarity to a target nucleic acid, said set of paired oligonucleotides comprising a ratio of (a) competitive oligonucleotides to (b) functional oligonucleotides, wherein the competitive oligonucleotides are configured to compete with the functional oligonucleotides for binding to the nucleic acid target during amplification of the first target nucleic acid with the polymerase chain reaction and are not extendible by a nucleic acid polymerase.

BRIEF DESCRIPTION OF THE DRAWINGS

[014] **Figure 1** illustrates three example probe sets with different ratios of competitive to fluorescent probes

[015] **Figure 2** illustrates an example PIV3 benchmark assay titration performed at four distinct coding levels.

[016] **Figures 3A-3B** illustrate an example RT-qPCR experiment with five distinct ratio levels of labeled and unlabeled competitive probes.

[017] **Figures 4A-4B** illustrate an example RT-qPCR experiment with three distinct ratio levels of labeled and unlabeled competitive probes.

DETAILED DESCRIPTION

[018] Without being bound by theory, Polymerase Chain Reaction is a method of exponential amplification of specific target DNA in a reaction mix with a DNA polymerase and primers. Primers are short single stranded DNA oligonucleotides which are complementary to the 3' sequences of the positive and negative strand of the target sequence. The reaction mix is cycled in repeated heating and cooling steps. The heating cycle denatures or split the double stranded DNA target into single stranded templates. In the cooling cycle, the primers bind to complementary sequence on the template. After the template is primed the DNA polymerase creates a copy of the original template. Repeated cycling exponentially amplifies the target 2 fold with each cycle leading to approximately a billion-fold increase of the target sequence in 30 cycles (Saiki et al 1988).

[019] Real-Time PCR is a process of monitoring a PCR reaction by recording the fluorescence generated either by an intercalating dye such as SYBR Green or a target-specific reporter probe at each cycle. This is generally performed on a Real-Time PCR instrument that executes thermal cycling of the sample to complete the PCR cycles and at a specified point in each cycle measures the fluorescence of the sample in each channel through a series of excitation/emission filter sets

[020] Frequently, the target specific reporter probe is a short oligonucleotide complementary to one strand of the amplified target. The probe lacks a 3' hydroxyl and therefore is not extendable by the DNA polymerase. TaqMan (ThermoFisher Scientific) chemistry is a common reporter probe method used for multiplex Real-Time PCR (Holland et al. 1991). The TaqMan oligonucleotide probe is covalently modified with a fluorophore and a quenching tag. In this configuration the fluorescence generated by the fluorophore is quenched and is not detected by the real time PCR instrument. When the target of interest is present, the probe oligonucleotide base pairs with the amplified target. While bound, it is digested by the 5'->3' exonuclease

activity of the Taq polymerase thereby physically separating the fluorophore from the quencher and liberating signal for detection by the real time PCR instrument. Other methods, such as molecular beacon probes, are also used in real-time PCR and amenable to this competitive probe technique.

[021] For certain qPCR, multiplexed qPCR, and coded qPCR methods, there is a need to adjust signal level in a predictable fashion. This is typically done by increasing or decreasing the number of fluorometric reporters added to a reaction. While this does adjust signal level to the first order, it also risks altering the chemical behavior of the reaction mixture as the total number of participating reactants has also changed. Often, this adversely affects the performance of other components within the reaction mixture. In particular, this is a critical failure mode for multiplexed qPCR. Disclosed herein are methods and compositions for detecting multiple target sequences with the same fluorophore within a single coded or multiplexed qPCR reaction.

[022] Disclosed herein are methods and compositions for adjusting fluorescent intensity without adjusting the reactant concentrations in a mixture. In some embodiments, combinations of both light-emitting and non-light-emitting chemical reporters are used to independently adjust the fluorescent intensity of a qPCR reaction without altering the chemical interactivity of the qPCR reaction mixture. In some embodiments, this is achieved by creating a mixture of non-fluorescent oligo probes and the TaqMan reporter probes for every target interrogated in a qPCR reaction.

[023] The herein disclosed competitive probes (also referred to as dimming probes, or non-fluorescing probes) are oligonucleotides similar to target probes with the exception of the covalently attached fluorophore and quencher. In some embodiments, the competitive probes do not have the fluorophore or the quencher. In some embodiments, the competitive probes have capped 5' and 3' ends to prevent enzymatic extension. In some embodiments, the competitive probes have a 5' end lacking a phosphate group. In some embodiments, the competitive probes have a 3' end lacking a hydroxyl group. In general, competitive probes behave identically to the specific target probes in binding to the amplified target sequence, and in the case of TaqMan style assay, the bound probes are digested by the 5'→3' exonuclease activity of Taq polymerase. However, because the competitive probes do not contain a fluorophore, no signal is generated.

[024] Disclosed herein are methods and compositions for using the disclosed non-fluorescing, competitive binding probes to adjust the signal intensity for a single gene by ratiometric titration of the competitive binding probe. In some examples, these methods are performed using a TaqMan probe. Figure 1 diagrams three different probes mixtures with differing proportions of competitive probes and signal probes. In some embodiments, the titration of the intensity allows

for control of intensity level ratio-metric coding and the ability to adjust the total probe concentration of a given target without affecting the intensity level coding.

[025] In some embodiments, during the probe annealing portion of the PCR reaction the competitive probes will bind to the amplified target at a rate proportional to the ratio of competitive probes to the total probes both labeled and unlabeled in the reaction yielding a fluorescent signal proportional to the ratio of labeled probes to the total probes.

[026] In some embodiments, the competitive probes are un-extendable probes. For example, the competitive probes are unable to initiate DNA synthesis by the Taq polymerase by employing a variety of 3' modifications including but not limited to the phosphorylation or the addition of a quenching molecule.

[027] In some embodiments, different purification methods and probe sequences are used to yield dramatically different fluorescence intensities. In some examples, differences in the quality of purification leads to more or less unlabeled product (similar to a competitive probe) reducing the fluorescence of a preparation. In some embodiments, the oligo sequence makeup of a probe affects the fluorescence, particularly guanine bases proximal to the fluorophore which can act as a quencher reducing the intensity of fluorescence. In some embodiments, the competitive probes are used to adjust the signal intensity to be consistent between preparations for more consistent lot to lot performance. In some embodiments, competitive probes are used if a dye's intensity is overloading the detector on a particular real-time PCR instrument.

[028] In some embodiments, competitive probes are used to engineer the optical characteristic of a PCR curve without altering its chemical characteristic. In some examples, this is achieved because competitive probes competitively interact with other sequences, along with their TaqMan probe equivalents. In these examples, this competitive interaction allows for the unique capability to fine tune the signal to noise of a PCR reaction without changing the chemical performance of the PCR itself.

[029] Disclosed herein are methods and compositions for using competitive primers for multiplex quantitation by normalizing the efficiency of multiplexed reactions. In some embodiments, the herein disclosed methods are used to drive the PCR reaction into producing more of one strand than the other which can lead to an earlier plateau phase of the PCR reaction.

[030] In some embodiments, competitive probes are used to replace a portion of priming oligonucleotides in a PCR reaction. In many of these examples, the competitive probes are rendered un-extendable by 3' phosphorylation or other methodology. In these cases, the competitive primer will compete with the extendable primer leading to a quantifiable fraction of template strands to fail to replicate during the extension phase of the PCR. One advantage of

these embodiments is that it allows the adjustment of amplification efficiency of either one or both of the replication strands.

[031] In some embodiments, the herein disclosed competitive probes are used in TaqMan style PCR. In some embodiments, the herein disclosed competitive probes are used in traditional PCR.

[032] Disclosed herein are methods for identifying a target by detecting a corresponding signal, such as a fluorescent signal. In some examples, multiple targets can be detected in a single experiment, wherein the corresponding signal from each is the same. For example, multiple targets can be identified, where each target has a corresponding fluorescent signal, wherein the signal generating molecule, such as a fluorophore, is the same among each of the different targets to be detected.

[033] Using methods disclosed herein, multiple targets can be detected using the same fluorophore by providing sets of target-binding molecules, such as probes, wherein each set of molecules comprises a unique ratio of competitive probes to signal probes. Signal probes can comprise a signal generating molecules, such as a fluorophore. Competitive probes often do not contain the signal generating molecule. In these cases, the total number of probes within each set of probes is the same, which the ratio of competitive probes to signal probes is unique for each target to be detected. This allows for the same level or concentration of total probes to be used for each target while still being able to distinguish each individual target.

[034] Disclosed herein are methods and compositions for detecting multiple target sequences with the same fluorophore. Also disclosed herein are methods and compositions for detecting multiple target specific reporter signals in a single reaction using fluorophore baring oligonucleotide probes in combination with non-fluorophore baring oligonucleotide probes. In some examples, each target sequence is detected using a set of oligonucleotide probes. Each oligonucleotide set can comprise two populations of oligonucleotide probes, each comprising the same nucleic acid sequence. In these examples, one population of oligonucleotides comprises a reporter fluorophore while the second population of oligonucleotides does not comprise a fluorophore. In some embodiments, up to five sets of oligonucleotide probes are used to detect up to five target sequences in a single reaction. In many of these examples, the fluorophore of the up to five sets of probes is the same. In some aspects, within each set of oligonucleotide probes the ratio of fluorophore baring and non-fluorophore baring oligonucleotides is set to a predetermined ratio. In some embodiments, the ratio within each set of oligonucleotides is different for each of the up to five sets of oligonucleotides. In these examples, the different ratios allow the reporter signal for each of the up to five targets to be distinguished from one another within a single reaction. In these examples, the total

concentration of probes in each set of probes is the same, while the ratio of the two populations of probes within the probe set are different between each of the sets of probes. A fluorophore bearing probe can be a signal probe. A non-fluorophore bearing probe can be a competitive probe.

[035] In some embodiments, up to 100 sets of oligonucleotide probes are used to detect up to 100 target sequences in a single reaction. For example, there can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 sets of oligonucleotide probes used to detect 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 target sequences in a single reaction. In many of these examples, the fluorophore of the up to 100 sets of probes is the same. In some aspects, within each set of oligonucleotide probes the ratio of fluorophore bearing and non-fluorophore bearing oligonucleotides is set to a predetermined ratio. In some embodiments, the ratio within each set of oligonucleotides is different for each of the up to 100 sets of oligonucleotides. In these examples, the different ratios allow the reporter signal for each of the up to 100 targets to be distinguished from one another within a single reaction. In these examples, the total concentration of probes in each set of probes is the same, while the ratio of the two populations of probes within the probe set are different between each of the sets of probes. A fluorophore bearing probe can be a signal probe. A non-fluorophore bearing probe can be a competitive probe.

[036] In some embodiments greater than 100 oligonucleotide probes are used to detect greater than 100 target sequences in a single reaction. In some embodiments, greater than 100 oligonucleotide probes are used to detect up to 100 target sequences in a single reaction. In some embodiments, up to 100 oligonucleotide probes are used to detect greater than 100 target sequences in a single reaction.

[037] Disclosed herein are methods and compositions for using an unlabeled analog to the corresponding labeled entity that generates signal for detection in the presence of a particular analyte in any system. In some examples, the system is an ELISA (enzyme-linked immunosorbent assay), in which case the labeled entity is a labeled antibody. In these cases, the unlabeled analog is an unlabeled antibody. In some embodiments, the ratio of labeled and unlabeled antibodies within a target set are predetermined as discussed previously. In some embodiments, multiple targets can be detected in a single ELISA reaction using multiple target

sets of antibodies. In many of these examples, each set of target antibodies has a different ratio of labeled and unlabeled antibodies which allow the independent target reporting signal to be distinguished from the others. In these examples, the total concentration of antibodies in each set of antibodies is the same, while the ratio of the two populations of antibodies within the antibodies set are different between each of the sets of antibodies. A fluorophore bearing antibody can be a signal antibody. A non-fluorophore bearing antibody can be a competitive antibody.

[038] In some embodiments, the total number of molecules in a reaction remains constant despite that the ratio of functional to competitive entities changes between different sets. In some aspects, the total number of molecules refers to the proteins, nucleic acids, solution volume, other entities required for the reaction to take place. In some aspects the reaction is a qPCR, ELISA, PCR, or sequencing reaction. In some aspects, the entities are oligonucleotides, proteins, antibodies, qPCR probes, aptamers, nucleic acid or other appropriate molecule. In some aspects function al comprises performing its intended purpose, such as PCR extension or emitting of fluorescence signal. In some aspects, competitive comprises binding to the target molecule, whether that be a protein, nucleic acid, or another molecule, but not being able to perform the function of the function molecule, such as extend during PCR or emit a fluorescence signal when excited.

Antibody labeling

[039] ELISA is a standard analytical technique for identifying proteins using fluorescence-labelled antibody markers. Without being bound by theory, in a sandwich ELISA, two antibodies, a primary antibody and a fluorescently labeled secondary antibody, are used to bind and identify a protein of interest (antigen). In general, the two antibodies bind to distinct regions of the protein. The primary antibody is adhered to a surface or a reaction vessel, with its binding domain exposed to allow for protein capture. When a sample is introduced, the protein of interest will diffuse through the chamber and favorably bind to the primary antibody. After a wash procedure, the fluorescently labeled secondary antibody is introduced. This fluorescently labeled secondary antibody binds to a distinct epitope of the fixed antigen. After washing the excess, unbound secondary antibody, the sample is interrogated. The presence and intensity of the fluorescence signal correlates with the concentration of antigen that has been captured. However, the binding dynamics of the antibodies are not readily engineerable. Unlike nucleic acids, antibodies cannot be inorganically assembled and even simple changes to the antibody structure result in dramatic changes in binding efficiencies. This critically limits the efficiency of antibody multiplexing for ELISA.

[040] Disclosed herein are methods and compositions for using competitive, non-reporting probes for fluorescence reporting using antibodies. In some embodiments, the fluorescence

reporting is in the context of an ELISA assay. In some embodiments, a primary antibody is adhered to a surface or a reaction vessel and the protein of interest favorably binds to the primary antibody. In these cases, after a wash procedure, a fluorescently labeled secondary antibody is introduced and binds to a distinct epitope of the fixed antigen. In these cases, after a wash procedure, the sample is interrogated and the presence and intensity of the fluorescence signal correlates with the concentration of antigen that has been captured.

[041] In some embodiments, a set of secondary antibodies is introduced and binds to the fixed antigen. In some aspects, the antigen is fixed by being bound to a primary antibody, which is in turn adhered to a surface or a reaction vessel. In some aspects, the set of secondary antibodies comprises a labeled population and an unlabeled population. In some aspects, the secondary antibodies in the labeled and unlabeled populations are identical except for the presence or absence of the label. In some aspects, secondary antibodies being identical means they comprise the same protein sequence. In some aspects, identical means they bind to the same epitope. In some aspects, identical means they bind to the same antigen. In some embodiments, the label is a fluorescent label. In some embodiments, the ratio of labeled and unlabeled secondary antibodies within a set of secondary antibodies is set to a predetermined ratio. In some aspects, the ratio of labeled to unlabeled secondary antibodies correlated to a signal intensity. In some aspects, by measuring the signal intensity, the signal intensity can be correlated back to the ratio of labeled to unlabeled secondary antibodies.

[042] In some embodiments, the ratio of labeled to unlabeled secondary antibodies within a set of secondary antibodies is used to set the signal level for an ELISA immunoassay. In some aspects, the ratio is set without changing the sensitivity or the number of antigens captured. In some embodiments, the primary antibody determines capture sensitivity and sensitivity. In some embodiments, multiple distinct antigens can be detected in a single ELISA reaction using the methods disclosed herein. In some aspects, a set of secondary antibodies able to bind to each of the multiple antigens is used in the ELISA. In many cases, the same label is used to label the labeled secondary antibody within each of the multiple sets of secondary antibodies. In some aspects the ratio of labeled to unlabeled secondary antibodies is unique within each of the multiple sets of secondary antibodies. In many examples, the unique ratios correspond to unique signal intensities. In these and other examples, the unique signal intensities are measured and correlated back to the predetermined ratios, thereby identifying the set of secondary antibodies that generated the signal intensity. In this way, a specific antigen can be identified in an ELISA assay based on the unique signal generated by the unique ratio of labeled to unlabeled secondary antibodies within the unique set of secondary antibodies used to bind to the specific antigen. In many examples, these methods disclosed herein allow identification of multiple antigens despite

using the same label on the labeled secondary antibodies for each of the multiple antigens. This is achieved by using unique ratios of labeled to unlabeled secondary antibodies for each of the multiple antigens. The unique ratios correlate to unique signal intensities, thereby allowing identification of each of the multiple antigens.

Sequencing bias

[043] While genomic sequencing is poised to be the predominant clinical tool in the coming future, it has particularly susceptible to sample-preparation bias. Without being bound by theory, before nucleic acids are sequenced, they are enriched using highly multiplexed broad-binding PCR primers. However, this enrichment generates a large bias in the amplified targets. This bias results in ineffective sequencing. For applications such as cancer tissue genotyping and tumor heterogeneity assessment, this makes the technique inapplicable for many clinical use cases.

[044] Disclosed herein are methods and compositions for using competitive oligonucleotides to reduce sequencing bias. In some embodiments, sequencing bias is reduced by using a set of oligonucleotide primers during amplification. In some aspects, amplification comprises a polymerase chain reaction. In some embodiments, a set of primers comprises a population of functional primers and a population of competitive primers. In some aspects, functional primers comprise primers capable of annealing to a complementary sequence and being extended during a polymerase chain reaction. In some aspects, competitive primers comprise are identical in sequence to their corresponding functional primer, are able to bind to the same complementary sequence, but are not able to be extended during a polymerase chain reaction. In some embodiments, the ratio of functional and competitive primers is adjusted to alter the level of sequencing bias without changing the chemical interactivity of the primer sets themselves. In some embodiments, reducing or increasing the ratio of functional primers to competitive primers, increase or decrease the enrichment of DNA to be sequenced. In some aspects, this ratio adjustment is performed without affecting the enrichment of other portions of the genome. In some aspects, this ratio adjustment allows a fine control of PCR bias. In some aspects, this ratio adjustment allows for the compensation of the deficiencies of current PCR methods for sequencing prep.

[045] In some embodiments, the ratio of functional primers to competitive primers is increased. In some aspects, the ratio of functional primers to competitive primers is increased in order to increase the sequencing bias of a target sequence. In some aspects, altering sequencing bias comprises increasing the amplification level of the target sequence. In some aspects, increasing the amplification level of the target sequence leads to an increased level of the target sequence in the final sample to be sequenced.

[046] In some embodiments, the ratio of functional primers to competitive primers is decreased. In some aspects, the ratio of functional primers to competitive primers is decreased in order to decrease the sequencing bias of a target sequence. In some aspects, altering sequencing bias comprises decreasing the amplification level of the target sequence. In some aspects, decreasing the amplification level of the target sequence leads to an decreased level of the target sequence in the final sample to be sequenced.

[047] In some embodiments, competitive primers are not able to be extended. In some aspects, they are not able to be extended due to a competitive, capped, or otherwise modified end, such as those described and disclosed herein.

EXAMPLES

Example 1

[048] A Real-Time PCR reaction with an equimolar ratio of competitive probes and fluorogenic probes would yield approximately half the fluorescent signal at each reading as compared to a reaction with a similar amount of fluorogenic probes. Figure 2 shows the same TaqMan Assay with 4 differing proportions of fluorogenic probes to total probes. The 100% fluorescent probe reaction in yellow (solid top line) has an amplitude approximately 8 fold that of the assay in grey (long dashed line) which has 1 part fluorescent probes with 7 parts competitive probes. See Table 1 for the reaction make up and Table 2 for the PCR cycling parameters.

Table 1: Reaction components for Figure 2 reactions

Reaction Mix	Part #	12.5% fluorescent probe	25% fluorescent probe	50% fluorescent probe	100% fluorescent probe	No Template
ABI TaqMan Universal Master Mix	4304437	10 μ L	11 μ L	12 μ L	13 μ L	14 μ L
Ambion Nuclease Free Water	AM9937	4 μ L	4 μ L	4 μ L	4 μ L	4 μ L
Promega TE 1x (diluant)	V6231	6 μ L	6 μ L	6 μ L	6 μ L	6 μ L
Oligo (concentration in final reaction)						
Forward Primer 1	FBF2	200 nM	200 nM	200 nM	200 nM	200 nM
Reverse Primer 1	FBR2	200 nM	200 nM	200 nM	200 nM	200 nM
FAM Labeled Probe 1	FBP2-1	25 nM	25 nM	25 nM	25 nM	25 nM

Blunt Probe 1	FBP2-0	175 nM	175 nM	175 nM	175 nM	175 nM
Forward Primer 2	FAF1	200 nM	200 nM	200 nM	200 nM	200 nM
Reverse Primer 2	FAR1	200 nM	200 nM	200 nM	200 nM	200 nM
FAM Labeled Probe 2	FAP1-1	50 nM	50 nM	50 nM	50 nM	50 nM
Blunt Probe 2	FAP1-0	150 nM	150 nM	150 nM	150 nM	150 nM
Forward Primer 3	RBF1	200 nM	200 nM	200 nM	200 nM	200 nM
Reverse Primer 3	RBR1	200 nM	200 nM	200 nM	200 nM	200 nM
FAM Labeled Probe 3	RBP1-1	100 nM	100 nM	100 nM	100 nM	100 nM
Forward Primer 4	RBP1-0	100 nM	100 nM	100 nM	100 nM	100 nM
Reverse Primer 4	RAF2	200 nM	200 nM	200 nM	200 nM	200 nM
FAM Labeled Probe 4	RAR2	200 nM	200 nM	200 nM	200 nM	200 nM
Blunt Probe 4	RAP2-1	200 nM	200 nM	200 nM	200 nM	200 nM
template (copies per reaction)						
Template 1	FBT2	1000 copies	-	-	-	-
Template 2	FAT1	-	1000 copies	-	-	-
Template 3	RBT1	-	-	1000 copies	-	-
Template 4	RAT2	-	-	-	1000 copies	-
total volume [μ L]		20 μ L	20 μ L	20 μ L	20 μ L	20 μ L

Table 2: Temperature Schedule on ABI Viia7 for figure 2 reactions

Temperature Schedule:	Temperature [$^{\circ}$ C]	Time [s]	Data Collection
UNGase	50	120	N
Initial Denature	95	600	N
Number of Cycles	75		
Denature	95	15	N
Anneal	60	30	Y

Example 2

[049] A Real-Time PCR reaction is performed to detect three nucleic acid sequences of interest within a single reaction. In addition to the other components required for the RT-PCR reaction,

three sets of probes are added. One of each of the sets of probed corresponds to and is able to uniquely bind to one of the three transcripts of interest.

[050] Set 1 comprises 25% unlabeled competitive probes and 75% labeled probes, which are labeled with a green fluorophore and a quencher molecule. The probes within Set 1 are identical other than the presence of the fluorescent label and quencher molecule.

[051] Set 2 comprises 50% unlabeled competitive probes and 50% labeled probes, which are labeled with a green fluorophore and a quencher molecule. The probes within Set 2 are identical other than the presence of the fluorescent label and quencher molecule. The green fluorophore used in Set 2 is the same green fluorophore used in Set 1.

[052] Set 3 comprises 75% unlabeled competitive probes and 25% labeled probes, which are labeled with a green fluorophore and a quencher molecule. The probes within Set 3 are identical other than the presence of the fluorescent label and quencher molecule. The green fluorophore used in Set 3 is the same green fluorophore used in Set 1 and Set 2.

[053] The probes within each of Sets 1-3 are distinct from each other since they are selected for their ability to specifically bind to their corresponding nucleic acid target. However, the fluorophore used to label the labeled probes within each set is the same. Also, the ratio of unlabeled to labeled probes within each set are different and unique compared to the other sets.

[054] When the reaction first starts, no fluorescent signal is detected since the signals from each fluorophore are blocked by the quencher molecule. As the reaction proceeds, unlabeled and labeled probes within each set compete for binding to their complementary nucleic acid target sequence. Because the labeled and unlabeled probes within each set are basically identical, their binding and annealing kinetics are basically identical. When a labeled probe anneals, it is subsequently degraded by the 5' to 3' exonuclease activity of the polymerase. This degradation step separates the fluorophore and the quenching molecule, allowing for a fluorescence signal to be emitted upon excitation by the appropriate excitation source.

[055] An excitation source excites the green fluorophores and a signal is emitted from each of the released green fluorophores. Three distinct signal intensities are detected and measured. Each signal intensity correlated to a specific ratio. Because each set of probes comprised a unique ratio of labeled to unlabeled probes, a unique signal intensity can be correlated to that specific set of probes. Therefore, each unique signal intensity is used to identify one of each of the three nucleic acid targets.

Example 3

[056] Figures 3A and 3B demonstrate an example experiment exemplifying consistent leveling with competitive probes versus diluted probes.

[057] Ten RT-qPCR Master Mixes were made for a system containing three TaqMan™ assays in a single color channel for human rhinovirus, respiratory syncytial virus A (RSV A), and respiratory syncytial virus B (RSV B). All three assays utilized probes with the same fluorophore quencher combination. The total probe concentration of the mixtures was conserved for five of these reactions. Table 3 summarizes how the fluorescent and competitive probes were diluted to the same concentration. Table 4 summarizes how the primer sets were diluted to the same concentration. Table 5 displays the master mix composition for reactions containing both fluorescent and competitive probes. Table 5 displays the master mix composition for reactions containing fluorescent probes only with no competitive probes. Table 7 displays the concentration of the three templates included in each reaction.

[058] The fluorescent probe concentrations for two of the targets (RSV A and RSV B) were fixed but different. For the final target (human rhinovirus, RH), the fluorescent probe concentration was varied from 25nM. All reactions contained the same enzyme formulation and primer concentrations, as shown in Tables 5 and 6.

[059] The qPCR protocol for each of the 10 master mixes were tested with known single-stranded DNA templates at DNA templates with a fixed concentration of 10,000 copies per reaction (Table 7). The qPCR reactions were performed on a standard qPCR instrument with the cycling parameters shown in Table 8. The resulting data is depicted in Figure 3A and Figure 3B.

Table 3. Oligo mix

Oligo mix	Oligo	μ M	Oligo volume to make 10x stock	TE Buffer	Total Volume
RH 100% Fluorescent	RHP1-1	101.1	2.47	997.53	1000.0
RH 100% Competitive	RHP1-0	95.2	2.63	997.37	1000.0
RSV B 100% Fluorescent	RBP1-1	101.1	2.24	997.76	1000.0
RSV A 25% Fluorescent	RAP2-1	95.9	0.50	999.50	1000.0
RSV A 75% Competitive	RAP2-0	93.8	2.00	998.00	1000.0

Table 4. 6plex primer mix

Primer set	Material	μM	Desired μM in 20X stock	Oligo volume to make 10x stock (μL)
Set 1	cdcFAUniF1	93.1	0.4	17.19
	cdcFAUniR1	99.2	0.4	16.13
Set 2	ccFAH1_2009F2	117	0.4	13.68
	ccFAH1_2009R2	114	0.4	14.04
Set 3	cdcFAH3F3	114.6	0.4	13.96
	cdcFAH3R3	97.23	0.4	16.46
Set 4	RHF1	97.7	0.4	16.38
	RHR1	96.7	0.4	16.55
Set 5	RAF2	102.2	0.4	15.66
	RAR2	94.7	0.4	16.90
Set 6	RBF1	94.2	0.4	16.99
	RBR1	94.7	0.4	16.90
Set 7	RBF1	94.2	0.4	42.46
	RBR1	94.7	0.4	42.24
	Promega V6231, TE	n/a	n/a	9.2

Table 5. Reaction Master Mix – with competitive probes

Material	100 FP	80 FP + 20 CP	60 FP + 40 CP	40 FP + 60 CP	20 FP + 80 CP
TaqPath ABI A28626	60 μL	60	60	60	60
20X 6plex primer mix	12	12	12	12	12
10X RH 100% Fluorescent	24	19.2	14.4	9.6	4.8
10X RH Competitive	0	4.8	9.6	14.4	19.2
10X RSV B 100% Fluorescent	24	24	24	24	24
10X RSV A 25% Fluorescent	24	24	24	24	24
10X RSV A Competitive	24	24	24	24	24
water RxBiosciences	12	12	12	12	12

Table 6. Reaction Master Mix – without competitive probes

Material	100 FP - BP	80 FP	60 FP	40 FP	20 FP
TaqPath ABI A28626	60 μ L	60	60	60	60
20X 6plex primer mix	12	12	12	12	12
10X RH 100% Fluorescent	24	19.2	14.4	9.6	4.8
10X RSV B 100% Fluorescent	24	24	24	24	24
10X RSV A 25% Fluorescent	24	24	24	24	24
water RxBiosciences	12	12	12	12	12
TE Promega V6231	24	28.8	33.6	38.4	43.2

Table 7. Templates

Name	Concentration
RHT1	2E3 copies/ μ L
RAT2	2E3 copies/ μ L
RBT1	2E3 copies/ μ L
NTC TE	NA (buffer only)

Table 8

Cycling Parameters

Step	time(seconds)	temperature($^{\circ}$ C)	
1	120	25	
2	600	53	
3	120	95	
4	3	95	
5	30	60	detection
	repeat step 4 & 5 90 cycles		

Example 4

[060] The following example experiments and resulting data displayed in Figure 4A and Figure 4B demonstrate the ability to deterministically generate a 3plex assay with the same

reporting fluorophore and quencher combination with the binary separation based upon initial 100% fluorescent probe data.

[061] To generate the initial fluorescence data a RT-qPCR Master Mix was made containing three TaqMan™ assays for human rhinovirus (Target A), respiratory syncytial virus A (Target B) and respiratory syncytial virus B (Target C). No non-fluorescent, competitive probes were used in these assays. All three assays were labeled with the same fluorophore quencher combination. The primer concentration was selected to be in excess of the total probe concentration. The same concentration of fluorogenic probe, and therefore total probe, was used for each target.

[062] Table 9 displays the composition of the oligo master mix, which contained the indicated fluorescent probes and no competitive probes. Table 10 displays the composition of the primer master mix. Table 11 displays the master mix composition for reactions containing only 100% fluorescent probes and no competitive probes. Table 12 displays the concentration of the three templates included in each reaction. Table 13 displays the cycling parameters used for the RT-qPCR reaction.

Table 9. 25X fluorescence probe Oligo Mix

Oligo Mix	Oligo	Stock Conc, μM	Conc. of 1X, μM	Vol, μL
Target A RH 100% Fluorescent	RHP1-1	79.9	0.025	3.13
Target B RSV A 100% Fluorescent	RAP2-1	95.9	0.025	2.61
Target C RSV B 100% Fluorescent	RBP1-1	101.1	0.025	2.47
	Promega V6231, TE	n/a	n/a	391.79
	Total			400.00

Table 10. 25X Primer Mix

Primer set	Name	Stock Conc, μM	Conc of 1X, μM	Vol, μL
Target A set	RHF1	97.7	0.1	10.24
	RHR1	96.7	0.1	10.34
Target B set	RAF2	102.2	0.1	9.78
	RAR2	94.7	0.1	10.56
Target C set	RBF1	94.2	0.1	10.62
	RBR1	94.7	0.1	10.56
	Promega V6231, TE	n/a	n/a	337.90
	Total			400.00

Table 11. Reaction Master Mix

Name	1 rxn	24 rxns
4x TaqPath with MP	5 μL	120
25X Primer mix	0.8	19.2
25X Fluorescent probe oligo mix	0.8	19.2
water	8.4	201.6

Table 12. Templates

Target	Name	Conc
Target A	RHT1	2E3 copies/ μL
Target B	RAT2	2E3 copies/ μL
Target C	RBT1	2E3 copies/ μL
No template control	NTC TE	NA

Table 13. Cycling parameters

Step	time(seconds)	temperature(°C)	
1	120	25	
2	600	53	
3	120	95	
4	3	95	
5	30	60	detection
	repeat step 4 & 5 90 cycles		

[063] The qPCR protocol for each of the 10 master mixes were tested with known single-stranded DNA templates at DNA templates with a fixed concentration of 10,000 copies per reaction (Table 12). The qPCR reactions were performed on a standard qPCR instrument with the parameters displayed in Table 13. The results from this experiment are displayed in Table 14 and in Figure 4A. Six replicates were performed along with a no template control (NTC).

Table 14. Data from RT-qPCR with no competitive probes

Target	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5	Replicate 6	NTC
Target A	0.4016	0.4028	0.3967	0.4018	0.4057	0.4056	0.0631
Target B	0.5291	0.5235	0.4415	0.5147	0.5324	0.5331	NA
Target C	0.4535	0.4415	0.4449	0.4445	0.4438	0.4381	0.0600

Table 15. Fluorescence score for experiment without competitive probes

Target	Name	Score
Target A	RH T1	0.4023
Target B	RSV A T2	0.5263
Target C	RSV B T1	0.4441

[064] To calculate the desired ratio of fluorogenic probe to completeive probe for a leveled mix, the median fluorescence level from the 6 replicates was determined from the data from the fluorescent probe only experiment (Table 15). Target C was set to the highest intensity level and left at 100% of the initial concentration of fluorogenic probe. The levels for Target A and Target B were engineered to be distinct by replacing a proportion of the fluorescent probe with a non-fluorescent, competitive probe. The total amount of probe (fluorescent plus non-fluorescent, competitive) was the same for every target and identical to the total probe concentrations for the experiments shown in Figure 4A.

[065] Table 16 displays the composition of the oligo master mix, which contained the indicated fluorescent probes and competitive probes. The primer master mix and reaction master mix used was the same as those displayed in Table 10 and Table 11 respectively. The templates and cycling parameters used were the same as those displayed in Tables 12 and Table 13 respectively.

[066] Six replicates along with a no template control (NTC) were performed. The resulting data is displayed in Table 17 and Figure 4B. The fluorescence score calculated for each target is displayed in Table 18.

Table 16. Fluorescent and Competitive probe oligo mix

Probe	Name	Stock Conc, μM	Conc of 1X, μM	Vol, μL
Target A - Fluorescent	RHP1-1	53.7	0.0125	4.67
Target A - Competitive	RHP1-0	90.2	0.0125	2.76
Target B - Fluorescent	RAP2-1	9.59	0.0044	9.11
Target B - Competitive	RAP2-0	93.8	0.0206	4.40
Target C - Fluorescent	RBP1-1	101.1	0.0250	4.95
	Promega V6231, TE	n/a	n/a	774.11
	Total			800.00

Table 17. Data from RT-qPCR with competitive probes

Target	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5	Replicate 6	NTC
Target A	0.2154	0.2121	0.2072	0.2129	0.2142	0.2161	0.0369
Target B	0.1192	0.1174	0.1145	0.1169	0.1129	0.1214	0.0359
Target C	0.4323	0.4247	0.4195	0.4190	0.4217	0.4243	0.0374

Table 18. Fluorescence score for experiment with competitive probes

Target	Name	Score
Target A	RH T1	0.2135
Target B	RSV A T2	0.1171
Target C	RSV B T1	0.4230

Example 5

[067] An ELISA assay is performed to detect three proteins of interest within a single reaction vessel. The sets of primary antibodies are adhered to a surface of a reaction vessel. One of each

sets is able to uniquely bind to one of each of the three proteins of interest. A sample comprising the three proteins of interest is added to the reaction vessel and each favorably binds to the corresponding primary antibody. A wash procedure is performed to removed unbound sample. Next, three sets of secondary antibodies are added to the reaction vessel, one of the sets corresponding to each of the three proteins of interest.

[068] Set 1 comprises 25% unlabeled competitive secondary antibodies and 75% labeled secondary antibodies, which are labeled with a green fluorophore. The secondary antibodies in Set 1 are identical other than the presence of the fluorescent label.

[069] Set 2 comprises 50% unlabeled competitive secondary antibodies and 50% labeled secondary antibodies, which are labeled with a green fluorophore. The secondary antibodies in Set 2 are identical other than the presence of the fluorescent label. The green fluorophore used in Set 2 is the same green fluorophore used in Set 1.

[070] Set 3 comprises 75% unlabeled competitive secondary antibodies and 25% labeled secondary antibodies, which are labeled with a green fluorophore. The secondary antibodies in Set 3 are identical other than the presence of the fluorescent label. The green fluorophore used in Set 3 is the same green fluorophore used in Set 1 and Set 2.

[071] The secondary antibodies within each of Sets 1-3 are distinct from each other since they are selected for their ability to specifically bind to their corresponding protein of interest. However, the fluorophore used to label the labeled secondary antibodies within each set is the same. Also, the ratio of unlabeled to labeled secondary antibodies within each set are different and unique compared to the other sets.

[072] The three sets of secondary antibodies are added to the reaction vessel. Labeled and unlabeled secondary antibodies from each set compete for binding to their corresponding protein epitope. Because the labeled and unlabeled secondary antibodies within each set are identical other than the presence of the fluorophore, their binding kinetics are basically the same. After the binding step, a wash procedure is performed to remove unbound secondary antibody.

[073] An excitation source excites the green fluorophores and a signal is emitted from each of the bound labeled secondary antibodies. Three distinct signal intensities are detected and measured. Each signal intensity correlated to a specific ratio. Because each set of secondary antibodies comprised a unique ratio of labeled to unlabeled secondary antibodies, a unique signal intensity can be correlated to that specific set of secondary antibodies. Therefore, each unique signal intensity is used to identify one of each of the three proteins of interest.

Example 6

[074] mRNA is isolated from a sample following an experiment and processed for sequencing. The mRNA sample reverse transcribed to cDNA and then PCR amplified in order to prepare the

sequencing library. Sequencing is performed to detect a target of interest and to determine how the target level has changed due to the conditions of the experiment compared to a control gene. Unfortunately, the target is from a very highly expressed gene and is over-represented in the sequencing library compared to the control gene. This makes it difficult to accurately calculate the impact of the experimental conditions on the expression level of the gene.

[075] The frustrated scientists go back to the original cDNA sample and re-amplify it by PCR. However, this time, they include a different set of target specific primers in the PCR reaction mix. The set of target primers comprises 25% functional primers capable of being extended during PCR, and 75% of competitive primers which cannot be extended due to a capped end.

[076] The resulting sequencing results has a lower level of the first transcript, which allows for a more accurate calculation of the transcription level caused by the conditions used in the original experiment.

CLAIMS

What is claimed is:

1. Method of identifying a first target nucleic acid comprising,
 - providing a sample comprising the first target nucleic acid,
 - providing a first set of paired oligonucleotides with complementarity to the first target nucleic acid, said first set of paired oligonucleotides comprising a first ratio of (a) first competitive oligonucleotides to (b) first signal oligonucleotides comprising a first signal tag,
 - wherein the first competitive oligonucleotides compete with the first signal oligonucleotides for binding to the first target nucleic acid,
 - amplifying the first target nucleic acid with the polymerase chain reaction, thereby degrading the first signal oligonucleotide and permitting generation of a first signal,
 - generating the first signal,
 - measuring intensity of the first signal, and
 - correlating the intensity of the first signal to the first ratio, thereby identifying the first target nucleic acid.
2. The method of claim 1, wherein the sample comprises a second target nucleic acid.
3. The method of claim 2, further comprising identifying the second target nucleic acid by:
 - providing a second set of paired oligonucleotides with complementarity to the second target nucleic acid, said second set of paired oligonucleotides comprising a second ratio of (a) second competitive oligonucleotides to (b) second signal oligonucleotides comprising a second signal tag,
 - wherein the second competitive oligonucleotides compete with the second signal oligonucleotides for binding to the second target nucleic acid,
 - amplifying the second target sequence with the polymerase chain reaction, thereby degrading the second signal oligonucleotide and permitting generation of a second signal,
 - generating the second signal,
 - measuring intensity of the second signal, and
 - correlating the intensity of the second signal to the second ratio, thereby identifying the second target nucleic acid.
4. The method of claim 3, wherein the first signal oligonucleotides and the second signal oligonucleotides comprise the same signal tag.

5. The method of any of claims 3-4, wherein the first target nucleic acid and the second target nucleic acid are identified from differing signal intensities generated during the same polymerase chain reaction.
6. The method of any of claims 3-5, wherein the first signal intensity is different from the second signal intensity.
7. The method of claim 1, wherein the paired oligonucleotides comprise qPCR probes.
8. The method of any of claims 1-7, wherein the signal oligonucleotides comprise a signal tag and a signal quenching tag.
9. The method of claim 8, wherein the signal tag comprises a fluorophore.
10. The method of and of claims 8-9, wherein the signal quenching tag blocks the generation of the signal corresponding with the signal tag until the signal quenching tag is separated from the signal tag by oligonucleotide degradation during the polymerase chain reaction, thereby permitting the signal to be generated.
11. The method of any of claims 1-10, wherein generating the signal comprises exciting the signal tag.
12. The method of claim 11, wherein exciting comprises directing at the signal tag at least one of illumination, electromagnetic radiation, and energy emitted from an energy source.
13. The method of any of claims 1-12 wherein the competitive oligonucleotides do not comprise a signal tag.
14. The method of any of claims 1-13, wherein the competitive oligonucleotides comprise at least one modified end.
15. The method of claim 14, wherein the modified end prevents enzymatic extension during amplification.
16. The method of claim 15, wherein the modified end comprises a 3' end lacking a hydroxyl group.
17. The method of claim 15, wherein the modified end comprises a 5' end lacking a phosphate group.
18. The method of claim 15, wherein the modified end comprises a 5' capped end.
19. The method of claim 15, wherein the modified end comprises a 3' capped end.
20. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:1.
21. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:2.

22. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:3.
23. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:4.
24. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:5.
25. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:6.
26. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:7.
27. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:8.
28. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:9.
29. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:10.
30. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:16.
31. The method of any of claims 1-19, wherein the ratio of signal oligonucleotides to competitive oligonucleotides or the ratio of competitive oligonucleotides to signal oligonucleotides is about 1:20.
32. The method of any of claims 1-19, wherein the paired oligonucleotides comprise only signal oligonucleotides or only competitive oligonucleotides.
33. The method of any of the claims 1-32, wherein the total concentration of the first set of paired oligonucleotides is the same as the total concentration of the second set of paired oligonucleotides.

34. The method of any of claims 1-33, further comprising identifying up to thirty target nucleic acid sequences using up to thirty sets of paired oligonucleotides,
- wherein each set of paired oligonucleotides possesses complementarity to one of the up to thirty target nucleic acid sequences,
 - wherein each of the up to thirty sets of paired oligonucleotides comprises a unique ratio of (a) competitive oligonucleotides to (b) signal oligonucleotides comprising a signal tag,
 - wherein the competitive oligonucleotides within each set compete with the signal oligonucleotides within the same set for binding to a corresponding target nucleic acid sequence,
 - amplifying the up to thirty target nucleic acid sequences with the polymerase chain reaction, thereby degrading the signal oligonucleotides and permitting generation of up to thirty signals,
 - generating up to thirty signals,
 - measuring the intensities of the up to thirty signals, and
 - correlating the intensities of the up to thirty signals to the up to thirty ratios,
- thereby identifying the up to thirty target nucleic acid sequences.
35. The method of claim 34, wherein the signal tag of the signal oligonucleotides within each of the up to thirty sets of paired oligonucleotides are the same.
36. The method of any of claims 34-35, wherein the up to thirty target nucleic acid sequences are identified from signal intensities generated during the same polymerase chain reaction.
37. The method of any of claims 34-36, wherein each of the up to thirty ratios differs from every other of the up to thirty ratios.
38. The method of any of claims 34-37, wherein each of the up to thirty signal intensities differs from every other of the up to thirty signal intensities.
39. The method of any of claims 1-38, wherein the polymerase chain reaction is carried out by use of a nucleic acid polymerase having 5' to 3' exonuclease activity, an exo-polymerase, or a molecular beacon.
40. The method of any of claims 34-39, wherein the up to five signals are measured in a single fluorescence channel.
41. The method of any of the claims 34-40, wherein the total concentration of each of the up to five sets of paired oligonucleotides is the same.
42. The method of any of claims 1-41, wherein a total number of molecules in the reaction remains constant

43. A composition comprising:
- a first set of paired oligonucleotides with complementarity to a first target nucleic acid,
 - said first set of paired oligonucleotides comprising a first ratio of (a) competitive oligonucleotides to (b) signal oligonucleotides comprising a signal tag,
 - wherein the competitive oligonucleotides are configured to compete with the signal oligonucleotides for binding to the first target nucleic acid during amplification of the first target nucleic acid with the polymerase chain reaction,
 - wherein the signal oligonucleotides are adapted to generate a first signal having a first intensity when degraded by a nucleic acid polymerase during the polymerase chain reaction and excited by an energy source,
 - and wherein the first intensity correlates with the first ratio.
44. Method of identifying a first target molecule comprising,
- providing a sample comprising the first target molecule,
 - providing a first set of paired capture molecules that bind to the first target molecule, said first set of paired capture molecules comprising a first ratio of (a) first competitive capture molecules to (b) first signal capture molecules comprising a signal tag,
 - wherein the competitive capture molecules compete with the signal capture molecules for binding to the first target molecule,
 - generating the first signal,
 - measuring intensity of the first signal, and
 - correlating the intensity of the first signal to the first ratio, thereby identifying the first target molecule.
45. The method of claim 44, wherein the sample comprises a second target molecule.
46. The method of claim 45, further comprising identifying the second target molecule by:
- providing a second set of paired capture molecules that binds to the second target molecule, said second set of paired capture molecules comprising a second ratio of (a) second competitive capture molecules to (b) second signal capture molecules comprising the signal tag,
 - wherein the second competitive capture molecules compete with the second signal capture molecules for binding to the second target molecule,
 - generating the second signal,
 - measuring intensity of the second signal, and

correlating the intensity of the second signal to the second ratio, thereby identifying the second target molecule.

47. The method of claim 46, wherein the first signal capture molecules and the second signal capture molecules comprise the same signal tag.
48. The method of any of claims 45-47, wherein the first target capture molecules and the second target capture molecules are identified from differing signal intensities generated.
49. The method of any of claims 45-48, wherein the first signal intensity is different from the second signal intensity.
50. The method of any of claims 44-49, wherein the signal tag comprises a fluorophore.
51. The method of any of claims 44-50, wherein generating the signal comprises exciting the signal tag.
52. The method of claim 51, wherein exciting comprises directing at the signal tag at least one of illumination, electromagnetic radiation, and energy emitted from an energy source.
53. The method of any of claims 44-52 wherein the competitive capture molecules do not comprise a signaling tag.
54. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:1.
55. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:2.
56. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:3.
57. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:4.
58. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:5.
59. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:6.

60. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:7.
61. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:8.
62. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:9.
63. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:10.
64. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:16.
65. The method of any of claims 44-53, wherein the ratio of signal capture molecules to competitive capture molecules or the ratio of competitive capture molecules to signal capture molecules is about 1:20.
66. The method of any of claims 44-53, wherein the first paired capture molecules comprise only signal capture molecules or only competitive capture molecules.
67. The method of any of claims 44-66, further comprising identifying up to one hundred target molecules using up to one hundred sets of paired capture molecules,
wherein each set of paired capture molecules that binds to one of the up to one hundred target molecules,
wherein each of the up to one hundred sets of paired capture molecules comprises a unique ratio of (a) competitive capture molecules to (b) signal capture molecules comprising a signal tag,
wherein the competitive capture molecules within each set compete with the signal capture molecules within the same set for binding to a corresponding target molecule,
generating up to one hundred signals,
measuring the intensities of the up to one hundred signals, and
correlating the intensities of the up to one hundred signals to the up to one hundred ratios, thereby identifying the up to one hundred target molecules.

68. The method of claim 67, wherein the signal tag of the signal capture molecules within each of the up to one hundred sets of capture molecules are the same.
69. The method of any of claims 67-68, wherein the up to one hundred target molecules are identified from signal intensities generated at the same time.
70. The method of any of claims 67-69, wherein each of the up to one hundred ratios differs from every other of the up to one hundred ratios.
71. The method of any of claims 67-70, wherein each of the up to one hundred signal intensities differs from every other of the up to one hundred signal intensities.
72. The method of any of claims 67-71, wherein the up to one hundred signals are measured in a single fluorescence channel.
73. The method of any of claims 44-72, wherein the target molecule is fixed to a surface.
74. The method of any of claims 44-73, wherein the capture molecule is fixed to a surface.
75. The method of any of claims 44-74, wherein the target molecule comprises a protein or a functional fragment thereof.
76. The method of any of claims 44-74, wherein the target molecule comprises an antigen.
77. The method of any of claims 44-74, wherein the target molecule comprises an aptamer.
78. The method of any of claims 44-74, wherein the target molecule comprises a nucleic acid.
79. The method of any of claims 44-78, wherein the capture molecule comprises a protein a protein or a functional fragment thereof.
80. The method of any of claims 44-78, wherein the capture molecule comprises an antibody a protein or a functional fragment thereof.
81. The method of any of claims 44-78, wherein the capture molecule comprises an aptamer.
82. The method of any of claims 44-81, wherein a total number of molecules in the reaction remains constant.
83. A composition comprising:
a first set of paired capture molecules with complementarity to a first target molecule,
said first set of paired capture molecules comprising a first ratio of (a) competitive capture molecules to (b) signal capture molecules comprising a signal tag, wherein the competitive capture molecules are configured to compete with the signal capture molecules for binding to the target molecule,
wherein the signal capture molecules are adapted to generate a first signal having a first intensity when excited by an energy source,
and wherein the first intensity correlates with the first ratio.

84. The composition of claim 83, wherein the target molecule comprises a protein or a functional fragment thereof.
85. The composition of claim 83, wherein the target molecule comprises a antigen.
86. The composition of claim 83, wherein the target molecule comprises a aptamer.
87. The composition of claim 83, wherein the target molecule comprises a nucleic acid.
88. The composition of any of claims 83-87, wherein the capture molecule comprises a protein or a functional fragment thereof.
89. The composition of any of claims 83-87, wherein the capture molecule comprises an antibody or a functional fragment thereof.
90. The composition of any of claims 83-87, wherein the capture molecule comprises an aptamer.
91. Method of altering amplification level of a target nucleic acid comprising,
providing a sample comprising the target nucleic acid,
providing a set of paired oligonucleotides with complementarity to the target nucleic acid, said set of paired oligonucleotides comprising a ratio of (a) competitive oligonucleotides to (b) functional oligonucleotides,
wherein the competitive oligonucleotides compete with the functional oligonucleotides for binding to the target nucleic acid and are not extendible by a nucleic acid polymerase, wherein the functional oligonucleotides are extendible by a nucleic acid polymerase as primers for amplifying the target nucleic acid by the polymerase chain reaction (PCR),
and
amplifying the target nucleic acid with PCR.
92. The method of claim 91, wherein the competitive oligonucleotides comprise at least one modified end.
93. The method of claim 92, wherein the modified end prevents enzymatic extension during amplification.
94. The method of claim 93, wherein the modified end comprises a 3' end lacking a hydroxyl group.
95. The method of claim 93, wherein the modified end comprises a 5' end lacking a phosphate group.
96. The method of claim 93, wherein the modified end comprises a 5' capped end.
97. The method of claim 93, wherein the modified end comprises a 3' capped end.
98. The method of any of claims 91-97, wherein a total number of molecules in the reaction remains constant.

99. A composition comprising:
- a set of paired oligonucleotides with complementarity to a target nucleic acid,
 - said set of paired oligonucleotides comprising a ratio of (a) competitive oligonucleotides to (b) functional oligonucleotides,
 - wherein the competitive oligonucleotides are configured to compete with the functional oligonucleotides for binding to the nucleic acid target during amplification of the first target nucleic acid with the polymerase chain reaction and are not extendible by a nucleic acid polymerase.

FIG. 1

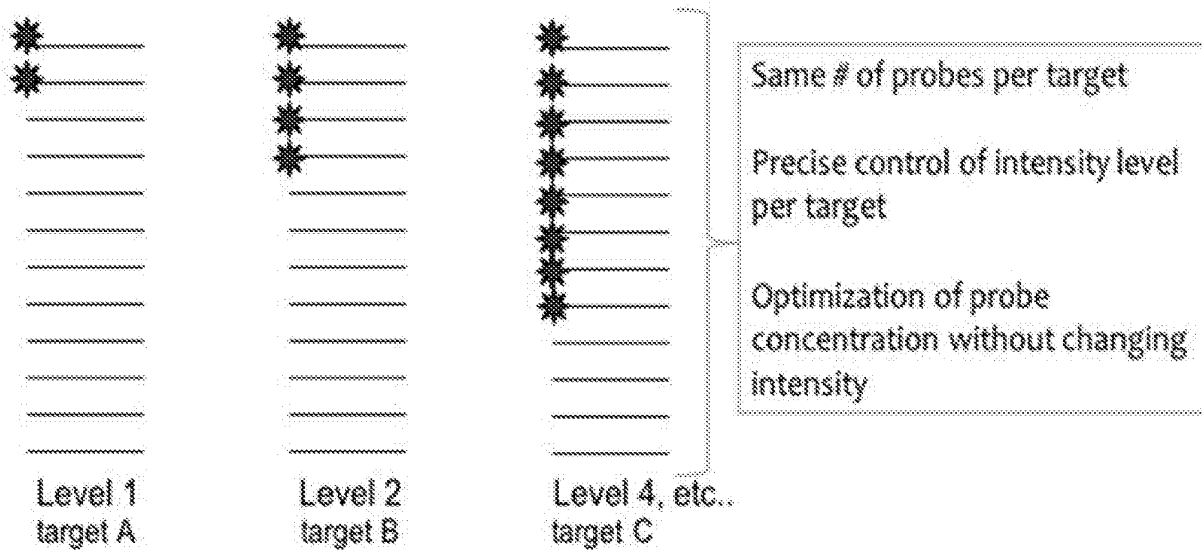


FIG. 2

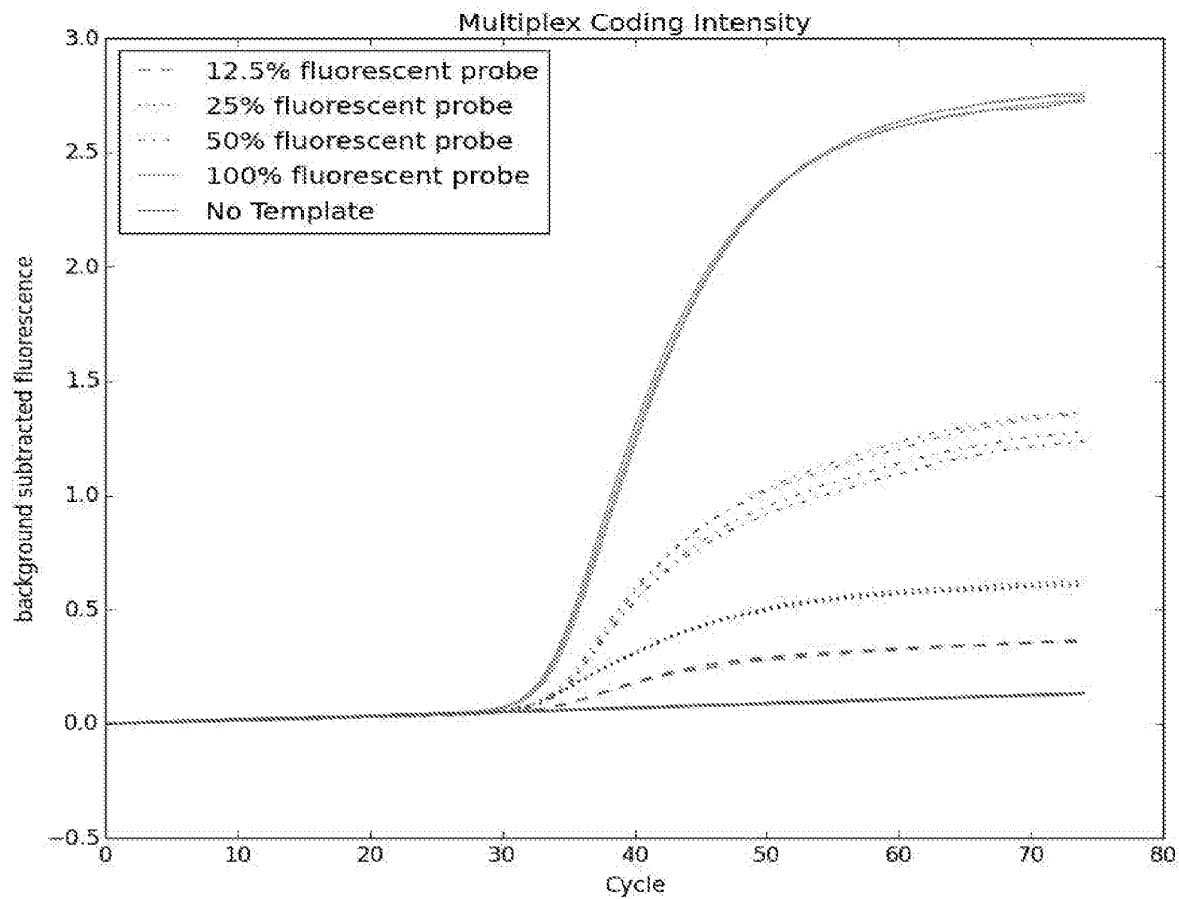


FIG. 3A

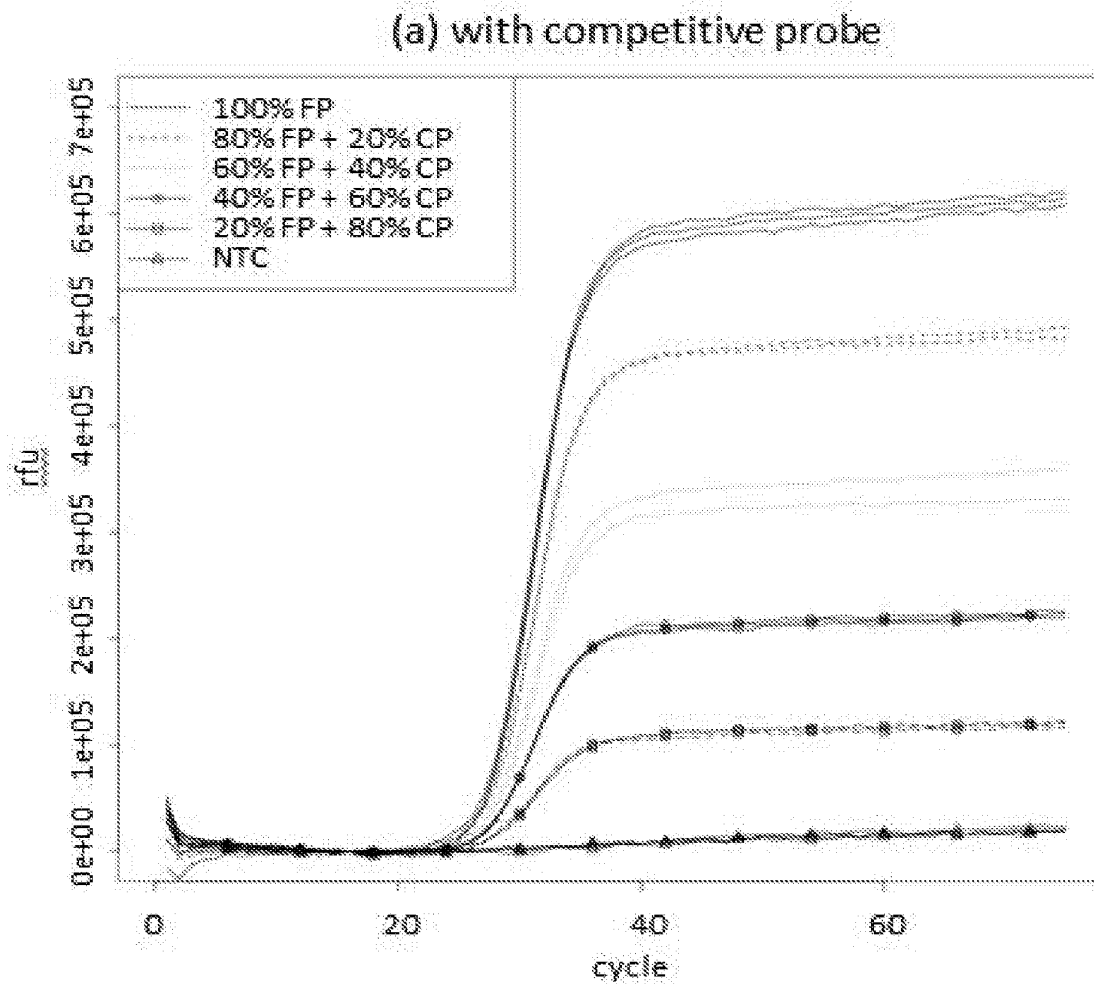


FIG. 3B

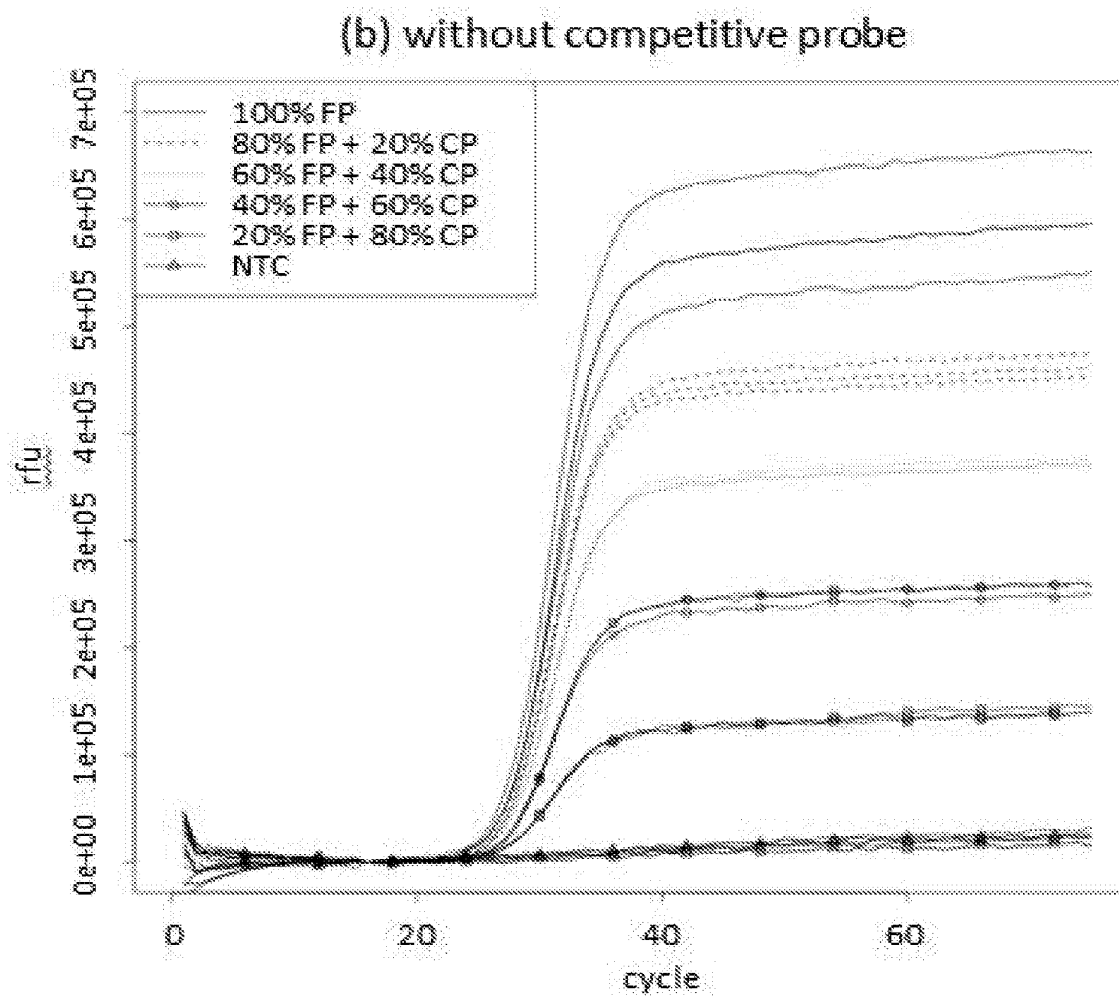


FIG. 4A

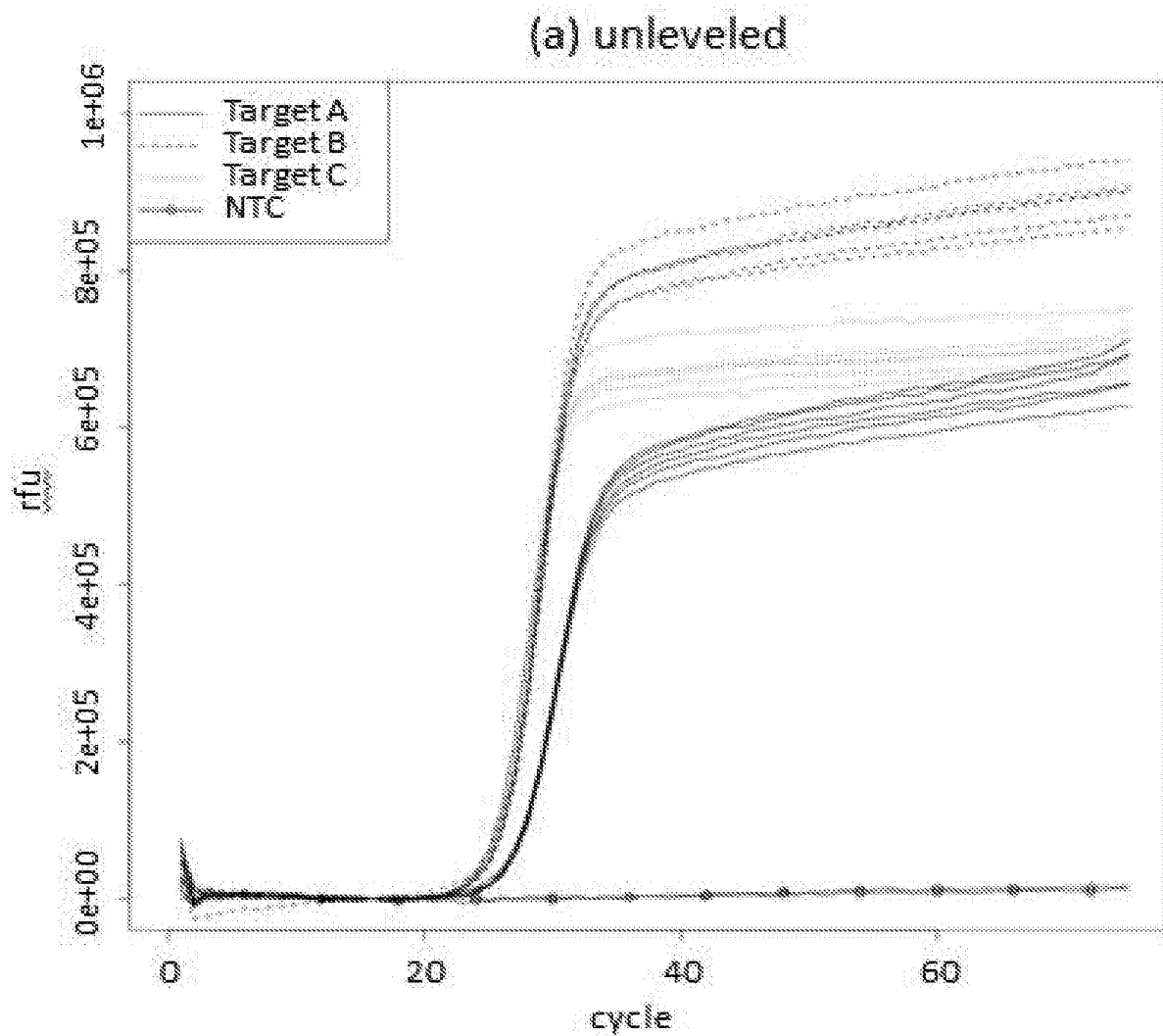
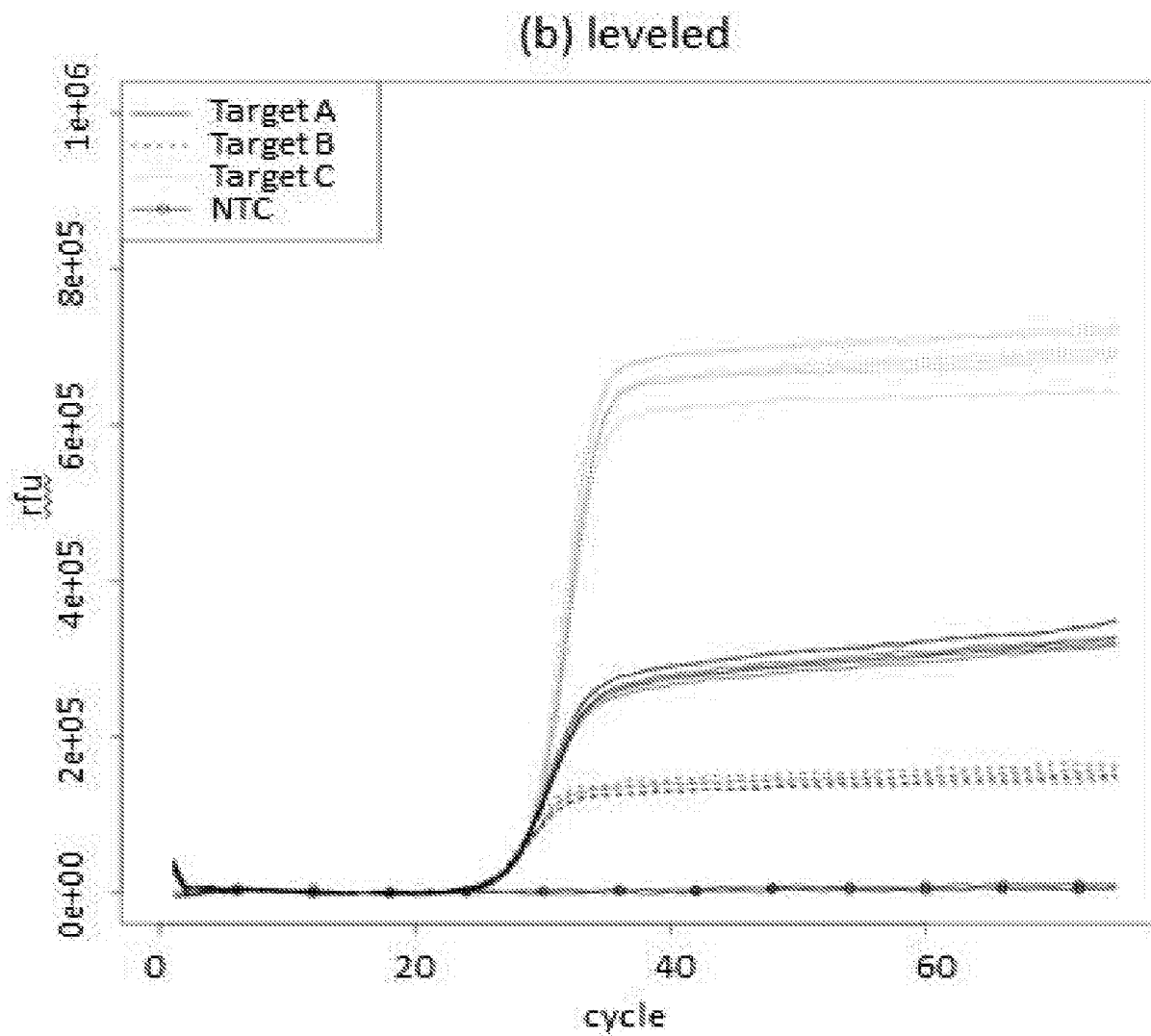


FIG. 4B



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/24933

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-42 and 49-82
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/24933

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C07H 21/00, 21/04; C12Q 1/68; G01N 21/64, 33/533, 33/542 (2017.01)

CPC - C07H 21/00, 21/04; C12Q 1/68, 1/686, 1/6806, 1/6851, 2561/101, 2561/113; G01N 21/64, 33/533, 33/542

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2006/079049 A2 (THIRD WAVE TECHNOLOGIES, INC.) 27 July 2006; page 2, lines 10-17 and 23-31; page 3, lines 13-20, page 5, lines 5-8 and 23-30; page 6, lines 18-30; page 7, lines 14-19; page 15, lines 24-31; page 16, lines 1-3 and 18-28	1-4, 5/3-4, 43-47, 48/45-47, 83-87, 88/83-87, 89/83-87, 90/83-87, 91-97, 98/91-97, 99
A	US 2009/0042735 A1 (BLAIR SM, et al.) 12 February 2009; paragraphs [0018], [0035], [0041]-[0043], [0072], [0074]	1-4, 5/1-3, 43-47, 48/45-47, 83-87, 88/83-87, 89/83-87, 90/83-87, 91-97, 98/91-97, 99
A	US 2014/0213471 A1 (CALIFORNIA INSTITUTE OF TECHNOLOGY) 31 July 2014; paragraphs [0003], [0010]-[0011], [0045], [0082]	1-4, 5/1-3, 43-47, 48/45-47, 83-87, 88/83-87, 89/83-87, 90/83-87, 91-97, 98/91-97, 99
A	US 2012/0040349 A1 (LODE PV, et al.) 16 February 2012; paragraphs [0030]-[0034], [0054]	1-4, 5/1-3, 43-47, 48/45-47, 83-87, 88/83-87, 89/83-87, 90/83-87, 91-97, 98/91-97, 99

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

3 June 2017 (03.06.2017)

Date of mailing of the international search report

22 JUN 2017

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

NUCLEIC ACID REACTIONS AND RELATED METHODS AND COMPOSITIONS**CROSS-REFERENCE**

This application claims the benefit of US Provisional Application No. 62/351,411, filed June 17, 2016, which is incorporated by reference herein in its entirety.

BACKGROUND

[0001] The polymerase chain reaction is a technique that is used for increasing the quantity of a nucleic acid target. By thermal cycling a reaction cocktail of short primer sequences, free nucleotides and DNA polymerase, a template strand can be copied. For the polymerase chain reaction (PCR), temperature plateaus are often required: a high temperature for denaturation of the amplicon of interest, a low temperature for the binding of primers to the template, and an intermediate temperature for the synthesis of a complement strand.

[0002] Specificity of the binding between the primers and the template allows for selective amplification of intended targets. For instance, a typical quantitative polymerase chain reaction is run to “saturation,” wherein the amplification reaction is run to the point at which at least one of the reactants such as primers, polymerase, probe, or free nucleotides is exhausted. When fluorescent reporters such as TaqMan probes, FRET probes, or intercalating dyes are used to interpret the extent of the DNA amplification, a fluorescence curve could be generated. The sigmoidal signature of a “saturation” fluorescence curve often occurs when the reporter probe concentrations are depleted at an ever-increasing rate.

[0003] However, this reaction (and by extension, its signature fluorescence) is limited by all reactant concentrations, including the PCR primers. As an example, when a single set of primers is used in equal concentrations, and when one primer is more efficiently extended than the other, the reaction becomes rate limited. This results in a linear, as compared to an exponential, amplification of DNA.

SUMMARY OF THE INVENTION

[0004] Described herein, in some embodiments, are methods of detecting the presence or absence of a target nucleic acid analyte in a sample, the methods comprising a) performing an amplification reaction on the sample; b) measuring a signal generated during the amplification reaction and generating a kinetic signature from the measured signal; and c) comparing the generated kinetic signature to a reference signature and determining whether

the generated kinetic signature corresponds with the reference signature, thereby detecting the presence or absence of the target nucleic acid analyte in the sample. In some embodiments, kinetic signature comprises a signal selected from an electromagnetic signal, a wavelength, and a fluorescence emission signal. In some embodiments, the kinetic signature is plotted as a curve. In some embodiments, the curve comprises at least one of an exponential region, a linear region, and a plateau region. In some embodiments, the kinetic signature comprises recorded values. In some embodiments, the values are recorded in a table. In some embodiments, the amplification reaction is a polymerase chain reaction (PCR). In some embodiments, the reference signature corresponds with a kinetic signature generated by amplification of the target nucleic acid analyte under one or more members of a set of constrained amplification parameters. In some embodiments, the set of constrained amplification parameters comprises any combination of: (i) a range of forward and/or reverse primer concentrations; (ii) a range of polymerase concentrations; (iii) polymerase type; (iv) a range of nucleotide concentrations; (v) chromophore concentration; and (vi) chromophore type; (vii) a number of thermocycles; (viii) rate of thermocycling; (ix) one or more thermocycle stage temperatures; and (x) one or more thermocycle stage time lengths; (xi) a range of target-specific molecular probes; (xii) primers associated with one or more target-specific probes; and (xiii) a concentration of one or more enzyme cofactors. In some embodiments, the amplification reaction is carried out within one or more of the constrained amplification parameters.

[0005] Described herein, in some embodiments, are methods of detecting the presence or absence of a target nucleic acid analyte in a sample, the method comprising a) performing an amplification reaction on the sample under a set of constrained amplification parameters, the set comprising any combination of one or more of (i) a range of forward and/or reverse primer concentrations; (ii) a range of polymerase concentrations; (iii) polymerase type; (iv) a range of nucleotide concentrations; (v) chromophore concentration; (vi) chromophore type; (vii) number of thermocycles; (viii) rate of thermocycling; (ix) one or more thermocycle stage temperatures; and (x) one or more thermocycler stage time lengths; (xi) a range of target-specific molecular probes; and (xii) primers associated with one or more target-specific probes; and (xiii) a concentration of one or more enzyme cofactors; b) measuring a signal generated during the amplification reaction and generating a kinetic signature from the measured signal; c) comparing the generated kinetic signature to a reference signature, the reference signature corresponding with amplification of the target nucleic acid analyte under

the set of constrained amplification parameters; and d) determining whether the generated kinetic signature corresponds with the reference signature, thereby detecting the presence or absence of the target nucleic acid analyte in the sample. In some embodiments, the amplification reaction is carried out within one or more of the constrained amplification parameters. In some embodiments, the one or more enzyme cofactors is selected from the group consisting of magnesium, potassium, and sodium. In some embodiments, the nucleic acid analyte is from a source selected from the group consisting of an animal, a plant, a bacterium, a parasite, and a virus. In some embodiments, the animal is a human. In some embodiments, the nucleic acid analyte is selected from the group consisting of human immunodeficiency virus, herpes simplex virus, human papilloma virus, Plasmodium, Mycobacterium, dengue virus, hepatitis virus, and influenza virus. In some embodiments, the nucleic acid analyte is selected from the group consisting of human immunodeficiency virus polyprotease, human immunodeficiency virus p17, human papilloma virus E6, and human papilloma virus E7. In some embodiments, the at least one of an exponential region, a linear region, and a plateau region occurs over at least 4 cycles. In some embodiments, the at least one of an exponential region, a linear region, and a plateau region occurs over at least 8 seconds.

[0006] Described herein, in some embodiments, are methods of detecting the presence or absence of each of a plurality of target nucleic acid analytes in a sample, the method comprising a) adding a primer set, the primer set comprising at least two primer pairs, to the sample, at least one of the primer pairs comprising a limiting primer, and each primer pair defining a region of one of the target nucleic acid analytes to be amplified; performing an amplification reaction on the sample; b) measuring a signal generated during the amplification reaction; and c) making a determination as to whether each of the plurality of target nucleic acid analytes is present based on the measured signal. In some embodiments, step c) further comprises a step of generating a kinetic signature from the measured signal, and step d) further comprises a step of comparing the kinetic signature to a reference signature, the reference signature corresponding with a kinetic signature generated by an amplification of one or more of the plurality of target nucleic acid analytes with the primer set, thereby making the determination. In some embodiments, the primer set consists of three primers. In some embodiments, the three primers consist of one forward primer and two reverse primers or two forward primers and one reverse primer. In some embodiments, the primer set comprises two or more nested primers. In some embodiments, the signal is

generated by exciting a molecule that binds an amplification product. In some embodiments, the molecule is a hybridization probe. In some embodiments, (A) the primer set comprises first and second primer pairs, the first primer pair having a first annealing temperature and the second primer pair having a second annealing temperature which is lower than the first annealing temperature, and the first primer pair defining a region of one of the target nucleic acid analytes, the second primer pair defining a region of another of the nucleic acid analytes; and (B) the amplification reaction comprises thermocycling of a set of temperature cycles comprising at least one temperature cycle having a minimum temperature which is both (i) not greater than the first annealing temperature, and (ii) greater than the second annealing temperature, the reaction further comprising at least one other temperature cycle having a minimum temperature which is not greater than the second annealing temperature. In some embodiments, the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the first annealing temperature and greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature. In some embodiments, the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature which is both (x) not greater than the first annealing temperature and (y) greater than the second annealing temperature. In some embodiments, step c) further comprises a step of generating a kinetic signature from the measured signal, and step d) further comprises a step of comparing the kinetic signature to a reference signature, the reference signature corresponding to a signal generated when one or more of the plurality of target nucleic acid analytes is amplified with the primer set by a thermocycling reaction comprising the set of temperature cycles, thereby making the determination. In some embodiments, the signal comprises a kinetic signature. In some embodiments, the kinetic signature comprises a signal selected from an electromagnetic signal, a wavelength, and a fluorescence emission signal. In some embodiments, the kinetic signature is plotted as a curve. In some embodiments, the curve comprises at least one of an exponential region, a linear region, or a plateau region. In some embodiments, the kinetic signature comprises recorded values. In some embodiments, the values are recorded in a table. In some embodiments, the limiting primer is present at a concentration sufficiently lower than the concentration of a non-limiting primer to generate a characteristic signal. In some embodiments, the limiting primer is present at a concentration sufficiently lower that

the concentration of a non-limiting primer to generate a characteristic kinetic signature. In some embodiments, the kinetic signature comprises a signal selected from an electromagnetic signal, a wavelength, and a fluorescence emission signal. In some embodiments, the kinetic signature is plotted as a curve. In some embodiments, the curve comprises at least one of an exponential region, a linear region, or a plateau region. In some embodiments, the kinetic signature comprises recorded values. In some embodiments, the values are recorded in a table. In some embodiments, the first and/or second annealing temperatures are determined by the limiting primer. In some embodiments, one or more primer of the primer pairs is linked to an antibody.

[0007] Described herein, in some embodiments, are methods of detecting the presence or absence of each of a plurality of target nucleic acid analytes in a sample, the methods comprising a) adding a primer set, the primer set comprising at least two primers, to the sample, at least one of the primer pairs comprising a limiting primer, and each primer pair defining a region of one of the target nucleic acid analytes to be amplified; b) performing a thermocycling amplification reaction on the sample, the reaction comprising a set of temperature cycles comprising at least one temperature cycle having a minimum temperature which is both (i) not greater than the first annealing temperature, and (ii) greater than the second annealing temperature, the reaction further comprising at least one other temperature cycle having a minimum temperature which is not greater than the second annealing temperature; c) measuring a signal generated during the amplification reaction; and d) making a determination as to whether each of the plurality of nucleic acid analytes is present based on the measured signal. In some embodiments, one or more primer of the primer pairs is linked to an antibody.

[0008] Described herein, in some embodiments, are methods of detecting the presence or absence of two or more of a plurality of target nucleic acid analytes in a sample, the methods comprising a) adding a primer set comprising first and second primer pairs to the sample, the first primer pair having a first annealing temperature and the second primer pair having a second annealing temperature which is lower than the first annealing temperature, and the first primer pair defining a region of one of the target nucleic acid analytes, the second primer pair defining a region of another of the nucleic acid analytes; b) performing a thermocycling amplification reaction on the sample, the reaction comprising a set of temperature cycles comprising at least one temperature cycle having a minimum temperature which is both (i) not greater than the first annealing temperature, and (ii) greater than the second annealing

temperature, the reaction further comprising at least one other temperature cycle having a minimum temperature which is not greater than the second annealing temperature; c) measuring a signal generated during the amplification reaction; and d) making a determination as to the presence or absence of the two or more of the nucleic acid analytes based on the generated signal. In some embodiments, step c) further comprises a step of generating a kinetic signature from the measured signal, and step d) further comprises a step of comparing the kinetic signature to a reference signature, the reference signature corresponding with a signal generated when one or more of the plurality of target nucleic acid analytes is amplified with the primer set by a thermocycling reaction comprising the set of temperature cycles, thereby making the determination. In some embodiments, the first and second primer pairs comprise three primers. In some embodiments, the three primers comprise one forward primer and two reverse primers. In some embodiments, the three primers comprise two forward primers and one reverse primer. In some embodiments, the primer set comprises at least one pair of nested primers. In some embodiments, the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the first annealing temperature and greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature. In some embodiments, the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature which is both (x) not greater than the first annealing temperature and (y) greater than the second annealing temperature. In some embodiments, at least one of the primer pairs comprises a limiting primer. In some embodiments, the limiting primer is present at a concentration of less than 1 μM . In some embodiments, a non-limiting primer is present at an initial concentration of at least two times greater than the limiting primer. In some embodiments, the limiting primer has a concentration-adjusted melting temperature that is equal to or greater than a non-limiting primer. In some embodiments, the signal is generated by a molecule that binds an amplification product. In some embodiments, the molecule is a hybridization probe. In some embodiments, the minimum temperature corresponds to the annealing temperature of first and second primer pairs. In some embodiments, the method comprises detecting the presence of n target nucleic acid analytes using n primer pairs, each primer pair having a different annealing temperature from that of the other primer pairs, and the set of temperature cycles

comprising a set of minimum temperatures that are interspersed between the different annealing temperatures.. In some embodiments, n is 3. In some embodiments, n is 4. In some embodiments, n is 5. In some embodiments, n is 6. In some embodiments, n is 7. In some embodiments, one or more primer of a primer pair is linked to an antibody.

[0009] Described herein, in some embodiments, are methods of detecting the presence or absence of each of a plurality of target nucleic acid analytes in a sample, the methods comprising a) adding a primer set comprising first and second primer pairs to the sample, the first primer pair having a first annealing temperature and the second primer pair having a second annealing temperature which is lower than the first annealing temperature, and the first primer pair defining a region of one of the target nucleic acid analytes, the second primer pair defining a region of another of the nucleic acid analytes; b) performing a thermocycling amplification reaction on the sample, the reaction comprising a set of temperature cycles comprising at least one temperature cycle having a minimum temperature which is both (i) not greater than the first annealing temperature, and (ii) greater than the second annealing temperature, the reaction further comprising at least one other temperature cycle having a minimum temperature which is not greater than the second annealing temperature; c) measuring a signal generated during the amplification reaction, thereby generating a kinetic signature from the measured signal; and d) comparing the kinetic signature to a reference signature, the reference signature corresponding with a signal generated when one or more of the plurality of target nucleic acid analytes is amplified in a sample comprising the primer set by a thermocycling reaction comprising at least one temperature cycle having a minimum temperature which is both (x) not greater than the first annealing temperature, and (y) greater than the second annealing temperature, the reaction further comprising at least one other temperature cycle having a minimum temperature which is not greater than the second annealing temperature, thereby making a determination as to the presence or absence of the at least two nucleic acid analytes.

[0010] In an aspect the present disclosure provides a method for detecting a nucleic acid analyte. The method comprises performing a nested asymmetric amplification reaction on a sample, measuring a signal generated during the nested asymmetric amplification reaction, and determining whether the analyte is present based on the signal. The nested asymmetric amplification reaction of the disclosure may comprise at least three nested primers, wherein at least one pair of primers of the at least three nested primer comprises a limiting primer and an excess primer. In some cases, the at least three nested primers bind to a single analyte. In

some examples, the at least three nested primers may comprise at least two pairs of nested primers. In some further examples, the at least three nested primers comprise at least three pairs of nested primers. In some cases, the at least three nested primers may comprise at least two pairs of nested primers that bind to the same analyte. In some cases, the at least three pairs of nested primers may bind to the same analyte. The limiting primer may be present at an initial concentration of less than $1\mu\text{M}$. The excess primer may be present at an initial concentration of at least 2 times greater than the limiting primer. The limiting primer may have a concentration-adjusted melting temperature that is equal to or greater than the excess primer.

[0011] The method of the disclosure may detect the presence or absence of a plurality of the analytes in a sample. In some cases, the methods of the disclosure may detect the presence or absence of at least 1 analyte in a sample. The method of the disclosure may further comprise performing a melting curve analysis.

[0012] The signal may be signal that is an electromagnetic signal. For example, the signal may be a fluorescence emission signal. The signal may comprise intensity and a frequency. An analyte may be encoded by a signal intensity at a particular wavelength. An analyte may be also be encoded by a plurality of signal intensities at a plurality of wavelengths. The plurality of wavelengths may comprise at least three wavelengths. In some examples, the analyte may be encoded by at least one additional value selected from the group consisting of a value from the signal and a value from an additional signal. The least one additional value may be selected from the group consisting of a Förster resonance energy transfer (FRET) emission intensity, a FRET emission wavelength, an electrochemical signal, a chemiluminescence wavelength, a chemiluminescence intensity, a fluorescence bleaching rate, and a chemiluminescence bleaching rate. In some cases the at least one additional value is Förster resonance energy transfer (FRET) emission intensity.

[0013] The signal may be generated by a molecule that binds an amplification product. For example the signal may be generated by a hybridization probe. The hybridization probe may comprise a fluorophore and a quencher. The amplification reaction may comprise a plurality of hybridization probes that are specific for different analytes and comprise an identical fluorophore. In some examples, the number of the hybridization probes may be greater than the number of the analytes.

[0014] The signal generated by a single analyte may be plotted as a function of a variable selected from the group consisting of time and cycle number, thereby generating a plot. The

plot may comprise at least two of the same regions selected from the group consisting of an exponential region, a linear region, and a plateau region. In some cases, both of the exponential regions occur over at least 4 cycles. In some cases, both of the exponential regions occur over at least 8 seconds. In some cases, both of the linear regions occur over at least 4 cycles. In some cases, both of the linear regions occur over at least 8 seconds. In some cases, both of the plateau regions occur over at least 4 cycles. In some cases, both of the plateau regions occur over at least 8 seconds.

[0015] The analyte may be selected from the group consisting of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The methods of the disclosure may further comprise reverse transcribing an RNA. The analyte may be present at less than 1,000 copies. The sample may further comprise a reference analyte. The analyte may be from a source selected from the group consisting of an animal, a plant, a bacteria, a fungus, a parasite, and a virus. The animal may be human. In some cases, the analyte is from a source selected from the group consisting of human immunodeficiency virus, herpes simplex virus, human papilloma virus, Plasmodium, Mycobacterium, dengue virus, hepatitis virus, and influenza virus. In some other examples, the analyte is selected from the group consisting of human immunodeficiency virus polyprotease, human immunodeficiency virus p17, human papilloma virus E6, and human papilloma virus E7. The analyte may also be a nucleic acid variant. The analyte may be a nucleic acid encoding a 16S RNA sequence. The sample may be selected from the group consisting of a clinical sample, a food sample, an environmental sample, a pharmaceutical sample, and a sample from a consumer product.

[0016] The method of the disclosure may further comprise transforming a signal intensity into a frequency domain, thereby generating a transformed signal. The transforming may be performed by a method selected from the group consisting of a Fourier transform, a fast Fourier transform, a discrete Fourier transform, a discrete-time Fourier transform, and a wavelet transform. The method may further comprise comparing the transformed signal to a reference signal. The comparing may be used to determine if the analyte is present or absent.

[0017] The methods of the disclosure may be useful for unambiguously determining whether one or more analyte is present or absent. The methods of the disclosure may also comprise determining the initial amount of the analyte in the sample. The method of disclosure may further comprising making a clinical decision based on whether the analyte is present or absent. The clinical decision may comprise selection of an agent for administration to a subject in need thereof. The agent may be a therapeutic agent. The therapeutic agent may be

selected from the group consisting of an antibiotic and an antiviral. The method the disclosure may further comprise transmitting information concerning the analyte through a computer network. The transmitting may be to a physician. In some examples, at least one step of the methods of the disclosure may be performed using instructions on a non-volatile computer readable medium. The instructions may be located on a remote server or on a thermal cycler.

[0018] The disclosure also provides a method of determining an amount of nucleic acid analyte. The method comprises performing an amplification reaction on a sample, measuring a signal generated during the amplification reaction, transforming the signal into a frequency domain thereby generating a transformed signal, comparing a value from the transformed signal to a value of reference signal, and determining the amount of the analyte based on the comparing. The transforming may be performed by a method selected from the group consisting of a Fourier Transform, a fast Fourier transform, a discrete Fourier transform, a discrete-time Fourier transform, and a wavelet transform. Transforming may be performed by an eigenbasis decomposition. The eigenbasis decomposition may be of a PCR curve. The eigenbasis decomposition may be of a characteristic PCR curve. The eigenbasis decomposition may comprise one or more eigenfunctions. The one or more eigenfunctions may comprise time-shifted, scaled sigmoidal, exponential, and/or linear curves corresponding to one or more primer sets. Transforming may be performed by an eigenbasis decomposition of a characteristic PCR curve comprising eigenfunctions consisting of time-shifted, scaled sigmoidal, exponential, and linear curves corresponding to specific primer-sets within a nested PCR amplification. The methods of the disclosure detect the presence or absence of a plurality of analytes. In some examples, each of the analytes produces a unique transformed signal amongst the analytes. The comparing may comprise comparing an intensity value at a particular frequency and time. In some cases, at least about 5 intensity values are generated during each cycle of the amplification reaction. The reference signal may be in a database of reference signals. The database may comprise at least 20 reference signals. The database may comprise at least 30 reference signals. The database may comprise at least 50 reference signals. The database may comprise less than 500 reference signals. The database may comprise less than 200 reference signals. The database may comprise less than 100 reference signals.

[0019] The disclosure also provides a kit for detecting a nucleic acid analyte. The kits of the disclosure comprise at least three nested primers. At least one pair of primers may comprise a

limiting primer and an excess primer. The kit may also comprise a real time detection agent. The kit may detect the presence or absence of a plurality of analytes. At least one of the primers may be lyophilized. In some cases each of the primers is lyophilized. The kit may further comprise reagents for the detection of an amplified analyte. The reagents for the detection of an amplified analyte may comprise a fluorophore. The reagents for the detection of an amplified analyte may comprise a hybridization probe.

[0020] In some examples the at least three nested primers in the kit may bind to a single analyte. In some cases the at least three nested primers of the kit comprise at least two pairs of nested primers. The at least three nested primers may comprise at least three pairs of nested primers. In some further examples, the at least three nested primers comprise at least two pairs of nested primers that bind to the same analyte. In some cases the at least three nested primers comprise at least three pairs of nested primers that bind to the same analyte. The limiting primer may be present at an initial concentration of less than 1 μM . The excess primer may be present at an initial concentration of at least 2 times greater than the limiting primer. The limiting primer may have a concentration-adjusted melting temperature that is equal to or greater than the excess primer.

INCORPORATION BY REFERENCE

[0021] All publications and patent applications mentioned in this specification are herein incorporated in their entireties by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0023] **Fig. 1** shows a diagram illustrating the method of nested PCR.

[0024] **Fig. 2** shows a nested PCR with three primer pairs (A, B, C). The targeted area of interest is shown in the box.

[0025] **Fig. 3** shows the fluorescence curve of a saturated, quantitative polymerase chain reaction.

[0026] **Fig. 4** shows the fluorescence curve of a biased-primer, quantitative polymerase chain reaction.

[0027] **Fig. 5** shows the fluorescence curve of a nested, quantitative polymerase chain reaction.

[0028] **Fig. 6** shows the characteristic fluorescence curves (curve ABC and curve DEF) of two polymerase chain reactions.

[0029] **Fig. 7** shows the characteristic fluorescence curves (curve ABC and curve DEF) of two polymerase chain reactions transformed into the frequency domain.

[0030] **Fig. 8** shows a block diagram illustrating a first example architecture of a computer system that can be used in connection with example embodiments of the present disclosure.

[0031] **Fig. 9** shows a diagram illustrating a computer network that can be used in connection with example embodiments of the present disclosure.

[0032] **Fig. 10** shows a block diagram illustrating a second example architecture of a computer system that can be used in connection with example embodiments of the present disclosure.

DETAILED DESCRIPTION OF THE INVENTION

I. General Overview

[0033] The present disclosure relates to methods, systems, compositions and kits relating to the detection and/or quantification of one or more analytes in a sample. The methods may be particularly useful for detection of a plurality of analytes in multiplexed reactions. Multiplexed reactions can permit performance of parallel reactions on the same sample, use of the same chamber to perform multiple reactions, and the ability to extract rich information from a sample in a fast and efficient manner. The methods disclosed herein may permit performance of multiplexed assays in a single solution. These methods can be used in place of or in addition to spectrally resolved fluorescence or chemiluminescence (e.g., PCR, ELISA), spatially resolved signals (e.g., microarrays, gel electrophoresis), temporally resolved signals (e.g., capillary electrophoresis), or combinations thereof (e.g., Sanger sequencing).

[0034] Specificity of the binding between PCR primers and template allows for selective amplification of intended targets. For instance, a typical quantitative polymerase chain reaction is run to “saturation,” wherein the amplification reaction is run to the point at which at least one of the reactants such as primers, polymerase, probe, or free nucleotides is

exhausted. When fluorescent reporters such as TaqMan probes, FRET probes, or intercalating dyes are used to interpret the extent of the DNA amplification, a fluorescence curve could be generated. The sigmoidal signature of a “saturation” fluorescence curve often occurs when the reporter probe concentrations are depleted at an ever-increasing rate.

[0035] However, this reaction (and by extension, its signature fluorescence) is limited by all reactant concentrations, including the PCR primers. As an example, when a single set of primers is used in equal concentrations, and when one primer is more efficiently extended than the other, the reaction becomes rate limited. This results in a linear, as compared to an exponential, amplification of DNA.

[0036] Generally, the methods of the disclosure comprise performing an amplification reaction, measuring a signal generated during the amplification reaction and determining the amount of the analytes present in the sample based on the measured signal. The method may additionally include transforming the measured signal to a second domain and comparing the transformed signal to a reference signal. The amount of the analyte present in the sample or the initial amount of analyte present in the sample may also be obtained based on the comparison of the transformed signal to the measured signal.

[0037] A signal may be anything capable of measurement. For example, a signal may be a fluorescence signal. The amplification reaction for the methods of the disclosure may comprise semi-nested or nested PCR reactions. In these PCR reactions, at least two sets of primers are used in at least two successive reactions. In the first PCR reaction, one pair of primers is used. The products from the first PCR reactions are then used as template in a second PCR reaction, using one (semi-nested PCR reactions) or two different primers (nested PCR reactions). Figure 1 shows a diagram illustrating the method of nested PCR and Figure 2 shows a nested PCR with three primer pairs (A, B, C).

[0038] The amplification reactions of the disclosure are performed so as to generate characteristic signal curves that may be indicative of the presence and/or absence of one or more analytes. The characteristic curves may also be used to measure the amount of one or more analytes present in a sample. Such characteristic curves may be generated by varying (i) the number of primer pairs per target analyte, (ii) the primer sequences used for target analytes, and/or (iii) the ratio of the forward and the reverse primers in each primer pair.

[0039] Typically the amplification reactions of the disclosure are run such that either the forward or the reverse primer in each primer pair is rate limited. This may be done by adding each primer set such that there are different ratios of reverse and forward primers. For

example, asymmetric PCR reactions, asymmetric nested PCR reactions, or asymmetric semi-nested PCR reactions may be used as amplification reactions in the methods of the disclosure. By choosing suitable ratios of the forward and the reverse primers, the amplification reactions of the disclosure may be tailored to generate curves that exhibit a characteristic linear amplification of the analytes. While this linear amplification may be non-ideal for efficient amplification of the target analytes, it may allow for discrimination of targets. Furthermore, tuning of primer concentration and the number of primers for a particular nested reaction allows for the engineering of a particular fluorescent signature for a given target. The fluorescent signal generated by an analyte may be plotted as a function of a variable selected from the group consisting of time and cycle number, thereby generating a plot or a characteristic curve. Preferably, the fluorescence signal is provided in AFU and is plotted against the number of cycles.

[0040] The characteristic curves generated by the methods of the disclosure may comprise one or more baseline regions, one or more exponential regions, one or more linear regions, and/or one or more plateau regions. In some cases the curves comprise at least two of the same regions selected from the group consisting of a baseline region, an exponential region, a linear region, and a plateau region. For example, the curves comprise one or more of at least two baseline regions, at least two exponential regions, at least two linear regions and/or at least two plateau regions. The amplification reactions of the disclosure may be conducted such that each target analyte in a sample generates a characteristic curve, which may indicate the presence or absence of the particular target analyte. A curve can be characterized by one or more of: (i) the number of baseline regions in a curve, (ii) the number of linear regions in a curve, (iii) the number of exponential regions in a curve, (iv) the number of plateau regions in a curve, (v) the slope of the one or more linear regions in a curve, (vi) the duration/length of one or more plateau regions in a curve, (vii) the duration/length of one or more linear regions in a curve, (viii) the duration/length of one or more exponential regions in a curve, (ix) the duration/length of one or more baseline regions in a curve, (x) the duration between two consecutive baseline regions in a curve, (xi) the duration between two consecutive exponential regions in a curve, (xii) the duration between two consecutive plateau regions in a curve, and/or (xiii) the duration between two consecutive linear regions in a curve. In some cases a curve is characterized by one or more of the preceding properties along with one or more additional properties not listed here. In some cases a curve is only characterized by one or more properties not listed here.

[0041] Once the characteristic curves are generated, these curves may be analyzed by curve fitting algorithms. These algorithms may determine the number of baseline regions in a curve, the number of linear regions in a curve, the number of exponential regions in a curve, the number of plateau regions in a curve, the slope of the one or more linear regions in a curve, the duration/length of one or more plateau regions in the curve, the duration/length of one or more linear regions in the curve, the duration/length of one or more exponential regions in the curve, the duration/length of one or more baseline regions in the curve, the duration between two consecutive baseline regions in the curve, the duration between two consecutive exponential regions in the curve, the duration between two consecutive plateau regions in the curve, and/or the duration between two consecutive linear regions in the curve. Furthermore, if these curves are in superposition (e.g. more than one target is in a single, multiplexed reaction), each target can be uniquely identified if the curves are in the linear dynamic range of the fluorometer. For example, Figure 6 shows two characteristic curves that may be generated from a sample comprising two target analytes by the methods of the disclosure.

[0042] The methods presented in this disclosure may be useful in a variety of applications. For example, the methods may be used to generate diagnostic data wherein these methods are employed to detect the presence or absence of one or more analytes that may be indicative of a disease or a condition. The methods are carried out such that each analyte, indicative of a disease, generates a characteristic curve. Exemplary diseases that may be diagnosed by the methods of the disclosure include cancer, autoimmune disease, cardio respiratory disease, liver disease, digestive disease, and so on.

[0043] In some cases, the methods may be used to detect infection with one or more pathogens. The methods provided herein may also be used to make a diagnosis and to make a clinical decision based on that diagnosis. For example, whether a particular target analyte is present or absent in a sample from an individual may assist in selection of an agent for administration into the subject. The agent is a therapeutic agent for example, antibiotic or antiviral agent. In some cases, the methods of the disclosure may indicate the presence of a bacterial polynucleotide in a sample taken from a subject, this may lead to the treatment of the subject by administering an antibiotic to the subject. In other cases, the methods of the disclosure may indicate the presence of a viral polynucleotide in a sample taken from a subject, this may lead to a clinical decision of administering an antiviral agent to the subject.

[0044] The methods may further be used to monitor the progress of a disease in an individual or to monitor the efficiency of a treatment method. Such methods include monitoring samples from the individual at regular time intervals and studying the changes in the characteristic curves generated by the corresponding one or more disease indicative analytes.

[0045] In other examples, the methods of the disclosure may also be used to identify genetic markers in an individual or a population of individuals. The genetic marker may be a molecular genetic marker, for example PCR based markers. Non-limiting examples of techniques that may be used for analysis of genetic markers by the methods of the disclosure include RAPD (random amplification of polymorphic DNA), SSRs (simple sequence repeats (SSR or microsatellite, or short tandem repeats)), SCAR (sequence characterized amplified region), AFLP (amplified fragment length polymorphism (AFLP)), SNP (single nucleotide polymorphism), CAPS (cleaved amplified polymorphic sequence), etc.).

II. Definitions

[0046] As used in this specification and the appended claims, the singular forms “a”, “an”, and “the” include plural references unless the context clearly dictates otherwise. Thus, for example, references to “the method” includes one or more methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

[0047] “About” as used herein when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of $\pm 20\%$ or $\pm 10\%$, or $\pm 5\%$, or even $\pm 1\%$ from the specified value, as such variations are appropriate for the disclosed compositions or to perform the disclosed methods.

[0048] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs.

[0049] The term “annealing temperature” as used herein refers to the maximum temperature at which a nucleic acid anneals to a target nucleic acid under given conditions sufficient for extension of the primer to proceed. When used to refer to the annealing temperature of a primer pair, this term refers to the maximum temperature, under given conditions, at which each member of a primer pair anneals to its corresponding sequence of a target template or analyte to be amplified, sufficient for extension of the primer to proceed. Where, as is usual, each member of a pair binds to a separate sequence, on an opposite strand, of a target double-stranded nucleic acid, the “annealing temperature” is the temperature at which both members of the pair is bound to its complementary sequence. Where one member of the pair binds at a

lower temperature than the other member, the annealing temperature of the pair will be that lower temperature. Thus, the term “annealing temperature” refers to the annealing temperature of the primer with a lower annealing temperature in a primer pair. The “annealing temperature” of a primer or primer pair may be below the melting temperature of the primer or primer pair. The “annealing temperature” may be lower than the “melting temperature” by any number of degrees Celsius, including, but not limited to 0.5°C, 1.0°C, 1.5°C, 2.0°C, 2.5°C, 3.0°C, 3.5°C, 4.0°C, 4.5°C, 5.0°C, 5.5°C, 6.0°C, 6.5°C, 7.0°C, 7.5°C, 8.0°C, 8.5°C, 9.0°C, 9.5°C, 10.0°C. In some embodiments, annealing temperatures recited herein are concentration-adjusted annealing temperatures.

[0050] The term “melting temperature” as used herein refers to the temperature at which half of a DNA or nucleic acid duplex becomes single-stranded. The term “primer melting temperature” refers to the temperature at which half of the primers associated with a nucleic acid analyte will dissociate.

[0051] As used herein, the term analyte encompasses a nucleic acid, a portion of a nucleic acid, an open reading frame, an exonic region, or an intronic region. The analyte may be present on a chromosome, a genome, a plasmid, or a cDNA. An analyte can be an overlapping nucleic acid sequence on the same strand of a chromosome or genome, for example.

[0052] As used herein, the term “temperature cycle” refers to the temperatures between which a sample is varied during a thermocycling amplification reaction. For example, in a standard thermocycling reaction, the sample might be repeatedly varied between an annealing temperature and a melting temperature. As used herein, each “temperature cycle” would comprise both the annealing and extension temperatures, as well as one or more temperature adjustment steps. As used herein, the term “n” refers to any positive integer.

III. Curve Signatures

[0053] Typically, the methods of the disclosure comprise plotting the fluorescent signal generated by an analyte as a function of a variable selected from the group consisting of time and cycle number, thereby generating a plot or a characteristic curve. Preferably, the fluorescence signal is provided in AFU and is plotted against cycle number. Often the AFU value may be normalized to a fixed proportional scale, for example to the scale of 0 to 100.

[0054] A typical fluorescent curve generated for a target analyte by the methods of the disclosure may comprise one or more baseline regions, plateau regions, linear regions and exponential regions. In some cases the curve comprises at least two of the same regions

selected from the group consisting of a baseline region, an exponential region, a linear region, and a plateau region. As mentioned herein, a curve can be characterized by any one or more of the following properties: (i) the number of baseline regions in the curve, (ii) the number of linear regions in the curve, (iii) the number of exponential regions in the curve, (iv) the number of plateau regions in the curve, (v) the slope of the one or more linear regions in the curve, (vi) the duration/length of one or more plateau regions in the curve, (vii) the duration/length of one or more linear regions in the curve, (viii) the duration/length of one or more exponential region in the curve, (ix) the duration/length of one or more baseline regions in the curve, (x) the duration between two consecutive baseline regions in the curve, (xi) the duration between two consecutive exponential regions in the curve, (xii) the duration between two consecutive plateau regions in the curve, and/or (xiii) the duration between two consecutive linear regions in the curve. In some cases a curve is characterized by one or more of the preceding property along with one or more additional properties not listed here. In some cases a curve is characterized only by one or more properties not listed here.

[0055] In some cases the characteristic curves may differ in any one of the above listed properties. For example, the curves may differ in number of baseline regions; or the curves may differ in number of plateau regions; or the curves may differ in the number of exponential regions; or the curves may differ in the number of linear regions; or the curves may differ in slope of the one or more linear regions; or the curves may differ in duration/length of one or more plateau regions; or the curves may differ in duration/length of one or more linear regions; or the curves may differ in duration/length of one or more exponential regions; or the curves may differ in the duration of one or more baseline regions; or the curves may only differ in the duration between any two consecutive baseline regions of the curves; or the curves may differ in the duration between any two consecutive exponential regions of the curve; or the curves may differ in the duration between any two consecutive plateau regions of a curve; or the curves may only differ in the duration between any two consecutive linear regions of a curve. For example, a curve with one linear region may indicate the presence of first target analyte in a sample while another curve with two linear regions may indicate the presence of a second target analyte in a sample. In another example, a curve with a positive slope of the linear region may indicate the presence of first target analyte in a sample while another curve with a negative slope in the linear region may indicate the presence of a second target analyte in a sample.

[0056] In some cases the curves will differ in any two of the above listed properties. For example the curves may differ in number of plateau regions and the number of linear regions; or the curves may differ by the slope of one or more linear regions and the number of plateau regions; or the curves differ in the length/duration of one or more linear regions and the length/duration of one or more plateau region.

[0057] In other cases the curves will differ in any 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 of the above listed properties. In some cases the curves will differ in all 13 of these properties. In some cases the curves will differ in a combination of these properties plus one or more additional properties not listed here.

[0058] The number of baseline regions, plateaus, exponential regions, and the number of linear regions in a curve, indicative of a particular target analyte, may be varied by varying the number of nested primers used for a target analyte. If a single primer pair is used for a target analyte, the resulting curve will typically have a single baseline region, a single plateau, a single exponential region and a single linear region. As the number of nested primer pairs is increased to 2, 3, 4, etc. the number of baseline regions, plateaus, exponential regions, and the number of linear regions may increase correspondingly.

[0059] The characteristic curves may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more baseline regions. For example the curve may comprise 15, 20, 25 or even more baseline regions. The characteristic curves may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more plateau regions. For example the curve may comprise 15, 20, 25 or even more plateau regions. The characteristic curves may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more exponential regions. For example the curve may comprise 15, 20, 25 or even more exponential regions. The characteristic curves may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more linear regions. For example the curve may comprise 15, 20, 25 or even more linear regions.

[0060] In some case the characteristic curves will comprise equal number of baseline regions, plateau regions, exponential regions, and linear regions. For example, a characteristic curve may comprise 1 baseline region, 1 exponential region, 1 linear region and 1 plateau region. Similarly a characteristic curve may comprise 2 baseline regions, 2 exponential regions, 2 linear regions and 2 plateau regions. Alternatively characteristic curve may comprise 3 baseline regions, 3 exponential region, 3 linear regions and 3 plateau regions. A characteristic curve may also comprise 4, 5, 6, 7, 8, 9, 10 or more of each baseline regions, exponential regions, linear regions and plateau regions. For example the curve may comprise 15, 20, 25

or even more of each baseline regions, exponential regions, linear regions and plateau regions.

[0061] In some cases the characteristic curves will comprise different numbers of baseline regions, plateau regions, exponential regions, and linear regions. In some cases the characteristic curves will comprise equal number of baseline regions, plateau regions, and exponential regions but a different number of linear regions. In some cases the characteristic curves will comprise equal number of baseline regions, exponential regions, and linear regions but a different number of plateau regions. In some cases the characteristic curves will comprise equal number of baseline regions, plateau regions and linear regions but a different number of exponential regions. In some cases the characteristic curves will comprise equal numbers of plateau regions, exponential regions, and linear regions, but a different number of baseline regions.

[0062] The curves may vary in the length or the duration of one or more plateau regions. The length/duration of a plateau region in a curve can for example be at least 2 cycles, for example at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, or 100 cycles. In some cases, the length or the duration of a plateau region will be at most 100 cycles, for example at most 50, 45, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases the length/duration of a plateau region in a curve is calculated in the unit of time. In such cases, the length/duration of a plateau region in a curve can for example be at least 30 seconds, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 minutes. In some cases, the length or the duration of a plateau region will be at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Further multiple plateau regions of a curve may all have same or different length/duration.

[0063] Similarly, the curves may also vary in the length or duration of the linear regions. The length/duration of a linear region in a curve can for example be at least 2 cycles, for example at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, or 100 cycles. In some cases, the length or the duration of a linear region will be at most 100 cycles, for example at most 50, 45, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases the length/duration of a linear region in a curve is calculated in the unit of time. In such cases, the length/duration of a linear region in a curve can for example be at least 30 seconds, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70,

75, 80, 85, 90, 95 or 100 minutes. In some cases, the length or the duration of a linear region will be at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Multiple linear regions of a curve may all have same or different length/duration.

[0064] The curves may vary in the length or the duration of one or more baseline regions. The length/duration of a baseline region in a curve can for example be at least 2 cycles, for example at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, or 100 cycles. In some cases, the length or the duration of a baseline region will be at most 100 cycles, for example at most 50, 45, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases the length/duration of a baseline region in a curve is calculated in the unit of time. In such cases, the length/duration of a baseline region in a curve can for example be at least 30 seconds, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 minutes. In some cases, the length or the duration of a baseline region will be at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Further multiple baseline regions of a curve may all have same or different length/duration.

[0065] The curves may vary in the length or the duration of one or more exponential regions. The length/duration of an exponential region in a curve can for example be at least 2 cycles, for example at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, or 100 cycles. In some cases, the length or the duration of an exponential region will be at most 100 cycles, for example at most 50, 45, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases the length/duration of an exponential region in a curve is calculated in the unit of time. In such cases, the length/duration of an exponential region in a curve can for example be at least 30 seconds, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 minutes. In some cases, the length or the duration of an exponential region will be at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Further multiple exponential regions of a curve may all have same or different length/duration.

[0066] The slope (or gradient) of a linear region in a curve may be varied by varying the ratio of the corresponding forward and reverse primers. The slope of one or more linear regions of the curves may have any value. Different linear regions of a curve may have same or

different values. In some case some of the multiple linear regions a curve have same values. In other cases multiple linear regions will all have different values. In some case a few of the multiple linear regions will have same value while the other will have different values. The slope may be steep (for example, a steep incline or steep decline) or shallow (for example, a shallow incline or a shallow decline). The value of the slope of the linear regions of the curves may be either positive or negative. In some cases the slope is positive with a numerical value in the range of 0-0.01, 0-0.1, 0-1, 0-2, 1-3, 0-4, 0-5, 0-6, 0-7, 0-8, 0-9, 0-10, 0-20, 0-30, 0-40, 0-50, 0-60, 0-70, 0-80, 0-90, or 0-100. In some cases the slope is positive with a numerical value greater than 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000. In some cases a slope is negative with a numerical value in the range of 0-0.01, 0-0.1, 0-1, 0-2, 1-3, 0-4, 0-5, 0-6, 0-7, 0-8, 0-9, 0-10, 0-20, 0-30, 0-40, 0-50, 0-60, 0-70, 0-80, 0-90, or 0-100. In some cases the slope is negative with a numerical value greater than 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000.

[0067] The curves may vary in the duration between two consecutive exponential regions. In some cases two consecutive exponential regions of a curve are separated by at least 1 cycles, for example at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 10,000 cycles. In some cases, the occurrence of two consecutive exponential regions of a curve are separated by at most 10,000 cycles, for example at most 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 45, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases two consecutive exponential regions of a curve are separated by at least 1 cycles, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 minutes. In some cases two consecutive exponential regions of a curve are separated by at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Further, if a curve comprises more than two exponential regions, each pair of consecutive exponential regions may be separated by a same or a different number of cycles.

[0068] The curves may vary in the duration between two consecutive baseline regions. In some cases two consecutive baseline regions of a curve are separated by at least 1 cycles, for example at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 10,000 cycles. In some cases, the occurrence of two consecutive baseline regions of a curve are separated by at most 10,000 cycles, for example at most 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 45, 40, 35,

30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases two consecutive baseline regions of a curve are separated by at least 30 seconds, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 minutes. In some cases two consecutive baseline regions of a curve are separated by at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Further, if a curve comprises more than two baseline regions, each pair of consecutive baseline regions may be separated by a same or a different number of cycles.

[0069] The curves may vary in the duration between two consecutive plateau regions. In some cases two consecutive plateau regions of a curve are separated by at least 1 cycles, for example at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 10,000 cycles. In some cases, the occurrence of two consecutive plateau regions of a curve are separated by at most 10,000 cycles, for example at most 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 45, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases two consecutive plateau regions of a curve are separated by at least 30 seconds, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 minutes. In some cases two consecutive plateau regions of a curve are separated by at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Further, if a curve comprises more than two plateau regions, each pair of consecutive plateau regions may be separated by a same or a different number of cycles.

[0070] The curves may vary in the duration between two consecutive linear regions. In some cases two consecutive linear regions of a curve are separated by at least 1 cycles, for example at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 10,000 cycles. In some cases, the occurrence of two consecutive linear regions of a curve are separated by at most 10,000 cycles, for example at most 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 45, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 cycle. In some cases two consecutive linear regions of a curve are separated by at least 30 seconds, for example at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 minutes. In some cases two consecutive linear regions of a curve are separated by at most 100 minutes, for example at most 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17,

16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 minute. Further, if a curve comprises more than two linear regions, each pair of consecutive linear regions may be separated by a same or a different number of cycles.

[0071] When the methods of the present disclosure are used with multiplexed PCR assay, the curves obtained may be a combination/addition of the component curves. This disclosure also provides methods of interpreting combined curves. For example, combination of linear regions of two or more curves will result in a linear region wherein the resulting slope is the sum of the individual slopes. This may allow two or more reactions to be coded and decoded by their slopes. Similarly, linear region of one curve and exponential region of another curve may combine to a signature that may neither be purely linear nor purely exponential. Such combination may be decoded by a mathematical fit to the experimental data to tell that both targets may be present and amplifying. Further, combining a plateau region of a curve with a linear region of another curve may produce a result that may bear uniquely decodable information about both, e.g. the slope may identify the latter, while the magnitude of the signal at the starting cycle of the combined region may help identify the plateau.

IV. Methods

[0072] The disclosure provides methods and materials for rapidly and simultaneously performing genotypic analysis for many genetic markers of an individual. The methods comprise obtaining a nucleic acid sample from the individual to be genotyped, subjecting the sample to the amplification methods disclosed herein to amplify segments containing the genetic markers to be examined, plotting the signal measured during the amplification reaction to generate characteristic curves, and determining the presence or absence of genetic markers in the sample based on the characteristic curves obtained. The amplification reactions may be conducted by contacting the sample with one or more nested primer pairs (for example, allele-specific primers). The number of nested primer pairs, the primer sequences and their concentrations may be adjusted so that each genetic marker may generate a characteristic signal curve. Accordingly, every genetic marker present in the nucleic acid sample of the individual will generate a characteristic curve. These methods may also be repeated for a large group of individuals to obtain a genetic profile of genetic markers in a population of individuals.

[0073] The disclosure also provides methods for the detection of one or more diseases in an individual. For example, the method may be used to detect specific mutations associated with the onset and progression of a disease. Non limiting examples of the diseases that may be

detected by the methods of the disclosure include cancer, cystic fibrosis, heart diseases, and leukemia. Other diseases that may be detected by the methods of this disclosure include Achondroplasia, Adrenoleukodystrophy, X-Linked, Agammaglobulinemia, X-Linked, Alagille Syndrome, Alpha-Thalassemia X-Linked Mental Retardation Syndrome, Alzheimer Disease, Alzheimer Disease, Early-Onset Familial, Amyotrophic Lateral Sclerosis Overview, Androgen Insensitivity Syndrome, Angelman Syndrome, Ataxia Overview, Hereditary, Ataxia-Telangiectasia, Becker Muscular Dystrophy also The Dystrophinopathies), Beckwith-Wiedemann Syndrome, Beta-Thalassemia, Biotinidase Deficiency, Branchiootorenal Syndrome, BRCA1 and BRCA2 Hereditary Breast/Ovarian Cancer, Breast Cancer, CADASIL, Canavan Disease, Cancer, Charcot-Marie-Tooth Hereditary Neuropathy, Charcot-Marie-Tooth Neuropathy Type 1, Charcot-Marie-Tooth Neuropathy Type 2, Charcot-Marie-Tooth Neuropathy Type 4, Charcot-Marie-Tooth Neuropathy Type X, Cockayne Syndrome, Colon Cancer, Contractural Arachnodactyly, Congenital, Cranio synostosis Syndromes (FGFR-Related), Cystic Fibrosis, Cystinosis,, Deafness and Hereditary Hearing Loss, DRPLA (Dentatorubral-Pallidoluysian Atrophy), DiGeorge Syndrome (also 22ql 1 Deletion Syndrome), Dilated Cardiomyopathy, X-Linked, Down Syndrome (Trisomy 21), Duchenne Muscular Dystrophy (also The Dystrophinopathies), Dystonia, Early-Onset Primary (DYT1), Dystrophinopathies, The, Ehlers-Danlos Syndrome, Kyphoscoliotic Form, Ehlers-Danlos Syndrome, Vascular Type, Epidermolysis Bullosa Simplex, Exostoses, Hereditary Multiple, Facioscapulohumeral Muscular Dystrophy, Factor V Leiden Thrombophilia, Familial Adenomatous Polyposis (FAP), Familial Mediterranean Fever, Fragile X Syndrome, Friedreich Ataxia, Frontotemporal Dementia with Parkinsonism-17, Galactosemia, Gaucher Disease, Hemochromatosis, Hereditary, Hemophilia A, Hemophilia B, Hemorrhagic Telangiectasia, Hereditary 55, Hearing Loss and Deafness, Nonsyndromic, DFNA (Connexin 26), Hearing Loss and Deafness, Nonsyndromic, DFNB 1 (Connexin 26), Hereditary Spastic Paraplegia, Hermansky-Pudlak Syndrome, Hexosaminidase A Deficiency (also Tay-Sachs), Huntington Disease, Hypochondroplasia, Ichthyosis, Congenital, Autosomal Recessive, Incontinentia Pigmenti, Kennedy Disease (also Spinal and Bulbar Muscular Atrophy), Krabbe Disease, Leber Hereditary Optic Neuropathy, Lesch-Nyhan Syndrome Leukemias, Li-Fraumeni Syndrome, Limb-Girdle Muscular Dystrophy, Lipoprotein Lipase Deficiency, Familial, Lissencephaly, Marfan Syndrome, MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis, and, Stroke-Like Episodes), Monosomies, Multiple Endocrine Neoplasia Type 2, Multiple Exostoses, Hereditary

Muscular Dystrophy, Congenital, Myotonic Dystrophy, Nephrogenic Diabetes Insipidus, Neurofibromatosis 1, Neurofibromatosis 2, Neuropathy with Liability to Pressure Palsies, Hereditary, Niemann-Pick Disease Type C, Nijmegen Breakage Syndrome, Norrie Disease, Oculocutaneous Albinism Type 1, Oculopharyngeal Muscular Dystrophy, Ovarian Cancer, Pallister-Hall Syndrome, Parkin Type of Juvenile Parkinson Disease, Pelizaeus-Merzbacher Disease, Pendred Syndrome, Peutz-Jeghers Syndrome, Phenylalanine Hydroxylase Deficiency, Prader-Willi Syndrome, PROP 1-Related Combined Pituitary Hormone Deficiency (CPHD), Prostate Cancer, Retinitis Pigmentosa, Retinoblastoma, Rothmund-Thorns on Syndrome, Smith-Lemli-Opitz Syndrome, Spastic Paraplegia, Hereditary, Spinal and Bulbar Muscular Atrophy (also Kennedy Disease), Spinal Muscular Atrophy, Spinocerebellar Ataxia Type 1, Spinocerebellar Ataxia Type 2, Spinocerebellar Ataxia Type 3, Spinocerebellar Ataxia Type 6, Spinocerebellar Ataxia Type 7, Stickler Syndrome (Hereditary Arthroophthalmopathy), Tay-Sachs (also GM2 Gangliosidosis), Trisomies, Tuberous Sclerosis Complex, Usher Syndrome Type I, Usher Syndrome Type II, Velocardiofacial Syndrome (also 22q11 Deletion Syndrome), Von Hippel-Lindau Syndrome, Williams Syndrome, Wilson Disease, X-Linked Adrenoleukodystrophy, X-Linked Agammaglobulinemia, X-Linked Dilated Cardiomyopathy (also The Dystrophinopathies), and X-Linked Hypotonic Facies Mental Retardation Syndrome. Methods, provided herein can also be applied to detect mutations during the course of therapeutic intervention in the treatment of disease that can decide the course and efficacy of a given treatment.

[0074] The methods may further be used to measure the efficacy of a treatment. In such cases samples from a patient will be monitored over a course of time for changes in the characteristic curves. Similarly, the methods described herein may be used to monitor the course of a disease.

[0075] The methods disclosed herein may be used with subjects at risk for developing a disease or disorder, subjects who may or may not have already been diagnosed with a disease or disorder and subjects undergoing treatment and/or therapies for a disease or disorder. The methods of the present disclosure can also be used to monitor or select a treatment regimen for a subject who has a disease or disorder.

[0076] The disclosure also provides methods for detecting the presence of one or more pathogens. In some cases, the methods may be used to determine the presence of one or more viruses, bacteria, or fungi. For example, the methods may be used for detection of DNA viruses, RNA viruses, retroviruses. In some cases the methods may be used to detect/quantify

dsDNA viruses, ssDNA viruses, dsRNA viruses, (+)ssRNA viruses, (-)ssRNA viruses, ssRNA-RT viruses or dsDNA-RT viruses. The method may be used to detect/quantify gram-positive (for example, gram-positive cocci and gram-positive bacilli) and gram-negative (for example, gram-negative cocci and gram-negative bacilli). These methods may further include extracting cellular components of the target pathogens such as genomic DNA, ribosomal RNA, transfer RNA, or messenger RNA. In these methods, the presence of each pathogen in a sample will be indicated by the corresponding characteristic curve. Non limiting examples of pathogens that may be detected by the methods of the disclosure include HIV-1, hepatitis B and C, human papillomavirus, chlamydia trachomatis, neisseria gonorrhoeae, cytomegalovirus, influenza, salmonella typhi, west nile virus, adenovirus, actinomyces, and mycobacterium tuberculosis. The methods of the disclosure also enable the identification of GMO (genetic modified organisms).

[0077] The methods of the disclosure would also be useful in tracking the presence and/or progression of mutations in patients and related model organisms. In some cases the patients may be cancer patients and the methods of the disclosure may be used to obtain information about the presence or absence of drug-resistant mutations. Such information may be of clinical meaning and utility, for e.g. a KRAS mutation may indicate colorectal cancer, and a T790M mutation of EGFR may indicate lung cancer. Similarly, the methods of the invention may be used to track genetic information in a model organism or a population of model organisms as they are exposed to different experimental drugs in pharmaceutical research. In some cases the model organism may be a mouse or a rat.

[0078] The methods of the disclosure may also be employed in genetic retroactive studies of archival tissue cross-referenced with patient histories and outcomes and implemented medication regimens. Such studies may generate new insights both in epidemiology and in fundamental understanding of diseases, for example cancer. The methods disclosed herein further provide for a means of multiplexed expression analysis in biological samples, e.g. blood, saliva, cerebrospinal fluid, etc., which can have both fundamental and clinical utilities.

V. Amplification Reactions

A. Polymerase Chain Reaction (PCR)

[0079] Amplification refers to any process by which the copy number of a target sequence is increased. Methods for amplification of target polynucleotides are known in the art, and include without limitation, methods based on PCR. Conditions favorable to the amplification of target sequences by PCR are known in the art, can be optimized at a variety of steps in the

process, and depend on characteristics of elements in the reaction, such as target type, target concentration, sequence length to be amplified, sequence of the target and/or one or more primers, primer length, primer concentration, polymerase used, reaction volume, ratio of one or more elements to one or more other elements, and others, some or all of which can be altered.

[0080] In general, PCR involves the steps of denaturation of the target to be amplified (if double stranded), hybridization of one or more primers to the target, and extension of the primers by a DNA polymerase, with the steps repeated (or “cycled”) in order to amplify the target sequence. Steps in this process can be optimized for various outcomes, such as to enhance yield, decrease the formation of spurious products, and/or increase or decrease specificity of primer annealing.

[0081] Methods of optimization are well known in the art and include adjustments to the type or amount of elements in the amplification reaction and/or to the conditions of a given step in the process, such as temperature at a particular step, duration of a particular step, and/or number of cycles. In some embodiments, an amplification reaction comprises at least 5, 10, 15, 20, 25, 30, 35, 50, or more cycles. In some embodiments, an amplification reaction comprises no more than 5, 10, 15, 20, 25, 35, 50, or more cycles. Cycles can contain any number of steps, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more steps. Steps can comprise any temperature or gradient of temperatures, suitable for achieving the purpose of the given step, including but not limited to, 3' end extension (e.g. adaptor fill-in), primer annealing, primer extension, and strand denaturation. Steps can be of any duration, including but not limited to about, less than about, or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 120, 180, 240, 300, 360, 420, 480, 540, 600, or more seconds, including indefinitely until manually interrupted. Cycles of any number comprising different steps can be combined in any order. In some embodiments, different cycles comprising different steps are combined such that the total number of cycles in the combination is about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 50, or more cycles. In some embodiments, amplification is performed following the fill-in reaction. Amplification can be performed before or after pooling of target polynucleotides from independent samples.

(i) Nested PCR reactions

[0082] In the methods of the present disclosure PCR reaction may be carried out in a traditional fashion where one set of specific primers is chosen for a signal nucleic acid target. However, strategies may also be used to increase PCR target yield. For example the

amplification reaction may be conducted under nested PCR conditions, wherein at least two pairs of primers are used, that is, a pair of outer primers and a pair of inner primers. Each primer pair comprises a forward primer and a reverse primer. Conventional nested PCR procedures utilize two sequential amplification processes. Specifically, the two sequential amplification processes include a first amplification process comprising at least one amplification step for amplifying an extended target sequence, and a second, subsequent amplification process comprising at least one amplification step for amplifying an internal sequence from the product of the first amplification process, wherein the internal sequence may or may not overlap one of the ends of the extended sequence. Each amplification step of the first amplification process employs an outer primer set typically comprising a pair of outer primers. Similarly, each amplification step of the second amplification process employs an inner primer set typically comprising a pair of inner primers. Figure 1 shows a diagram illustrating the methods of nested PCR reactions and Figure 2 shows a nested PCR with three primer pairs (A, B, C). The targeted area of interest is shown in the box. As used herein, nested PCR reactions are also meant to include semi-nested or hemi-nested PCR reactions wherein at least one of the primers used in the second amplification process is same as in the first amplification process.

[0083] The nested PCR methods of the present disclosure may be performed using plurality of primers, for example at least 3 nested primers, at least 4 nested primers, at least 5 nested primers, at least 6 nested primers, at least 7 nested primers, at least 8 nested primers, at least 9 nested primers, at least 10 nested primers, at least 11 nested primers, at least 12 nested primers, at least 13 nested primers, at least 14 nested primers, or at least 15 nested primers. In some example the methods may employ more than 15 nested primers for example 20, 25, 30, 35 or even more nested primers. In some embodiments the nested PCR methods of the present disclosure may be performed using at most 35 nested primers, 30 nested primers, 25 nested primers, 20 nested primers, 15 nested primers, 14 nested primers, 13 nested primers, 12 nested primers, 11 nested primers, 10 nested primers, 9 nested primers, 8 nested primers, 7 nested primers, 6 nested primers, 5 nested primers, 4 nested primers, 3 nested primers or 2 nested primers. In some example the methods may 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 nested primers.

[0084] The plurality of primers may comprise at least 2 pairs of nested primers, at least 3 pairs of nested primers, at least 4 pairs of nested primers, at least 5 pairs of nested primers, at least 6 pairs of nested primers, at least 7 pairs of nested primers, at least 8 pairs of nested

primers, at least 9 pairs of nested primers, at least 10 pairs of nested primers, at least 11 pairs of nested primers, at least 12 pairs of nested primers, at least 13 pairs of nested primers, at least 14 pairs of nested primers, at least 15 pairs of nested primers, at least 16 pairs of nested primers, at least 17 pairs of nested primers, at least 18 pairs of nested primers, at least 19 pairs of nested primers or at least 20 pairs of nested primers. In some embodiments the plurality of primers may comprise at most 20 pairs of nested primers, 19 pairs of nested primers, 18 pairs of nested primers, 17 pairs of nested primers, 16 pairs of nested primers, 15 pairs of nested primers, 14 pairs of nested primers, 13 pairs of nested primers, 12 pairs of nested primers, 11 pairs of nested primers, 10 pairs of nested primers, 9 pairs of nested primers, 8 pairs of nested primers, 7 pairs of nested primers, 6 pairs of nested primers, 5 pairs of nested primers, 4 pairs of nested primers, 3 pairs of nested primers, or 2 pairs of nested primers.

[0085] For instance, as shown in Figure 2, three nested primer pairs A, B, and C may be used. In this case, all three primers are effective at amplifying the targeted area of interest. By choosing primer sequences carefully for binding energy, they may exhibit different melt temperatures. This will allow the reaction kinetics for each primer to be tuned as a function of temperature.

[0086] By choosing the temperatures of the nested primer pairs, and by controlling the annealing temperature of the PCR amplification, the fluorescence signature of the amplification reaction may be tuned to generate multiple plateaus. For instance, as shown in Figure 5, the three nested primer pairs from Figure 2 may generate a unique fluorescence signature. Because the concentrations of primer sets A, B, and C may be carefully selected at the beginning of the reaction, the concentration of the nucleic acid target at plateaus A, B, and C may also be deterministic. This information may be used to select at which concentration of targets the linear PCR regions should be run.

(ii) Asymmetric PCR reactions

[0087] The nested PCR reactions of the present disclosure may further be run in such a way that either the forward or reverse primer in the nested primer pairs may be rate limited (asymmetric PCR conditions). This may be done by adding each primer set such that there are different ratios of reverse and forward primers. In some cases this may be done by titrating each primer set such that there are different ratios of reverse and forward primers. Accordingly, it may be possible to generate curves that exhibit a characteristic linear amplification of DNA. While this is linear amplification may be non-ideal for the efficient

amplification of DNA, it may be suitable for the careful discrimination of DNA targets. Figure 3 shows the fluorescence curve for a saturated, quantitative PCR reaction and Figure 4 shows the fluorescence curve of an asymmetric (biased primer) quantitative PCR reaction. Furthermore, tuning of primer concentrations and the number of primers for a particular nested reaction may allow for the engineering of particular fluorescent signatures for a given target.

[0088] In certain embodiments, the starting molar concentration of the limiting primer is less than the starting molar concentration of the excess primer. The ratio of the starting concentrations of the excess primer and the limiting primer may be at least 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 25:1, 30:1, 35:1, 40:1, 45:1, 50:1, 55:1, 60:1, 65:1, 70:1, 75:1, 80:1, 85:1, 90:1, 95:1, 100:1, 150:1, 200:1, 250:1, 300:1, 350:1, 400:1, 450:1, 500:1, 550:1, 600:1, 650:1, 700:1, 750:1, 800:1, 850:1, 900:1, 950:1, 1000:1 or more.

[0089] In certain embodiments, the starting molar concentration of the limiting primer is less than the starting molar concentration of the excess primer. The ratio of the starting concentrations of the excess primer and the limiting primer may be at most 1000:1, 950:1, 900:1, 850:1, 800:1, 750:1, 700:1, 650:1, 600:1, 550:1, 500:1, 450:1, 400:1, 350:1, 300:1, 250:1, 200:1, 150:1, 100:1, 95:1, 90:1, 85:1, 80:1, 75:1, 70:1, 65:1, 60:1, 55:1, 50:1, 45:1, 40:1, 35:1, 30:1, 25:1, 20:1, 19:1, 18:1, 17:1, 16:1, 15:1, 14:1, 13:1, 12:1, 11:1, 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1, 3:1, 2:1, 1:1 or less.

[0090] In certain embodiments, the forward primer comprises a limiting primer and the reverse primer comprises an excess primer, wherein the excess primer is added to the combined sample at a concentration at least 2 times, 3 times, 4 times, 5 times, 6 times, 7 times, 8 times, 9 times, 10 times, 11 times, 12 times, 13 times, 14 times, 15 times, 16 times, 17 times, 18 times, 19 times, 20 times, 21 times, 22 times, 23 times, 24 times, 25 times, 30 times, 35 times, 40 times, 45 times, 50 times, 55 times, 60 times, 65 times, 70 times, 75 times, 80 times, 85 times, 90 times, 95 times, 100 times, 150 times, 200 times, 250 times, 300 times, 350 times, 400 times, 450 times, 500 times, 550 times, 600 times, 650 times, 700 times, 750 times, 800 times, 850 times, 900 times, 950 times, or 1000 times that of the limiting primer, and wherein the amplification conditions comprise asymmetric PCR conditions.

[0091] In certain embodiments, the forward primer comprises a limiting primer and the reverse primer comprises an excess primer, wherein the excess primer is added to the

combined sample at a concentration at most 1000 times, 900 times, 950 times, 850 times, 800 times, 750 times, 700 times, 650 times, 600 times, 550 times, 500 times, 450 times, 400 times, 350 times, 300 times, 250 times, 200 times, 150 times, 100 times, 95 times, 90 times, 85 times, 80 times, 75 times, 70 times, 65 times, 60 times, 55 times, 50 times, 45 times, 40 times, 35 times, 30 times, 25 times, 24 times, 23 times, 22 times, 21 times, 20 times, 19 times, 18 times, 17 times, 16 times, 15 times, 14 times, 13 times, 12 times, 11 times, 10 times, 9 times, 8 times, 7 times, 6 times, 5 times, 4 times, 3 times, or 2 times that of the limiting primer, and wherein the amplification conditions comprise asymmetric PCR conditions.

[0092] In some embodiments, the forward primer comprises an excess primer and the reverse primer comprises a limiting primer, wherein the excess primer is added to the combined sample at a concentration at least 2 times, 3 times, 4 times, 5 times, 6 times, 7 times, 8 times, 9 times, 10 times, 11 times, 12 times, 13 times, 14 times, 15 times, 16 times, 17 times, 18 times, 19 times, 20 times, 21 times, 22 times, 23 times, 24 times, 25 times, 30 times, 35 times, 40 times, 45 times, 50 times, 55 times, 60 times, 65 times, 70 times, 75 times, 80 times, 85 times, 90 times, 95 times, 100 times, 150 times, 200 times, 250 times, 300 times, 350 times, 400 times, 450 times, 500 times, 550 times, 600 times, 650 times, 700 times, 750 times, 800 times, 850 times, 900 times, 950 times, or 1000 times that of the limiting primer, and wherein the amplification conditions comprise asymmetric PCR conditions.

[0093] In certain embodiments, the forward primer comprises an excess primer and the reverse primer comprises a limiting primer, wherein the excess primer is added to the combined sample at a concentration at most 1000 times, 900 times, 950 times, 850 times, 800 times, 750 times, 700 times, 650 times, 600 times, 550 times, 500 times, 450 times, 400 times, 350 times, 300 times, 250 times, 200 times, 150 times, 100 times, 95 times, 90 times, 85 times, 80 times, 75 times, 70 times, 65 times, 60 times, 55 times, 50 times, 45 times, 40 times, 35 times, 30 times, 25 times, 24 times, 23 times, 22 times, 21 times, 20 times, 19 times, 18 times, 17 times, 16 times, 15 times, 14 times, 13 times, 12 times, 11 times, 10 times, 9 times, 8 times, 7 times, 6 times, 5 times, 4 times, 3 times, or 2 times that of the limiting primer, and wherein the amplification conditions comprise asymmetric PCR conditions.

[0094] Primer length and sequence may be adjusted or modified, such that the concentration-adjusted melting temperature of the limiting primer at the start of the reaction is greater than or equal to the concentration-adjusted melting point of the excess primer. In some

embodiments, the difference in the concentration-adjusted melting temperature of the limiting and the excess primer may be at least 0 °C, 0.5 °C, 1 °C, 1.5 °C, 2 °C, 2.5 °C, 3 °C, 3.5 °C, 4 °C, 4.5 °C, 5 °C, 5.5 °C, 6 °C, 6.5 °C, 7 °C, 7.5 °C, 8 °C, 8.5 °C, 9 °C, 9.5 °C, 10 °C or more. In some embodiments, the difference in the concentration-adjusted melting temperature of the limiting and the excess primer may be at most 10 °C, 9.5 °C, 9 °C, 8.5 °C, 8 °C, 7.5 °C, 7 °C, 6.5 °C, 6 °C, 5.5 °C, 5 °C, 4.5 °C, 4 °C, 3.5 °C, 3 °C, 2.5 °C, 2 °C, 1.5 °C, 1.0 °C, 0.5 °C, 0 °C or less.

[0095] Amplifications and assays according to embodiments of the present disclosure may be performed with initial reaction mixtures having ranges of concentrations of target molecules and primers. The methods may be used to assay samples containing less than 100 copies, 1000 copies, 10,000 copies, 20,000 copies, 30,000 copies, 40,000 copies, 50,000 copies, 60,000 copies, 70,000 copies, 80,000 copies, 90,000 copies, 100,000 copies, 200,000 copies, 300,000 copies, 400,000 copies, 500,000 copies, 600,000 copies, 70,000 copies, 800,000 copies, 900,000 copies or 1000,000 copies of the target analyte.

[0096] The methods may be used to assay samples containing more than 1000,000 copies, 900,000 copies, 800,000 copies, 700,000 copies, 600,000 copies, 500,000 copies, 400,000 copies, 300,000 copies, 200,000 copies, 100,000 copies, 90,000 copies, 80,000 copies, 70,000 copies, 60,000 copies, 50,000 copies, 40,000 copies, 30,000 copies, 20,000 copies, 10,000 copies, 1,000 copies or 100 copies of the target analyte.

[0097] The methods may be employed to assay sample comprising 0-100 copies, 0-1000 copies, 0-10,000 copies, 0-20,000 copies, 0-30,000 copies, 0-40,000 copies, 0-50,000 copies, 0-60,000 copies, 0-70,000 copies, 0-80,000 copies, 0-90,000 copies, 0-100,000 copies, 0-200,000 copies, 0-300,000 copies, 0-400,000 copies, 0-500,000 copies, 0-600,000 copies, 0-70,000 copies, 0-800,000 copies, 0-900,000 copies or 0-1000,000 copies of the target analyte.

[0098] The concentration of the limiting primer may range from a few to a several hundred nanomolars (nM). For example, the concentration of the limiting primer may be in the range of 1-50 nM, 1-100 nM, 1-150 nM, 1-200 nM, 1-250 nM, 1-300 nM, 1-350 nM, 1-400 nM, 1-450 nM, 1-500 nM, 50-100 nM, 50-150 nM, 50-200 nM, 50-250 nM, 50-300 nM, 50-350 nM, 50-400 nM, 50-450 nM, 50-500 nM, 100-150 nM, 100-200 nM, 100-250 nM, 100-300 nM, 100-350 nM, 100-400 nM, 100-450 nM, 100-500 nM, 150-200 nM, 150-250 nM, 150-300 nM, 150-350 nM, 150-400 nM, 150-450 nM, 150-500 nM, 200-250 nM, 200-300 nM, 200-350 nM, 200-400 nM, 200-450 nM, 200-500 nM, 250-300 nM, 250-350 nM, 250-400 nM, 250-450 nM, 250-500 nM, 300-350 nM, 300-400 nM, 300-450 nM, 300-500 nM, 350-

400 nM, 350-450 nM, 350-500 nM, 400-450 nM, 400-500 nM, or 450-500 nM. Those of skill in the art appreciate that the concentration of the limiting primer may fall within any range bounded by any of these values.

[0099] In some examples the concentration of the limiting primer is less than 10 nM, 50 nM, 100 nM, 150 nM, 200 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM or 500 nM. In some examples the concentration of the limiting primer is greater than 10 nM, 50 nM, 100 nM, 150 nM, 200 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM or 500 nM.

[00100] The asymmetric nested PCR methods of the present disclosure include repeated thermal cycling through the steps of strand melting, primer annealing and primer extension. In a first-stage polymerase chain reaction, specific outer primers are added to the sample nucleic acid mixture, and the resulting mixture is subjected to an initial denaturation step to obtain single-stranded DNA templates. Following denaturation, the mixture is subjected to an initial annealing step, where the outer primers hybridize to opposite strands of the target nucleotide sequence. The temperature is then raised to allow for extension or replication of the specific segment of DNA across the region between the two primers by a thermostable DNA polymerase. The reaction is then thermocycled to allow for repeated denaturation, annealing, and extension, so that at each cycle, the amount of DNA representing the targeted nucleotide sequence between the two outer primers is doubled.

[00101] The first denaturation temperature may be in the range of 70-110 °C, for example in the range of 70-75°C, 75-80°C, 80-85°C, 85-90°C, 90-95 °C, 95-100 °C, 100-105 °C, or 105-110 °C. The first annealing temperature is may be in the range of in the range of 35-65 °C, for example in the range of 35-40 °C, 40-45 °C, 45-50 °C, 50-55 °C, 55-60 °C, or 60-65 °C. The first extension temperature may be in the range of in the range of 40-80 °C, for example in the range of 40-45 °C, 45-50 °C, 50-55 °C, 55-60 °C, 60-65 °C, 65-70 °C, 70-75 °C, or 75-80 °C.

[00102] At the end of the first-stage polymerase chain reaction, the resulting mixture is carried over into a second-stage polymerase chain reaction. In this second-stage reaction, the products of the first-stage reaction are combined with specific inner primers. This mixture is again subjected to initial denaturation, annealing, and extension steps, followed by thermocycling as before to allow for repeated denaturation, annealing, and extension or the targeted nucleotide sequence.

[00103] The second denaturation temperature may be in the range of 70-110 °C, for example in the range of 70-75°C, 75-80°C, 80-85°C, 85-90°C, 90-95 °C, 95-100 °C, 100-105

°C, or 105-110 °C. The second annealing temperature may be in the range of 35-65 °C, for example in the range of 35-40 °C, 40-45 °C, 45-50 °C, 50-55 °C, 55-60 °C, or 60-65 °C. The second extension temperature may be in the range of in the range of 40-80 °C, for example in the range of 40-45 °C, 45-50 °C, 50-55 °C, 55-60 °C, 60-65 °C, 65-70 °C, 70-75 °C, or 75-80 °C.

[00104] Other PCR techniques that can be used in the methods of the provided disclosure include, e.g., AFLP (amplified fragment length polymorphism) PCR (see e.g.: Vos et al. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-14), allele-specific PCR (see e.g., Saiki R K, Bugawan T L, Horn G T, Mullis K B, Erlich H A (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes *Nature* 324: 163-166), Alu PCR, assembly PCR (see e.g., Stemmer W P, Cramer A, Ha K D, Brennan T M, Heyneker H L (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides *Gene* 164: 49-53), asymmetric colony PCR, helicase dependent PCR (see e.g., Myriam Vincent, Yan Xu and Huimin Kong (2004). Helicase-dependent isothermal DNA amplification *EMBO reports* 5 (8): 795-800), hot start PCR, inverse PCR (see e.g., Ochman H, Gerber A S, Hartl D L. *Genetics*. 1988 November; 120(3):621-3), in situ PCR, intersequence-specific PCR or IS SR PCR, digital PCR, linear-after-the-exponential-PCR or Late PCR (see e.g., Pierce K E and Wagh L T (2007). Linear-after-the-exponential polymerase chain reaction and allied technologies Real-time detection strategies for rapid, reliable diagnosis from single cells *Methods Mol. Med.* 132: 65-85), long PCR, nested PCR, real-time PCR, duplex PCR, multiplex PCR, quantitative PCR, or single cell PCR. Further examples of PCR techniques that may be used with the methods of the disclosure include, but are not limited to, quantitative PCR, quantitative fluorescent PCR (QF-PCR), multiplex fluorescent PCR (MF-PCR), real time PCR, reverse transcriptase PCR (RT-PCR), single cell PCR, restriction fragment length polymorphism PCR (PCR-RFLP), PCR-RFLP/RT-PCR-RFLP, hot start PCR, nested PCR, in situ polonony PCR, in situ rolling circle amplification (RCA), bridge PCR, picotiter PCR and emulsion PCR. In some examples, the nested PCR reactions of the disclosure are conducted by having all primers added at the very beginning of PCR reaction and by programming the PCR machine to cycle among different sets of PCR temperatures over the course of the reaction, so that different sets of primers are activated in their respective phases only. For example, a nested PCR reaction may be carried out with three pairs of nested primers, say A, B, and C present at the

same time. The reaction may be programmed in such a way that it runs for the first third of the cycles at hybridization temperature for the primers of sequence A, which is too high for the primers of B and C to hybridize. The second third of the cycles is then run at the temperature suitable for the primers of B but still too hot for hybridization of primer C. Finally, the final third of the cycles is run at temperature low enough for the primer pair C to hybridize. The lever of control in designing the methods of the disclosure include the sequence information and the programming of the temperatures and durations of the PCR cycles. The methods of the invention may in some cases involve varying the annealing and extension temperatures during the reaction, so that different primers get activated for extension at different times during the same PCR reaction.

[00105] In some cases, it may also be possible to program Taqman probes to activate at different times in the reaction as well. For example, the outer nesting primers may have annealing temperatures that are higher than the inner nested primers and the Taqman probe. This may mean that the Taqman probes may be protected from hydrolysis during the “pre-amplification”, because while the template is being amplified, the probe is not hybridized and thus cannot be hydrolyzed. As a result no signal would be generated during the first stage of amplification, but when the inner pair of primers is engaged by lowering the annealing temperature, the probe would hybridize as well, and signal generation will be commenced. This characteristic may be used to keep the signal clean and help diagnose problems with the assay.

B. SPIA Amplification

[00106] Amplification of the analytes of interest employing a linear amplification method such as the single primer isothermal amplification (SPIA) can be used. SPIA enables generation of multiple copies of the strand specific sequence regions of interest and employs a single amplification primer, thus reducing the complexity associated with multiple oligonucleotide design and manufacturing, enables the use of a generic amplification primer, and can be linear.

[00107] Amplification by SPIA can occur under conditions permitting composite primer hybridization, primer extension by a DNA polymerase with strand displacement activity, cleavage of RNA from a RNA/DNA heteroduplex and strand displacement. In so far as the composite amplification primer hybridizes to the 3'-single-stranded portion (of the partially double stranded polynucleotide which is formed by cleaving RNA in the complex comprising a RNA/DNA partial heteroduplex) comprising, generally, the complement of at

least a portion of the composite amplification primer sequence, composite primer hybridization may be under conditions permitting specific hybridization. In SPIA, all steps are isothermal (in the sense that thermal cycling is not required), although the temperatures for each of the steps may or may not be the same. It is understood that various other embodiments can be practiced given the general description provided above. For example, as described and exemplified herein, certain steps may be performed as temperature is changed (e.g., raised, or lowered).

[00108] Although generally only one composite amplification primer is described above, it is further understood that the SPIA amplification methods can be performed in the presence of two or more different first and/or second composite primers that randomly primer template polynucleotide. In addition, the amplification polynucleotide products of two or more separate amplification reactions conducted using two or more different first and/or second composite primers that randomly prime template polynucleotide can be combined.

[00109] The composite amplification primers are primers that are composed of RNA and DNA portions. In the amplification composite primer, both the RNA and the DNA portions are generally complementary and can hybridize to a sequence in the amplification-ready product to be copied or amplified. In some embodiments, a 3'-portion of the amplification composite primer is DNA and a 5'-portion of the composite amplification primer is RNA. The composite amplification primer is designed such that the primer is extended from the 3'-DNA portion to create a primer extension product. The 5'-RNA portion of this primer extension product in a RNA/DNA heteroduplex is susceptible to cleavage by RNase H, thus freeing a portion of the polynucleotide to the hybridization of an additional composite amplification primer. The extension of the amplification composite primer by a DNA polymerase with strand displacement activity releases the primer extension product from the original primer and creates another copy of the sequence of the polynucleotide. Repeated rounds of primer hybridization, primer extension with strand displacement DNA synthesis, and RNA cleavage create multiple copies of the strand-specific sequence of the polynucleotide.

[00110] In some embodiments, the composite amplification primer is generated in the amplification reaction mixture from a stem-loop chimeric pro-primer. The amplification reaction mixture can comprise a target partial duplex nucleic acid, for example a target partial duplex DNA, a chimeric stem-loop pro-primer, DNA polymerase with strand displacement activity, and an RNase targeting RNA in a RNA/DNA heteroduplex, for example RNase H.

The RNA portion of the RNA/DNA heteroduplex at the stem of the chimeric stem-loop primer can be cleaved by RNase H to generate, for example, a linear composite primer comprising a 3'-DNA and 5'-RNA. The linearized amplification primer can hybridize to a 3'-single stranded DNA portion (overhang) of a target partial duplex and can be extended by the DNA polymerase with strand displacement activity. The RNA portion of the hybridized primer in a heteroduplex can be cleaved by RNase H to free a portion of the primer binding site. A second linear composite amplification primer can hybridize to the freed primer binding site, and can be extended along the target DNA strand. The previously synthesized primer extension product (amplification product) can be displaced by the newly extended primer. Repeated cycles of primer hybridization, primer extension by strand displacement DNA polymerase, and cleavage of the RNA portion of the hybridized primer can generate multiple copies of a target nucleic acid.

C. Other Amplification Methods

[00111] Another method for amplification involves amplification of a single stranded polynucleotide using a single oligonucleotide primer. The single stranded polynucleotide that is to be amplified contains two non-contiguous sequences that are substantially or completely complementary to one another and, thus, are capable of hybridizing together to form a stem-loop structure. This single stranded polynucleotide already may be part of a polynucleotide analyte or may be created as the result of the presence of a polynucleotide analyte.

[00112] Another method for achieving the result of an amplification of nucleic acids is known as the ligase chain reaction (LCR). This method uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[00113] Another method for achieving nucleic acid amplification is the nucleic acid sequence based amplification (NASBA). This method is a promoter-directed, enzymatic process that induces in vitro continuous, homogeneous and isothermal amplification of a specific nucleic acid to provide RNA copies of the nucleic acid. The reagents for conducting NASBA include a first DNA primer with a 5'-tail comprising a promoter, a second DNA primer, reverse transcriptase, RNase-H, T7 RNA polymerase, NTP's and dNTP's.

[00114] Another method for amplifying a specific group of nucleic acids is the Q-beta-replicase method, which relies on the ability of Q-beta-replicase to amplify its RNA substrate

exponentially. The reagents for conducting such an amplification include “midi-variant RNA” (amplifiable hybridization probe), NTP's, and Q-beta-replicase.

[00115] Another method for amplifying nucleic acids is known as 3SR and is similar to NASBA except that the RNase-H activity is present in the reverse transcriptase. Amplification by 3SR is an RNA specific target method whereby RNA is amplified in an isothermal process combining promoter directed RNA polymerase, reverse transcriptase and RNase H with target RNA. See for example Fahy et al. PCR Methods Appl. 1:25-33 (1991).

[00116] Another method for amplifying nucleic acids is the Transcription Mediated Amplification (TMA) used by Gen-Probe. The method is similar to NASBA in utilizing two enzymes in a self-sustained sequence replication. See U.S. Pat. No. 5,299,491 herein incorporated by reference.

[00117] Another method for amplification of nucleic acids is Strand Displacement Amplification (SDA) (Westin et al 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), which is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3'-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[00118] Another method for amplification of nucleic acids is Rolling Circle Amplification (RCA) (Lizardi et al. 1998, Nature Genetics, 19:225-232). RCA can be used to amplify single stranded molecules in the form of circles of nucleic acids. In its simplest form, RCA involves the hybridization of a single primer to a circular nucleic acid. Extension of the primer by a DNA polymerase with strand displacement activity results in the production of multiple copies of the circular nucleic acid concatenated into a single DNA strand.

[00119] In some embodiments of the disclosure, RCA is coupled with ligation. For example, a single oligonucleotide can be used both for ligation and as the circular template for RCA. This type of polynucleotide can be referred to as a “padlock probe” or a “RCA probe.” For a padlock probe, both termini of the oligonucleotide contain sequences complementary to a domain within a nucleic acid sequence of interest. The first end of the padlock probe is substantially complementary to a first domain on the nucleic acid sequence

of interest, and the second end of the padlock probe is substantially complementary to a second domain, adjacent to the first domain near the first domain. Hybridization of the oligonucleotide to the target nucleic acid results in the formation of a hybridization complex. Ligation of the ends of the padlock probe results in the formation of a modified hybridization complex containing a circular polynucleotide. In some cases, prior to ligation, a polymerase can fill in the gap by extending one end of the padlock probe. The circular polynucleotide thus formed can serve as a template for RCA that, with the addition of a polymerase, results in the formation of an amplified product nucleic acid. The methods of the disclosure described herein can produce amplified products with defined sequences on both the 5'- and 3'-ends. Such amplified products can be used as padlock probes.

[00120] Some aspects of the disclosure utilize the linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

[00121] In some embodiments, amplification methods can be solid-phase amplification, polony amplification, colony amplification, emulsion PCR, bead RCA, surface RCA, surface SDA, etc., as will be recognized by one of skill in the art. In some embodiments, amplification methods that results in amplification of free DNA molecules in solution or tethered to a suitable matrix by only one end of the DNA molecule can be used. Methods that rely on bridge PCR, where both PCR primers are attached to a surface (see, e.g., WO 2000/018957 and Adessi et al., *Nucleic Acids Research* (2000): 28(20): E87) can be used. In some cases the methods of the disclosure can create a "polymerase colony technology," or "polony," referring to a multiplex amplification that maintains spatial clustering of identical amplicons (see Harvard Molecular Technology Group and Lipper Center for Computational Genetics website). These include, for example, in situ polonies (Mitra and Church, *Nucleic Acid Research* 27, e34, Dec. 15, 1999), in situ rolling circle

amplification (RCA) (Lizardi et al., Nature Genetics 19, 225, July 1998), bridge PCR (U.S. Pat. No. 5,641,658), picotiter PCR (Leamon et al., Electrophoresis 24, 3769, November 2003), and emulsion PCR (Dressman et al., PNAS 100, 8817, Jul. 22, 2003). The methods of the disclosure provide new methods for generating and using colonies.

[00122] Other suitable amplification methods include transcription amplification, self-sustained sequence replication, selective amplification of target polynucleotide sequences, consensus sequence primed polymerase chain reaction (CP-PCR), arbitrarily primed polymerase chain reaction (AP-PCR), degenerate oligonucleotide-primed PCR (DOP-PCR) and nucleic acid based sequence amplification (NABSA). Other amplification methods that can be used herein include those described in U.S. Pat. Nos. 5,242,794; 5,494,810; 4,988,617; and 6,582,938. In some embodiments, amplification of target nucleic acids may occur on a bead. In other embodiments, amplification does not occur on a bead.

Multiplex Assays

[00123] In some embodiments, the present disclosure provides a multiplexed assay for simultaneous amplification, detection, and or/quantification of multiple target analytes in a sample. Multiplex assays or grammatical equivalents refer to the detection, analysis or amplification of more than one target analyte of interest in a single reaction mixture. In some embodiments the methods of the disclosure may be used to detect and/or quantify at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000, 2000, 3000, or more different target analytes in a sample. In some embodiments the methods of the disclosure may be used to detect and/or quantify at most 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000, 2000, 3000, or less different target analytes a sample. Accordingly, in some cases, the methods described herein may be used to simultaneously detect multiple diseases at once. In other cases the methods may be used to simultaneously detect multiple infectious agents in a sample. In some other cases the methods may be used to simultaneously detect multiple genetic markers in a sample.

[00124] In some embodiments the methods of the disclosure involve multiplexed PCR reactions. In some embodiments the methods may involve multiplexed nested PCR reactions or multiplexed asymmetric PCR reactions.

VI. Analytes

[00125] An analyte can be any molecule detectable by the methods provided herein. An analyte may be a polynucleotide, such as DNA, RNA, and any hybrid thereof, where the

polynucleotide contains any combination of deoxyribo- and/or ribo-nucleotides. Polynucleotides may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. Polynucleotides may contain any combination of nucleotides, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine, hypoxanthine, isocytosine, isoguanine and any nucleotide derivative thereof. As used herein, the term “nucleotide” may include nucleotides and nucleosides, as well as nucleoside and nucleotide analogs, and modified nucleotides, including both synthetic and naturally occurring species.

[00126] The analyte may comprise RNA. For example an analyte may comprise messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), nuclear RNA (nRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), small Cajal body-specific RNA (scaRNA), microRNA (miRNA), double stranded (dsRNA), ribozyme, riboswitch 16SRNA or viral RNA. In some cases the methods of the invention may further comprise reverse transcription of the RNA analyte into its DNA complement (cDNA) through the use of reverse transcriptase. Subsequently, the cDNA may be amplified and assayed by the methods of the disclosure.

[00127] The analyte may comprise DNA, or complex DNA, for example genomic DNA. Analytes comprising genomic DNA may be obtained from any source and using any methods known in the art. Genomic DNA may be obtained from naturally occurring or genetically modified organisms or from artificially or synthetically created genomes. Genomic DNA may be isolated with or without amplification. Amplification may include PCR amplification, multiple displacement amplification (MDA), rolling circle amplification and other amplification methods. Genomic DNA may also be obtained by cloning or recombinant methods, such as those involving plasmids and artificial chromosomes or other conventional methods (see Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, cited *supra*.). The analyte may also comprise cDNA. The cDNA can be generated from RNA, e.g., mRNA. The DNA can be cDNA made from a mixture of genomes of different species. The DNA can be mitochondrial DNA (mtDNA). The DNA can be cell-free DNA. The cell-free DNA can be obtained from, e.g., a serum or plasma sample. The DNA can comprise one or more chromosomes. For example, if the DNA is from a human, the DNA can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. The DNA can be from a linear or circular genome. The

DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC).

[00128] The DNA can be of a specific species, for example, human, rat, mouse, other animals, plants, bacteria, algae, viruses, and the like. The DNA can also be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The DNA can be from more than one individual or organism. The DNA can be double stranded or single stranded. The DNA can be part of chromatin. The DNA can be associated with histones. The methods described herein can be applied to high molecular weight DNA, such as is isolated from tissues or cell culture, for example, as well as highly degraded DNA, such as cell-free DNA from blood and urine and/or DNA extracted from formalin-fixed, paraffin-embedded tissues, for example.

[00129] An analyte may be a polynucleotide from a particular pathogen, such as a bacteria, virus, fungus, or helminth. For example, an analyte may be a polynucleotide from human immunodeficiency virus (HIV, for example human immunodeficiency virus p17), *Plasmodium*, herpes simplex virus, dengue virus, hepatitis virus, human papilloma virus (for example human papilloma virus E6 or human papilloma virus E7), influenza virus, *Mycobacterium*, Adenoviridae, Picornaviridae, Herpesviridae, Hepadnaviridae, Flaviviridae, Retroviridae, Orthomyxoviridae, Paramyxoviridae, Papovaviridae, Polyomavirus, Rhabdoviridae, Togaviridae, *Streptococcus*, *Pseudomonas*, *Shigella*, *Campylobacter*, *Salmonella*, or *Escherichia*.

[00130] An analyte may also be encoding an enzyme, a protein or a polypeptide. The analyte may be indicative of presence and/or quantity of a carbohydrate, sugar, small molecule, a drug or metabolite (e.g., anti-cancer drug, chemotherapeutic drug, anti-viral drug, antibiotic drug, or any other molecule, such as a co-factor, receptor, receptor ligand, hormone, cytokine, blood factor, antigen, steroid, or antibody).

[00131] The samples may be derived from animals, plants, bacterias, fungi, parasites or viruses. In some cases, the samples are derived from humans. The different samples from which the analytes are derived can comprise multiple samples from the same individual, samples from different individuals, or combinations thereof. In some embodiments, a sample comprises a plurality of polynucleotides from a single individual. In some embodiments, a sample comprises a plurality of polynucleotides from two or more individuals. An individual is any organism or portion thereof from which target polynucleotides can be derived, non-limiting examples of which include plants, animals, fungi, protists, monerans, viruses,

mitochondria, and chloroplasts. Sample polynucleotides can be isolated from a subject, such as a cell sample, tissue sample, or organ sample derived there from, including, for example, cultured cell lines, biopsy, blood sample, or fluid sample containing a cell. Cells may comprise a specific cell type, such as a somatic cell, germline cell, wild-type cell, cancer or tumor cells, or diseased or infected cell. A cell may refer to a cell derived from a particular tissue or a particular locus in a target organism. A cell may comprise whole intact cells, or cell preparations. In some instances, samples may be obtained from bodily fluids which may include blood, urine, serum, lymph, saliva, mucosal secretions, perspiration, or semen. The subject may be an animal, including but not limited to, an animal such as a cow, a pig, a mouse, a rat, a chicken, a cat, a dog, etc., and is usually a mammal, such as a human.

[00132] Samples may also be obtained from manufactured products (such as cosmetics and foods), pharmaceutical products, or consumer products. In some instances, samples may be obtained from environmental samples including air, agricultural products, water, and soil. Samples can also be artificially derived, such as by chemical synthesis. In other instances samples may be the products of experimental manipulation including, recombinant cloning, polynucleotide amplification, polymerase chain reaction (PCR) amplification, purification methods (such as purification of genomic DNA or RNA), and synthesis reactions

[00133] Methods for the extraction and purification of nucleic acids are well known in the art. For example, nucleic acids can be purified by organic extraction with phenol, phenol/chloroform/isoamyl alcohol, or similar formulations, including TRIzol and TriReagent. Other non-limiting examples of extraction techniques include: (1) organic extraction followed by ethanol precipitation, e.g., using a phenol/chloroform organic reagent (Ausubel et al., 1993), with or without the use of an automated nucleic acid extractor, e.g., the Model 341 DNA Extractor available from Applied Biosystems (Foster City, Calif.); (2) stationary phase adsorption methods (U.S. Pat. No. 5,234,809; Walsh et al., 1991); and (3) salt-induced nucleic acid precipitation methods (Miller et al., (1988), such precipitation methods being typically referred to as “salting-out” methods. Another example of nucleic acid isolation and/or purification includes the use of magnetic particles to which nucleic acids can specifically or non-specifically bind, followed by isolation of the beads using a magnet, and washing and eluting the nucleic acids from the beads (see e.g. U.S. Pat. No. 5,705,628). In some embodiments, the above isolation methods may be preceded by an enzyme digestion step to help eliminate unwanted protein from the sample, e.g., digestion with proteinase K, or other like proteases. See, e.g., U.S. Pat. No. 7,001,724. If desired, RNase inhibitors may be

added to the lysis buffer. For certain cell or sample types, it may be desirable to add a protein denaturation/digestion step to the protocol. Purification methods may be directed to isolate DNA, RNA, or both. When both DNA and RNA are isolated together during or subsequent to an extraction procedure, further steps may be employed to purify one or both separately from the other. Sub-fractions of extracted nucleic acids can also be generated, for example, purification by size, sequence, or other physical or chemical characteristic. In addition to an initial nucleic isolation step, purification of nucleic acids can be performed after any step in the methods of the disclosure, such as to remove excess or unwanted reagents, reactants, or products.

VII. Primers

[00134] According to the present disclosure, the term primer relates to a nucleic acid sequence, which is capable of hybridizing e.g. to the target nucleic acid sequence, to a nucleic acid sequence in the vicinity of the target sequence, or to a nucleic acid sequence overlapping with the target nucleic acid sequence. Alternatively, a primer may be capable of hybridizing to either the complementary sequence of the target nucleic acid sequence, to the complementary sequence of a nucleic acid sequence in the vicinity of the target nucleic acid sequence, or to the complementary sequence of a nucleic acid sequence overlapping with the target nucleic acid sequence.

[00135] The primers typically comprise oligonucleotides. The primers may comprise 5-50 nucleotides, for example 5-10 nucleotides, 5-15 nucleotides, 5-20 nucleotides, 5-25 nucleotides, 5-30 nucleotides, 5-35 nucleotides, 5-40 nucleotides, 5-45 nucleotides, 5-50 nucleotides, 10-15 nucleotides, 10-20 nucleotides, 10-25 nucleotides, 10-30 nucleotides, 10-35 nucleotides, 10-40 nucleotides, 10-45 nucleotides, 10-50 nucleotides, 15-20 nucleotides, 15-25 nucleotides, 15-30 nucleotides, 15-35 nucleotides, 15-40 nucleotides, 15-45 nucleotides, 15-50 nucleotides, 20-25 nucleotides, 20-30 nucleotides, 20-35 nucleotides, 20-40 nucleotides, 20-45 nucleotides, 20-50 nucleotides, 25-30 nucleotides, 25-35 nucleotides, 25-40 nucleotides, 25-45 nucleotides, 25-50 nucleotides, 30-35 nucleotides, 30-40 nucleotides, 30-45 nucleotides, 30-50 nucleotides, 35-40 nucleotides, 35-45 nucleotides, 35-50 nucleotides, 40-45 nucleotides, 40-50 nucleotides, or 45-50 nucleotides. In some embodiments the primers comprise 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides. The primers may also comprise more than 50 nucleotides, for example 55, 60, 65, 70 or even more nucleotides.

[00136] A primer can be purified from a restriction digest by conventional methods, or it can be produced synthetically. The primer is preferably single-stranded for maximum efficiency in amplification, but the primer can be double-stranded. Double-stranded primers may be denatured, i.e., treated to separate the strands. One method of denaturing double stranded nucleic acids is by heating.

[00137] The methods of designing and preparing primers that are appropriate for amplification reaction are well known to the person of ordinary skill in the art. Important features when designing oligonucleotides to be used as amplification primers include, but are not limited to, an appropriate size amplification product to facilitate detection (e.g., by electrophoresis), melting temperatures for the members of a pair of primers, and the length of each primer (i.e., the primers need to be long enough to anneal with sequence-specificity and to initiate synthesis but not so long that fidelity is reduced during oligonucleotide synthesis). The length of the amplification primers depends on several factors including the nucleotide sequence identity and the temperature at which these nucleic acids are hybridized or used during in vitro nucleic acid amplification. Primers can be designed using, for example, a computer program such as OLIGO (Molecular Biology Insights, Inc., Cascade, Colo.).

[00138] In some embodiments primers may include a detectable label, although a label may also be introduced by other means (for example via a labeled probe) during the amplification assay. In the former case, the labeled primer will result in a labeled amplification product, as the labeled primers are incorporated into the products. Labeled primers are available from commercial suppliers such as CPG, Inc. (Lincoln Park, N.J.), or they can be prepared via conjugating a label to an unlabeled primer. Labeling oligonucleotides such as primers can be carried out using conventional coupling procedures. Specific coupling procedures, however, will vary depending on the reactive group or groups present on the label and/or on the oligonucleotide. Preferred labels include those moieties detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. Such labels include, without limitation, fluorescence, chemiluminescence, dyes, biotin, haptens, enzymes, enzyme substrates, enzyme cofactors, enzyme inhibitors, enzyme subunits, metal ions, electron-dense reagents, and radioactive isotopes (e.g., ^{32}P). The label moiety can be directly or indirectly attached to the primer.

[00139] For example, if a primer is labeled at the 5' ends with a fluorophore, the amplified target sequence will be labeled with the detectable fluorescent material, and the intensity of fluorescence emitted from the fluorescent material may be measured using a

fluorescence spectrophotometer. Suitable fluorophores that may be used with the primers of the disclosure include those described above. For example the primers may be labeled with 6-FAM; Alexa fluor 405, 430, 488, 532, 546, 555, 568, 594, 633, 647, or 660; Cy2; Cy3; Cy3.5; Cy5; Cy5.5; Cy7; hydroxycoumarin; methoxycoumarin; aminocoumarin; fluorescein; HEX; R-phycoerythrin; rhodamine Red-X; ROX; Red 613; Texas Red; allophycocyanin; TruRed; BODIPY 630/650; BODIPY 650/665; BODIPY-FL; BODIPY-R6G; BODIPY-TMR; BODIPY-TRX; carboxyfluorescein; Cascade Blue; 6-JOE; Lissamine rhodamine B; Oregon Green 488, 500, or 514; Pacific Blue; REG; Rhodamine Green; SpectrumAqua; TAMRA; TET; and Tetramethylrhodamine.

VIII Signals

[00140] The methods presented in this disclosure involve the measurement of a signal. The signal may comprise a intensity and/or a frequency. In many of the exemplary embodiments described in this disclosure, a fluorescent signal may be used. The fluorescent signal is often described as being measured in color (i.e., wavelength) and intensity. Color may be measured by using a spectrophotometer or by using a band pass filter that only allows light of a certain wavelength to reach a photodetector. Such band pass filters are commonly used in qPCR machines. Intensity is measured by using a photo detector. In some cases each analyte is encoded by a signal intensity at a particular wavelength. In some cases, each analyte is encoded by a plurality of signal intensities at a plurality of wavelengths. For example, each analyte may be labeled with a single intensity of 1, 2, 3, 4, 5 or more wavelengths. In some cases an analyte is encoded by at least one additional value selected from the group consisting of a value from said signal and a value from an additional signal. In some cases, the one additional value is selected from the group consisting of a Förster resonance energy transfer (FRET) emission intensity, a FRET emission wavelength, an electrochemical signal, a chemiluminescence wavelength, a chemiluminescence intensity, a fluorescence bleaching rate, and a chemiluminescence bleaching rate. For example, at least one additional value is Förster resonance energy transfer (FRET) emission intensity.

[00141] Although many of the exemplary embodiments described in this disclosure are described in terms of a fluorescent signal, this is by no means limiting. The methods provided in this disclosure may be used with any signal, for example Forster resonance energy transfer (FRET) emission intensity, a FRET emission wavelength, a chemiluminescent wavelength, a chemiluminescent intensity, a electromagnetic signal, a fluorescence bleaching rate, or a chemiluminescence bleaching rate.

[00142] Some of the methods provided in this disclosure utilize a reagent that generates a signal in the presence of an analyte. In some cases, these reagents are probes. The probes may be hybridization probes. The hybridization probes may be an oligonucleotide probe attached to a fluorophore and a quencher (e.g., a TAQMAN probe).

[00143] The methods of the disclosure may use one or more reagents or probes to detect the presence of each analyte. For example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more reagents or probes may be used to detect the presence of an analyte. In some cases, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more reagents or probes may be used to detect the presence of an analyte. In some cases, fewer than 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more reagents or probes may be used to detect the presence of an analyte.

[00144] As described above, oligonucleotide probes attached to a fluorophore and a quencher may be used to detect the presence of an analyte in a polynucleotide amplification assay. So long as the quencher and the fluorophore are in proximity, the quencher quenches the fluorescence emitted by the fluorophore upon excitation by a light source. The sequence of the oligonucleotide probe is designed to be complementary to a polynucleotide sequence present in an analyte, and therefore capable of hybridizing to the polynucleotide sequence present in the analyte. Hybridization of the oligonucleotide probe may be performed in a nucleic acid amplification reaction comprising primers (e.g., a polymerase chain reaction). Upon extension of the primers by a DNA polymerase, the 5' to 3' exonuclease activity of the polymerase degrades the probe, releasing the fluorophore and the quencher into the medium. The proximity between the fluorophore and the quencher is broken and the signal from the fluorophore is no longer quenched. Thus, the amount of fluorescence detected is a function of the amount of analyte present. If no analyte is present, the probe will not hybridize to an analyte, and the fluorophore and quencher will remain in close proximity. Little or no signal will be produced.

[00145] Oligonucleotide probes may have one or a plurality of fluorophores and quenchers per probe. For example, in some embodiments an oligonucleotide probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more fluorophores. An oligonucleotide probe may comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 fluorophores. An oligonucleotide probe may comprise fewer than 2, 3, 4, 5, 6, 7, 8, 9, or 10 fluorophores. An oligonucleotide probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more quenchers. An oligonucleotide probe may comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 quenchers. An oligonucleotide probe may comprise fewer than 2, 3, 4, 5, 6, 7, 8, 9, or 10 quenchers. Attachment of probes and

quenchers to a probe may be performed in the same reaction or in serial reactions. A series of reactions may be performed to label probes with at least one fluorophore and the reaction products may be mixed to generate a mixture of probes with different fluorophores.

[00146] Suitable fluorophores for use in the present disclosure include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue, Texas Red, and others described in the 11th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference in its entirety.

[00147] Commercially available fluorescent nucleotide analogues readily incorporated into the labeling oligonucleotides include, for example, Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy5- dUTP (GE Healthcare), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY®R-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-1 4-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-1 4-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, and Alexa Fluor® 546-1 4-UTP (Invitrogen).

[00148] Other fluorophores available for post-synthetic attachment include, inter alia, Alexa Fluor® 350, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647, BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Invitrogen), and Cy2, Cy3.5, Cy5.5, and Cy7 (GE Healthcare).

[00149] If an oligonucleotide probe comprises two or more fluorophores, these fluorophores may be arranged to allow Förster (Fluorescence) resonance energy transfer (FRET) to occur between the fluorophores. Briefly, FRET is a mechanism of energy transfer between fluorophores. Using FRET-based probes allows one excitation source to generate two different fluorescence emission signals by excitation of a single fluorophore. For

example, the methods described herein may use paired probes, where one probe in the pair is attached to a fluorophore and a quencher and a second probe is attached to two fluorophores (in close enough proximity for FRET to occur) and a quencher. Any combination of fluorophores and quenchers that provide for FRET may be used. Such an approach doubles the number of fluorescent probes that may be used with an excitation source, because a single excitation source can be used to produce two signals: one from the probe with one fluorophore and one from the probe with two fluorophores in close enough proximity to provide for FRET. For example, using the encoding method described in Table 8, the number of unambiguously detectable analytes could be increased from 16 to 32 by pairing each of the probes in Table 8 with a corresponding FRET probe.

[00150] In one example, a first probe may be labeled with the fluorophore 5'-fluorescein amidite (5'-FAM) and the quencher black hole quencher 1 (BHQ-1). A second probe may be labeled with 5'-FAM in close proximity (e.g., attached to) to cyanine 5.5 (Cy5.5) and the quencher black hole quencher 3 (BHQ-3). Upon digestion of the probe, via the nuclease activity of the polymerase, two fluorescent signals are generated from a single excitation wavelength (e.g., 470 nm). The fluorophore from the first probe will fluoresce at about 520 nm. The fluorophores on the second probe undergo FRET. The donor fluorophore, FAM is excited by the excitation light source (e.g., 470 nm) and transfers its energy to the acceptor fluorophore, Cy5.5. The acceptor fluorophore emits at about 705 nm. These fluorophores and quenchers are merely exemplary. Any fluorophores that can undergo FRET and any quenchers that can quench fluorescence are suitable for use with the disclosure. Methods for producing FRET-based probes are described in Jothikumar et al., *BioTechniques*, 2009, 46(7):519-524.

[00151] Although many aspects of the disclosure are exemplified using nucleic acid-based probes, one of ordinary skill in the art will readily recognize that other forms of probes would work equally well with the disclosure described in this disclosure. For example, a binding molecule specific to an analyte could be used as a probe. Non-limiting exemplary binding molecules include an antibody recognizing an analyte, and generating a signal in the presence of an analyte.

[00152] Similarly, although many aspects of the disclosure are exemplified using fluorescent labels attached to probes, many other types of labels could also be used with the methods of the disclosure. The disclosure can be used with any label capable of generating a measurable signal.

[00153] For multicolor imaging, signals of different wavelength can be obtained by multiple acquisitions or by simultaneous acquisition by splitting the signal, using RGB detectors or analyzing the whole spectrum (Richard Levenson, Cambridge Healthtech Institutes, Fifth Annual meeting on Advances in Assays, Molecular Labels, Signaling and Detection, May 17-18th Washington D.C.). Several spectral lines can be acquired by the use of a filter wheel or a monochromator. Electronic tunable filters such as acoustic-optic tunable filters or liquid crystal tunable filters can be used to obtain multispectral imaging (e.g. Oleg Hait, Sergey Smirnov and Chieu D. Tran, 2001, Analytical Chemistry 73: 732-739). An alternative method to obtain a spectrum is hyperspectral imaging (Schultz et al., 2001, Cytometry 43:239-247).

Methods and apparatus for signal detection and processing of intensity data are disclosed in, for example, U.S. Pat. Nos. 5,143,854, 5,547,839, 5,578,832, 5,631,734, 5,800,992, 5,834,758; 5,856,092, 5,902,723, 5,936,324, 5,981,956, 6,025,601, 6,090,555, 6,141,096, 6,185,030, 6,201,639; 6,218,803; and 6,225,625, 7,689,022 and in WO99/47964, each of which also is hereby incorporated by reference in its entirety for all purposes. The fluorophores that may be used with the disclosure are not limited to any of the fluorophores described herein. For example, fluorophores with improved properties are continually developed, and these fluorophores could readily be used with the methods provided in this disclosure. Such improved fluorophores include quantum dots, which may emit energy at different wavelengths after being excited at a single wavelength. The advantage of using such fluorophores is that only a single excitation source is needed, but many different signals may be quantified, for example in terms of color and intensity.

IX. Other Additives

[00154] The PCR reagents may furthermore comprise additives such as a DNA polymerase (e.g. Taq polymerase), deoxynucleotide triphosphates (e.g. dATP, dCTP, dTTP, and dGTP), salts (e.g. KCl and MgCl₂), buffer (e.g. Tris-HCl), biopreservatives (e.g. sodium azide), PCR enhancers (e.g. Betaine, Trehalose, etc.), and inhibitors (e.g. RNase inhibitors), detergents (e.g. Triton X-100, Triton X-114, NP-40, Tween 20, Tween 80 and similar non ionic detergents). Other additives may include 2-pyrrolidone, acetamide, N-methylpyrrolidone (NMP), B-hydroxyethylpyrrolidone (HEP), propionamide, NN-dimethylacetamide (DMA), N-methylformamide (MMP), NN-dimethylformamide (DMF), formamide, N-methylacetamide (MMA), dimethyl sulfoxide (DMSO), polyethylene glycol, betaine, tetramethylammonium chloride (TMAC), 7-deaza-2'-deoxyguanosine, T4 gene 32 protein, 2-

mercaptoethanol or glycerol. Paragraph 104 and 105 remove MIP 00187 for polymerase; 00188 for different types of PCR

X. Analysis of the characteristic curves

[00155] Once characteristic curves are generated for a set of reaction, these curves can be analyzed by curve-fitting algorithms. The curves may be fitted using, for example, linear regression, non-linear regression (for example Levenberg-Marquardt (LM) method), or may simply connect fluorescence data from adjacent cycles with a smoothed or non-smoothed line. The fitted curves may be then used to determine the slope of the amplification reaction in its linear regime. In the embodiments where two or more of these curves are in superposition (i.e. where two or more targets are present in multiplexed fashion) each target may be uniquely identified as long as the curves are in the linear dynamic range of the fluorometer. The multiplexed discrimination of multiple targets in a single reaction by means of reaction rates allows for another axis of orthogonality for PCR multiplexing.

[00156] In some cases the curves generated by the methods of the disclosure may be analyzed by taking the derivative of the signal with respect to cycle number and generating a derivative plot. In the derivative plot, the plateaus will show up as zeros or near-zeros. The normal saturating amplification will show up as a sinusoidal-like peak between two zeros. The linear amplification will show up as plateaus, where the magnitude will be informative of the slope and will be proportional to the concentration of respective primer.

[00157] Another strategy for discriminating curves is to transform them into a transform domain thereby generating a transformed signal. For example, the signal may be transformed into a frequency domain. The transforming may be performed by any suitable method including Fourier transform, fast Fourier transform (FFT), discrete-time Fourier transform (DTFT), discrete-time Fourier transform (DTFT), short-time Fourier transform, fractional Fourier transform, chirplet transforms, or wavelet transform. In general, any eigenvalue transformation that is a conformal map can be used to extract the natural frequencies of the PCR curve. Once this is obtained, the frequency space signature (Figure 7) of each reaction can be used to resolve which targets were present. Careful design of the primers and probes will allow reaction rate information to be resolved as well, allowing for a built-in gauge of the quality or efficiency of the PCR reaction. Furthermore, this frequency space signature can be cross-correlated against a database of known target combinations to determine the most likely mix of targets in a particular reaction.

[00158] Accordingly, in another aspect the disclosure provides a method of detecting a nucleic acid analyte in a sample. The method comprises performing an amplification reaction on a sample, measuring a signal generated during said amplification reaction, transforming the signal into a transform domain thereby generating a transformed signal, comparing the transformed signal to a reference signal, and based on the comparison determining whether the analyte is present or absent.

[00159] The reference signal may be a signal in a database of reference signals. The database may comprise at least 1, 5, 10, 50, 100, 200, 300, 400, 500 or more signals. In some cases, the database comprises at most 500, 400, 300, 200, 100, 50, 10, 5 or 1 signal.

XI. Kits

[00160] In one aspect, the disclosure provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, a kit comprises a composition of the disclosure, in one or more containers. The kits may further comprise additional agents, such as those described above, for use according to the methods of the disclosure. The kit elements may be provided in any suitable container, including but not limited to test tubes, vials, flasks, bottles, ampules, syringes, or the like.

[00161] An exemplary kit of the disclosure may comprise at least three nested primers, at least one pair of primers comprising a limiting primer and an excess primer, and a real time detection agent. The kits may further comprise one or more PCR reagents, for example salts, deoxynucleotide triphosphates, buffers, detergents, PCR inhibitors, PCR enhancers and nucleic acid polymerases. The kit may also comprise additives such as 2-mercaptoethanol, 2-pyrrolidone, acetamide, N-methylpyrrolidone (NMP), B-hydroxyethylpyrrolidone (HEP), propionamide, NN-dimethylacetamide (DMA), N-methylformamide (MMP), NN-dimethylformamide (DMF), formamide, N-methylacetamide (MMA), dimethyl sulfoxide (DMSO), polyethylene glycol, betaine, tetramethylammonium chloride (TMAC), 7-deaza-2'-deoxyguanosine, T4 gene 32 protein, or glycerol. A kit may additionally include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions can include variations that can be implemented.

[00162] The containers of the kits can generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquoted. Where there is more than one component in the kit, the kit also can generally contain a second, third or other additional container into which the additional

components can be separately placed. However, various combinations of components can be comprised in a container.

[00163] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit can be provided as dried powder(s). Agents may be provided in aliquots for single-use or as stocks from which multiple uses, such as in a number of reaction, may be obtained. When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent. Elements of the kit can further be provided, without limitation, in any suitable amounts and/or using any of the combinations (such as in the same kit or same container) described above or any other suitable combination known in the art.

[00164] The agents may be provided in a form that may be directly used in the methods of the disclosure, or in a form that requires preparation prior to use, such as in the reconstitution of lyophilized agents. For example, primers in kits provided herein may be in a lyophilized or non-lyophilized state. In some embodiments at least one primer provided in the kits is lyophilized. In some cases more than one, for example two, three, four, five, six or even more of the primers provided in a kit are in lyophilized state. In some examples, all of the primers contained in the kits of the present disclosure are lyophilized.

[00165] In certain embodiments, the concentration of the excess primer in the kits is at least 2 times, 3 times, 4 times, 5 times, 6 times, 7 times, 8 times, 9 times, 10 times, 11 times, 12 times, 13 times, 14 times, 15 times, 16 times, 17 times, 18 times, 19 times, 20 times, 21 times, 22 times, 23 times, 24 times, 25 times, 30 times, 35 times, 40 times, 45 times, 50 times, 55 times, 60 times, 65 times, 70 times, 75 times, 80 times, 85 times, 90 times, 95 times, 100 times, 150 times, 200 times, 250 times, 300 times, 350 times, 400 times, 450 times, 500 times, 550 times, 600 times, 650 times, 700 times, 750 times, 800 times, 850 times, 900 times, 950 times, or 1000 times that of the limiting primer.

[00166] In certain embodiments, the concentration of the excess primer in the kits is at most 1000 times, 900 times, 950 times, 850 times, 800 times, 750 times, 700 times, 650 times, 600 times, 550 times, 500 times, 450 times, 400 times, 350 times, 300 times, 250 times, 200 times, 150 times, 100 times, 95 times, 90 times, 85 times, 80 times, 75 times, 70 times, 65 times, 60 times, 55 times, 50 times, 45 times, 40 times, 35 times, 30 times, 25 times, 24 times, 23 times, 22 times, 21 times, 20 times, 19 times, 18 times, 17 times, 16 times, 15 times, 14 times, 13 times, 12 times, 11 times, 10 times, 9 times, 8 times, 7 times, 6 times, 5 times, 4 times, 3 times, or 2 times that of the limiting primer.

[00167] The length and sequence of the primers in the kits may be adjusted or modified such that the concentration-adjusted melting temperature of the limiting Primer is greater than or equal to the concentration-adjusted melting point of the excess primer. For example, the difference in the concentration-adjusted melting temperature of the limiting and the excess primer may be at least 0 °C, 1 °C, 1.5 °C, 2 °C, 2.5 °C, 3 °C, 3.5 °C, 4 °C, 4.5 °C, 5 °C, 5.5 °C, 6 °C, 6.5 °C, 7 °C, 7.5 °C, 8 °C, 8.5 °C, 9 °C, 9.5 °C, 10 °C or more. In some embodiments, the difference in the concentration-adjusted melting temperature of the limiting and the excess primer may be at least 10 °C, 9.5 °C, 9 °C, 8.5 °C, 8 °C, 7.5 °C, 7 °C, 6.5 °C, 6 °C, 5.5 °C, 5 °C, 4.5 °C, 4 °C, 3.5 °C, 3 °C, 2.5 °C, 2 °C, 1.5 °C, 1 °C, 0 °C or less.

[00168] The initial concentration of the limiting primers in the kits may be less than 500 nM. For example the initial concentration of the limiting primer may be less than 10 nM, 50 nM, 100 nM, 150 nM, 200 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM or 500 nM. In some embodiments the initial concentration of the limiting primers in the kits may be more than 500 nM. For example the initial concentration of the limiting primer may be more than 500 nM, 450 nM, 400 nM, 350 nM, 300 nM, 250 nM, 200 nM, 150 nM, 100 nM, 50 nM or 10 nM.

[00169] The at least one real time detection agent in the kits provided herein may be any reagent that generates a signal in the presence of an analyte. In some cases, these reagents are probes. The probes may be hybridization probes. The hybridization probes may be an oligonucleotide probe attached to a fluorophore and a quencher (e.g., a TAQMAN probe).

[00170] The kits may be for detection of pathogens, such as bacteria, virus, fungus, or helminth. For example, the kits provided herein may be used for detection of human immunodeficiency virus, *Plasmodium*, herpes simplex virus, dengue virus, hepatitis virus, human papilloma virus, influenza virus, *Mycobacterium*, Adenoviridae, Picornaviridae, Herpesviridae, Hepadnaviridae, Flaviviridae, Retroviridae, Orthomyxoviridae, Paramyxoviridae, Papovaviridae, Polyomavirus, Rhabdoviridae, Togaviridae, *Streptococcus*, *Pseudomonas*, *Shigella*, *Campylobacter*, *Salmonella*, *Neisseria gonorrhoeae*, or *Escherichia*.

XII. Analytical Techniques and Instrumentation

[00171] The methods provided herein are suitable for use with a variety of detection methods. For example, the methods may be applied using an analytical technique that measures the wavelength and intensity of a fluorescent signal. This may be accomplished by measuring the intensity of a signal across a spectrum of wavelengths, or by using band pass

filters that restrict the passage of certain wavelengths of light, thereby allowing only light of certain wavelengths to reach a photodetector. Many real-time PCR and quantitative PCR instruments comprise an excitation light source and band pass filters that enable the detection of fluorescent signals in four colors (e.g., blue, green, yellow, and red). Therefore, the methods of the disclosure can be readily applied using instruments widely used in the art.

[00172] The methods are also compatible with end point PCR and digital PCR methods. Digital PCR methods (e.g., DROPLET DIGITAL (BIORAD) and DYNAMIC ARRAY (FLUIDIGM)) produce highly sensitive quantification of polynucleotide copy numbers. The methods provided herein can be easily integrated into these systems to drastically expand their throughput by allowing multiplexing in a droplet or a dynamic array.

[00173] In some cases, instruments may be modified or constructed, for example, to provide additional excitation light sources, at multiple wavelength, and/or to provide additional band pass filters or a capability of determining a complete spectrum. Including additional excitation sources would allow for the excitation of a larger variety of fluorophores. Including additional band pass filters, or modifying or constructing an instrument capable of determining an entire spectrum allows detection of a wider variety of emissions from fluorophores. These techniques can be used to increase the number of fluorophores that can be used with the methods described in this disclosure and, accordingly, to increase the number of analytes that be simultaneously detected.

XIII. Systems and Software

[00174] The methods of this disclosure may also be performed as part of a system. For example, one or more steps of the methods described herein may be performed by software contained on a computer readable medium. The software may be used to convert measurements made on an instrument (e.g., a PCR machine) into the presence or absence of an analyte. The software may be used to calculate the slopes of the linear region of the generated characteristic curves. The software may also be used to transform the measured signal into a transform domain and to thereby generate a transformed signal. The software may compare the transformed signal to a reference signal and based on the comparison determine the amount of analyte present in the sample. In some cases the determined amount of the analyte in the sample may be used to conclude whether a said analyte is present or absent in the sample.

[00175] In one example, the software is embedded in the instrument making the measurement. In another example, the software is installed on a computer attached to the

instrument making the measurement. In another example, the software is in the “cloud” -- i.e., on a computer that another computer may communicate with through a network. Each of these elements may be connected to a controller that communicates with each of the elements and coordinates their actions. For example, a controller may initiate the measurement of the signal and generation of characteristic curves. The controller can execute software to transform the signal into a transform domain, thereby generating a transformed signal. The controller may also initiate the comparison of the measured signal or the transformed signal to a reference signal.

[00176] Alternatively, the controller may operate an oligonucleotide synthesis instrument that can synthesize the appropriate probes. After analysis of the sample on the instrument (which may also be automated), the controller can process the results of the analysis using the reference signal values. The controller may then provide these results to a user.

[00177] Figure 8 is a block diagram illustrating an example architecture of a computer system **100** that can be used in connection with example embodiments of the present disclosure. As depicted in Figure 8, the example computer system can include a processor **102** for processing instructions. Non-limiting examples of processors include: Intel Xeon™ processor, AMD Opteron™ processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0™ processor, ARM Cortex-A8 Samsung S5PC100™ processor, ARM Cortex-A8 Apple A4™ processor, Marvell PXA 930™ processor, or a functionally-equivalent processor. Multiple threads of execution can be used for parallel processing. In some embodiments, multiple processors or processors with multiple cores can also be used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[00178] As illustrated in Figure 8, a high speed cache **104** can be connected to, or incorporated in, the processor **102** to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor **102**. The processor **102** is connected to a north bridge **106** by a processor bus **108**. The north bridge **106** is connected to random access memory (RAM) **110** by a memory bus **112** and manages access to the RAM **110** by the processor **102**. The north bridge **106** is also connected to a south bridge **114** by a chipset bus **116**. The south bridge **114** is, in turn, connected to a peripheral bus **118**. The peripheral bus can be, for example, PCI, PCI-X, PCI Express, or other peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data

transfer between the processor, RAM, and peripheral components on the peripheral bus **118**. In some alternative architectures, the functionality of the north bridge can be incorporated into the processor instead of using a separate north bridge chip.

[00179] In some embodiments, system **100** can include an accelerator card **122** attached to the peripheral bus **118**. The accelerator can include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator can be used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

[00180] Software and data are stored in external storage **124** and can be loaded into RAM **110** and/or cache **104** for use by the processor. The system **100** includes an operating system for managing system resources; non-limiting examples of operating systems include: Linux, WindowsTM, MACOSTM, BlackBerry OSTM, iOSTM, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example embodiments of the present disclosure.

[00181] In this example, system **100** also includes network interface cards (NICs) **120** and **121** connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that can be used for distributed parallel processing.

[00182] Figure 9 is a diagram showing a network **200** with a plurality of computer systems **202a**, and **202b**, a plurality of cell phones and personal data assistants **202c**, and Network Attached Storage (NAS) **204a**, and **204b**. In example embodiments, systems **202a**, **202b**, and **202c** can manage data storage and optimize data access for data stored in Network Attached Storage (NAS) **204a** and **204b**. A mathematical model can be used for the data and be evaluated using distributed parallel processing across computer systems **202a**, and **202b**, and cell phone and personal data assistant systems **202c**. Computer systems **202a**, and **202b**, and cell phone and personal data assistant systems **202c** can also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) **204a** and **204b**. Figure 9 illustrates an example only, and a wide variety of other computer architectures and systems can be used in conjunction with the various embodiments of the present disclosure. For example, a blade server can be used to provide parallel processing. Processor blades can be connected through a back plane to provide parallel processing.

Storage can also be connected to the back plane or as Network Attached Storage (NAS) through a separate network interface.

[00183] In some example embodiments, processors can maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In other embodiments, some or all of the processors can use a shared virtual address memory space.

[00184] Figure 10 is a block diagram of a multiprocessor computer system **300** using a shared virtual address memory space in accordance with an example embodiment. The system includes a plurality of processors **302a-f** that can access a shared memory subsystem **304**. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) **306a-f** in the memory subsystem **304**. Each MAP **306a-f** can comprise a memory **308a-f** and one or more field programmable gate arrays (FPGAs) **310a-f**. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs **310a-f** for processing in close coordination with a respective processor. For example, the MAPs can be used to evaluate algebraic expressions regarding the data model and to perform adaptive data restructuring in example embodiments. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP can use Direct Memory Access (DMA) to access an associated memory **308a-f**, allowing it to execute tasks independently of, and asynchronously from, the respective microprocessor **302a-f**. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[00185] The above computer architectures and systems are examples only, and a wide variety of other computer, cell phone, and personal data assistant architectures and systems can be used in connection with example embodiments, including systems using any combination of general processors, co-processors, FPGAs and other programmable logic devices, system on chips (SOCs), application specific integrated circuits (ASICs), and other processing and logic elements. In some embodiments, all or part of the computer system can be implemented in software or hardware. Any variety of data storage media can be used in connection with example embodiments, including random access memory, hard drives, flash memory, tape drives, disk arrays, Network Attached Storage (NAS) and other local or distributed data storage devices and systems.

[00186] In example embodiments, the computer system can be implemented using software modules executing on any of the above or other computer architectures and systems.

In other embodiments, the functions of the system can be implemented partially or completely in firmware, programmable logic devices such as field programmable gate arrays (FPGAs) as referenced in Figure 10, system on chips (SOCs), application specific integrated circuits (ASICs), or other processing and logic elements. For example, the Set Processor and Optimizer can be implemented with hardware acceleration through the use of a hardware accelerator card, such as accelerator card **122** illustrated in Figure 8.

XIV. Services

[00187] The methods provided herein may also be performed as a service. For example, a service provider may obtain the identity of a series of analytes that a customer wishes to analyze. The service provider may then encode each analyte to be detected by any of the methods described herein and provide appropriate reagents to the customer for the assay. The customer may perform the assay and provide the results to the service provider for decoding. The service provider may then provide the decoded results to the customer. The customer may also encode analytes, generate probes, and/or decode results by interacting with software installed locally (at the customer's location) or remotely (e.g., on a server reachable through a network). Exemplary customers include clinical laboratories, physicians, manufacturers, and the like.

EXAMPLES

Example 1: General Materials & Methods

1. DNA Sequences, Primers, and Probes

[00188] Synthesis of the oligonucleotides for the following examples is done by INTEGRATED DNA TECHNOLOGIES (Coralville, Iowa). HSV-1 and HSV-2 sequences from the Los Alamos National Laboratory reference sequence are selected for the exemplary study.

[00189] The 198-mer HSV-2 oligonucleotide is reconstituted with TE buffer at a concentration of 1.7 μM . Forward and reverse primers A for the HSV-2 oligonucleotide are reconstituted with TE buffer at a concentration of 100 μM and 1 μM respectively. Forward and reverse primers B for the HSV-2 oligonucleotide are reconstituted with TE buffer at a concentration of 40 μM and 1 μM respectively. Forward and reverse primers C for the HSV-2 oligonucleotide are reconstituted with TE buffer at a concentration of 3 μM and 1 μM respectively. A TAQMAN sense probe ("HSV-2_probe1") is synthesized with a 6-FAM fluorophore at the 5' end and a BHQ-1 quencher at the 3' end.

[00190] The 199-mer HSV-1 oligonucleotide is reconstituted with TE buffer at a concentration of 1.5 μM . Forward and reverse primers D for the HSV-1 oligonucleotide are reconstituted with TE buffer at a concentration of 1 μM and 100 μM respectively. Forward and reverse primers E for the HSV-1 oligonucleotide are reconstituted with TE buffer at a concentration of 1 μM and 40 μM respectively. Forward and reverse primers F for the HSV-1 oligonucleotide are reconstituted with TE buffer at a concentration of 1 μM and 3 μM respectively. A TAQMAN sense probe (“HSV-1_probe1”) is synthesized with a 6-FAM fluorophore at the 5’ end and a BHQ-1 quencher at the 3’ end. An additional TAQMAN sense probe (“HSV-1_probe2”) is synthesized with Cy3 at the 5’ end and BHQ-2 at the 3’ end.

2. Polymerase Chain Reactions

[00191] All PCR reactions are performed in a ROCHE 480 LIGHTCYCLER instrument. Forty-five thermal cycles are performed, with a 60 second hot-start at 95°C. The cycling conditions for denaturation, annealing, and extension are 45 seconds, 50 seconds, and 60 seconds at 95°C, 65°C, and 70°C, respectively. Each experiment is run in quintuplicate, with a reaction volume of 15 μL . Fluorescence measurements are obtained in the following channels after every annealing step: 483 nm to 533 nm (FAM), 523 nm to 568 nm (Cy3), 558 nm to 610 nm (ROX), and 615 nm to 670 nm (Cy5).

Example 2: Detection of Polynucleotide Markers from Pathogens

[00192] This example describes the detection of polynucleotides from HSV-1 and HSV-2.

1. PCR Reaction Components

[00193] Three PCR reactions are assembled, using the reagents described in Table 1.

Table 1. Reagents for the three PCR reactions.

Reaction No.	Type of Reagent	Reagent	Concentration	Volume
1		UltraPure Water	-	34 μL
		Taq 5x Master Mix	-	42 μL
	<i>Template</i>	HSV-2 Template	17 nM	2 μL
	<i>Primers</i>	HSV-2 FWD Primer A	100 nM	2 μL
		HSV-2 RVS Primer A	1 nM	2 μL

		HSV-2FWD Primer B	40 nM	2 μ L
		HSV-2 RVS Primer B	1 nM	2 μ L
		HSV-2 FWD Primer C	3 nM	2 μ L
		HSV-2 RVS Primer C	1 nM	2 μ L
	<i>Probe</i>	HSV-2 Probe 1	1 nM	2 μ L
2		UltraPure Water	-	34 μ L
		Taq 5x Master Mix	-	42 μ L
	<i>Template</i>	HSV-1 Template	17 nM	2 μ L
	<i>Primers</i>	HSV-1 FWD Primer D	1 nM	2 μ L
		HSV-1 RVS Primer D	100 nM	2 μ L
		HSV-1 FWD Primer E	1 nM	2 μ L
		HSV-1 RVS Primer E	40 nM	2 μ L
		HSV-1 FWD Primer F	1 nM	2 μ L
		HSV-1 RVS Primer F	3 nM	2 μ L
	<i>Probe</i>	HSV-1 Probe 1	1 nM	2 μ L
3		UltraPure Water	-	34 μ L
		Taq 5x Master Mix	-	42 μ L
	<i>Template</i>	HSV-2 Template	17 nM	2 μ L
		HSV-1 Template	17 nM	2 μ L
	<i>Primers</i>	HSV-2 FWD Primer A	100 nM	2 μ L
		HSV-2 RVS Primer A	1 nM	2 μ L
		HSV-2 FWD Primer B	40 nM	2 μ L
		HSV-2 RVS Primer B	1 nM	2 μ L
		HSV-2 FWD Primer C	3 nM	2 μ L
		HSV-2 RVS Primer C	1 nM	2 μ L
		HSV-1 FWD Primer D	1 nM	2 μ L
		HSV-1 RVS Primer D	100 nM	2 μ L
		HSV-1 FWD Primer E	1 nM	2 μ L
		HSV-1 RVS Primer E	40 nM	2 μ L
		HSV-1 FWD Primer F	1 nM	2 μ L
		HSV-1 RVS Primer F	3 nM	2 μ L
	<i>Probe</i>	HSV-2 Probe 1	1 nM	2 μ L

		HSV-1 Probe 1	1 nM	2 μ L
--	--	---------------	------	-----------

[00194] The signal generated from PCR reactions 1-3 is measured in arbitrary fluorescence units (AFU) and plotted as a function of number of cycles. A typical curve generated by a uniplex PCR reaction (for example reaction 1 and 2) is shown in Figure 5.

[00195] A typical curve generated from a multiplex PCR reaction of the disclosure (for example, from reaction 3 where two targets are present in a single multiplexed reaction) is shown in Figure 6. The curves generated are analyzed by curve fitting algorithms. These algorithms calculate the slope of the reaction in its linear regime. Each target present in these multiplexed reactions can be uniquely identified as long as the curves are in the linear dynamic range of the fluorometer.

Example 3: Generation of characteristic curves

[00196] Figure 5 shows three consecutive plateaus separated by normal amplification regions. Such plot may be generated by the different reactions depending on the starting reagents.

- (i) Three independent sequences, A, B, and C are present at the same time and addressed by standard PCR with a single intercalator dye and symmetric primers optimized for maximal exponential performance but designed in such a way that they hybridize at different temperatures. The reaction is programmed in such a way that it runs for the first third of the cycles at hybridization temperature for the primers of sequence A, which is too high for the primers of B and C to hybridize. The second third of the cycles is then run at the temperature suitable for the primers of B to hybridize but still too hot for the primers of C. Finally, the final third of the cycles are run at temperature low enough for C to hybridize. As there is normal amplification in the first third of the cycles, the conclusion is that A is present. As there is normal amplification in the second third of the cycles, the conclusion is that B is present. As there is normal amplification in the final third of the cycles, the conclusion is that C is present. Thus this curve decodes to positive results for A, B, and C under this setup. If any of the plateaus were missing, we would conclude that the respective sequences were missing.

- (ii) A multiplex PCR reactions contains a nested set of two pairs of symmetric primers A and B for sequence #1 and a normal set of primer C for sequence #2. The outer primers of sequence #1 A have the highest melting temperature, followed by the inner primers sequence #1 B, followed by the primer C of sequence #2. The PCR program runs the first third at hybridization temperature that allows only the outer set of primers to sequence #1 to hybridize, leading to normal symmetric exponential amplification for the outer set of sequence #1. The reaction saturates by the end of the first third of the cycles. Then the program lowers the hybridization temperature so that the inner pair hybridizes but it is still too hot for the primers of sequence #2. Thus the second third sees normal symmetric exponential amplification of the inner primers. Finally, in the end third of the cycles, the temperature is low enough for sequence #2 to amplify symmetrically and exponentially and saturate by the end of the reaction. Looking at the curve and the reagent setup, it is concluded that both sequences are present.
- (iii) Three pairs of nested primers A, B, and C are used for the same sequence. A, B, and C in the graph indicate the correct workings of the assay and the successful detection of the sequence. The use of three pairs of nested primers as opposed to just one primer pair may provide increased reliability. For example, if a single pair of primer is used, and the primers misprime on another sequence, e.g. in genomic DNA, amplification would still happen and the intercalator dye would report it, resulting in a false positive. This may not be possible when nested primer pairs are used because not all plateaus will be present. For example, if the sequence is not present but the outermost pair misprimes, A will be seen as normal, but B and C will not happen, so the results would be negative.

[00197] Figure 6 can be interpreted in very similar terms as described above for Figure 5, except that the amplification is linear asymmetric, so the respective coding is by asymmetrically chosen primer pairs. The difference between A and D is the concentration of the limiting primer, which can be used as a coding method. For example, if the curve looks like DEF instead of ABC, then a different set of target sequences has been detected using the same multiplexed assay.

Example 4: Discrimination of curves by basis transformation

The qPCR signals generated from a 3-target multiplexed qPCR assay similar to the one in Example 2 are transformed using a basis transformation to obtain frequency space signatures for the reactions. These frequency space signatures may be used to resolve which targets are present. In this example, Target 1 has a unique signature that is reproducible across qPCR reaction replicates. The target is identified by a specific (height, width) tuple that is specific only to Target 1. For this multiplexed reaction chemistry, a single peak signature of (0.096, 14) identifies Target 1. Furthermore, Target 2 has a distinct signature from Target 1. When Target 2 is present with this qPCR reaction chemistry, the kinetic signature observed is a dual peak curve with (height, width) tuples of (0.01, 7) and (0.1, 10). Furthermore, when Targets 1 and 2 are present in a qPCR reaction with this multiplex assay, the generated kinetic signature is unique and highly reproducible. In this example, the signature is dual peaked with (height, width) tuples of (0.018, 7) and (0.09, 12). With this example assay set, we demonstrate the proof-of-principle of unambiguous discrimination of targets using basis-transformed qPCR signals. Another typical frequency space signature obtained from the multiplexed PCR reactions of the disclosure is shown in Figure 7.

Example 5: Detection of one or more nucleic acid analytes

[00198] This example illustrates a method for the detection of one or more of a nucleic acid analyte in a sample.

[00199] In one aspect, the presence or absence of a target nucleic acid analyte in a sample is detected. An amplification reaction is performed on the sample and a signal generated during the amplification reaction is measured. A kinetic signature is generated from the measured signal, which is compared to a reference signature to determine whether the generated kinetic signature corresponds to the reference signature. Comparison of the generated kinetic signature to the reference signature allows for detection of the presence or absence of the target nucleic acid analyte in the sample. For example, a nucleic acid analyte is determined to be present in a sample if the generated kinetic signature corresponds to the reference signature. Conversely, a nucleic acid analyte is determined to be absent if the generated kinetic signature does not correspond to the reference signature.

[00200] In another aspect, the presence or absence of each of a plurality of target nucleic acid analytes in a sample is detected. A primer set is added to the sample that comprises at least two primer pairs. At least one of the primer pairs comprises a limiting primer, and each primer pair defines a region of one of the target nucleic acid analytes to be

amplified. An amplification reaction is performed on the sample and a signal generated during the amplification reaction is measured. The measured signal is used to determine the presence or absence of each of the plurality of target nucleic acid analytes. In some aspects, the concentration of a limiting primer may be sufficiently lower relative to other primers in the amplification reaction so as to affect the generated signal. For example, where the generated signal is plotted as an amplification curve, the concentration of the limiting primer may be sufficiently lower such that the generated amplification curve is not sigmoidal.

[00201] In yet another aspect, the presence or absence of two or more of a plurality of target nucleic acids in a sample is detected. A first and second primer pair is added to the sample, each primer pair defining a region of a nucleic acid analyte and having a first annealing temperature and a second annealing temperature, respectively. A thermocycling amplification reaction is performed on the sample. The thermocycling amplification reaction comprises a set of temperature cycles. At least one temperature cycle is performed that has a minimum temperature, *i.e.* an annealing temperature, that is not greater than the annealing temperature of the first primer pair that corresponds to the first annealing temperature, but is greater than the annealing temperature of the second primer pair, *i.e.* the second annealing temperature. At least one other temperature cycle is performed that has a minimum temperature, *i.e.* an annealing temperature, that is not greater than the annealing temperature of the second primer pair, *i.e.* the second annealing temperature. A signal generated during the amplification reaction is measured, and the measured signal is used to determine the presence or absence of each of the two or more of a plurality of target nucleic acid analytes.

[00202] Unless the context requires otherwise, throughout the present specification and claims, the word "comprise" and variations thereof, such as, "comprises" and "comprising," which is used interchangeably with "including," "containing," or "characterized by," is inclusive or open-ended language and does not exclude additional, unrecited elements or method steps. The phrase "consisting of" excludes any element, step, or ingredient not specified in the claim. The phrase "consisting essentially of" limits the scope of a claim to the specified materials or steps and those that do not materially affect the basic and novel characteristics of the claimed invention. The present disclosure contemplates embodiments of the invention compositions and methods corresponding to the scope of each of these phrases. Thus, a composition or method comprising recited elements or steps contemplates particular embodiments in which the composition or method consists essentially of or consists of those elements or steps.

[00203] Reference throughout this specification to "one embodiment" or "an embodiment" or "an aspect" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, the appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[00204] The various embodiments described above can be combined to provide further embodiments. These and other changes can be made to the embodiments in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

[00205] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A method of detecting the presence or absence of a target nucleic acid analyte in a sample, the method comprising:
 - a) performing an amplification reaction on the sample;
 - b) measuring a signal generated during the amplification reaction and generating a kinetic signature from the measured signal; and
 - c) comparing the generated kinetic signature to a reference signature and determining whether the generated kinetic signature corresponds with the reference signature, thereby detecting the presence or absence of the target nucleic acid analyte in the sample.
2. The method of claim 1, wherein the kinetic signature comprises a signal selected from an electromagnetic signal, a wavelength, and a fluorescence emission signal.
3. The method of claim 1, wherein the kinetic signature is plotted as a curve.
4. The method of claim 1, wherein the reference signature corresponds with a kinetic signature generated by amplification of the target nucleic acid analyte under one or more members of a set of constrained amplification parameters, the set of constrained amplification parameters comprising any combination of: (i) a range of forward and/or reverse primer concentrations; (ii) a range of polymerase concentrations; (iii) polymerase type; (iv) a range of nucleotide concentrations; (v) chromophore concentration; and (vi) chromophore type; (vii) a number of thermocycles; (viii) rate of thermocycling; (ix) one or more thermocycle stage temperatures; and (x) one or more thermocycle stage time lengths; (xi) a range of target-specific molecular probes; (xii) primers associated with one or more target-specific probes; and (xiii) a concentration of one or more enzyme cofactors.
5. The method of claim 4, wherein the amplification reaction is carried out within one or more of the constrained amplification parameters.
6. The method of claim 3, wherein the at least one of an exponential region, a linear region, and a plateau region of the curve occurs over at least 4 thermocycles or over at least 8 thermocycles.

7. A method of detecting the presence or absence of each of a plurality of target nucleic acid analytes in a sample, the method comprising:
 - a) adding a primer set, the primer set comprising at least two primer pairs, to the sample, at least one of the primer pairs comprising a limiting primer, and each primer pair defining a region of one of the target nucleic acid analytes to be amplified;
 - b) performing an amplification reaction on the sample;
 - c) measuring a signal generated during the amplification reaction; and
 - d) making a determination as to whether each of the plurality of target nucleic acid analytes is present based on the measured signal.
8. The method of claim 7, wherein step c) further comprises a step of generating a kinetic signature from the measured signal, and wherein step d) further comprises a step of comparing the kinetic signature to a reference signature, the reference signature corresponding with a kinetic signature generated by an amplification of one or more of the plurality of target nucleic acid analytes with the primer set, thereby making the determination.
9. The method of claim 7, wherein the primer set comprises one forward primer and two reverse primers or two forward primers and one reverse primer.
10. The method of claim 7, wherein the primer set comprises two or more nested primers.
11. The method of claim 7, wherein the signal is generated by exciting a molecule that binds an amplification product.
12. The method of claim 7, wherein
 - (A) the primer set comprises first and second primer pairs, the first primer pair having a first annealing temperature and the second primer pair having a second annealing temperature which is lower than the first annealing temperature, and the first primer pair defining a region of one of the target nucleic acid analytes, the second primer pair defining a region of another of the nucleic acid analytes; and wherein
 - (B) the amplification reaction comprises thermocycling of a set of temperature cycles comprising at least one temperature cycle having a

minimum temperature which is both (i) not greater than the first annealing temperature, and (ii) greater than the second annealing temperature, the reaction further comprising at least one other temperature cycle having a minimum temperature which is not greater than the second annealing temperature.

13. The method of claim 12, wherein the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the first annealing temperature and greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature.
14. The method of claim 12, wherein the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature which is both (x) not greater than the first annealing temperature and (y) greater than the second annealing temperature.
15. The method of claim 7, wherein step c) further comprises a step of generating a kinetic signature from the measured signal, and wherein step d) further comprises a step of comparing the kinetic signature to a reference signature, the reference signature corresponding to a signal generated when one or more of the plurality of target nucleic acid analytes is amplified with the primer set by a thermocycling reaction comprising the set of temperature cycles, thereby making the determination.
16. The method of claim 12, wherein the limiting primer is present at a concentration sufficiently lower than the concentration of a non-limiting primer to generate a characteristic kinetic signature.
17. The method of claim 12, wherein the first and/or second annealing temperatures are determined by the limiting primer.
18. A method of detecting the presence or absence of two or more of a plurality of target nucleic acid analytes in a sample, the method comprising:
 - a) adding a primer set comprising first and second primer pairs to the sample, the first primer pair having a first annealing temperature and the second primer pair having a second annealing temperature which is lower than the first annealing temperature, and the first primer pair defining a region of

- one of the target nucleic acid analytes, the second primer pair defining a region of another of the nucleic acid analytes;
- b) performing a thermocycling amplification reaction on the sample, the reaction comprising a set of temperature cycles comprising at least one temperature cycle having a minimum temperature which is both (i) not greater than the first annealing temperature, and (ii) greater than the second annealing temperature, the reaction further comprising at least one other temperature cycle having a minimum temperature which is not greater than the second annealing temperature;
 - c) measuring a signal generated during the amplification reaction; and
 - d) making a determination as to the presence or absence of the two or more of the nucleic acid analytes based on the generated signal.
19. The method of claim 18, wherein step c) further comprises a step of generating a kinetic signature from the measured signal, and wherein step d) further comprises a step of comparing the kinetic signature to a reference signature, the reference signature corresponding with a signal generated when one or more of the plurality of target nucleic acid analytes is amplified with the primer set by a thermocycling reaction comprising the set of temperature cycles, thereby making the determination.
20. The method of claim 18, wherein the first and second primer pairs comprise one forward primer and two reverse primers.
21. The method of claim 18, wherein the primer set comprises at least one pair of nested primers.
22. The method of claim 18, wherein the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the first annealing temperature and greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature.
23. The method of claim 18, wherein the amplification reaction comprises at least one temperature cycle which has a minimum temperature not greater than the second annealing temperature, followed by at least one temperature cycle which has a minimum temperature which is both (x) not greater than the first annealing temperature and (y) greater than the second annealing temperature.

24. The method of claim 18, wherein at least one of the primer pairs has a limiting primer, the limiting primer having a concentration-adjusted melting temperature that is equal to or greater than a non-limiting primer.
25. The method of claim 18, wherein the method comprises detecting the presence of n target nucleic acid analytes using n primer pairs and wherein the minimum temperature of each temperature cycle is lower than the minimum temperature of the previous temperature cycle.
26. The method of claim 18, wherein one or more primer of a primer pair is linked to an antibody.
27. A method of detecting a nucleic acid analyte, comprising:
 - a. performing a nested asymmetric amplification reaction on a sample, wherein said nested asymmetric amplification reaction comprises:
at least three nested primers, wherein at least one pair of primers of said at least three nested primer comprises a limiting primer and an excess primer;
 - b. measuring a signal generated during said nested asymmetric amplification reaction; and
 - c. determining whether said analyte is present based on said signal.
28. The method of claim 27, wherein said at least three nested primers bind to a single analyte.
29. The method of claim 27, wherein said limiting primer has a concentration-adjusted melting temperature that is equal to or greater than said excess primer.
30. The method of claim 27, wherein said analyte is encoded by at least one additional value selected from the group consisting of a value from said signal and a value from an additional signal.
31. The method of claim 27, wherein said at least one additional value is selected from the group consisting of a Förster resonance energy transfer (FRET) emission intensity, a FRET emission wavelength, an electrochemical signal, a chemiluminescence wavelength, a chemiluminescence intensity, a fluorescence bleaching rate, and a chemiluminescence bleaching rate, wherein said at least one additional value is Förster resonance energy transfer (FRET) emission intensity.
32. A method of determining an amount of nucleic acid analyte, comprising:

- a. performing an amplification reaction on a sample;
 - b. measuring a signal generated during said amplification reaction;
 - c. transforming said signal into a frequency domain, thereby generating a transformed signal;
 - d. comparing a value from said transformed signal to a value of reference signal; and
 - e. determining said amount of said analyte based on said comparing.
33. The method of claim 32, wherein said transforming is performed by a method selected from the group consisting of a Fourier Transform, a fast Fourier transform, a discrete Fourier transform, a discrete-time Fourier transform, and a wavelet transform, wherein said method detects the presence or absence of a plurality of analytes, and wherein said reference signal is in a database of reference signals.

FIG. 1

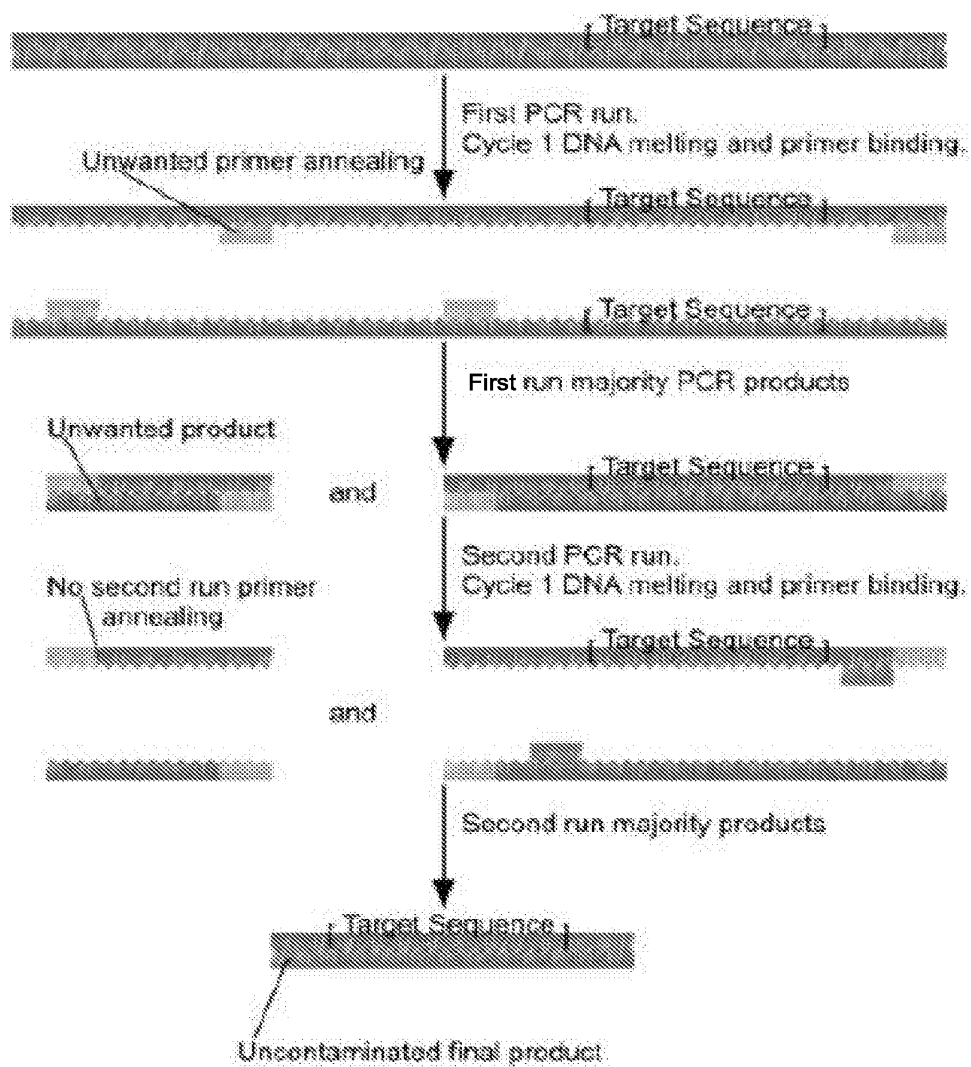


FIG. 2

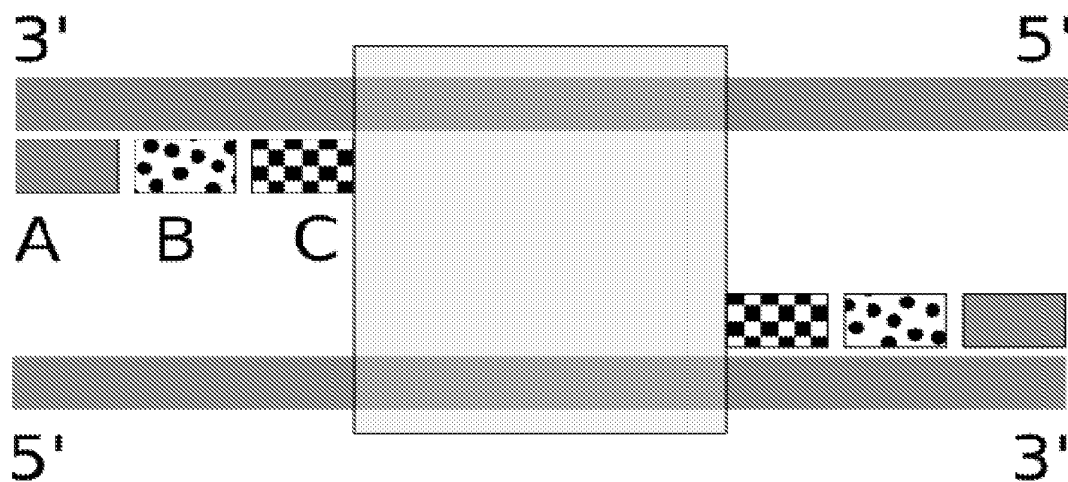


FIG. 3

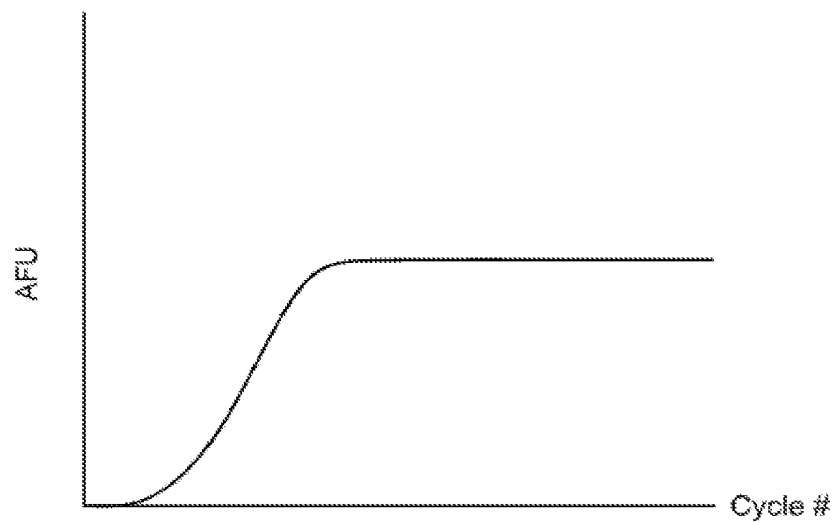


FIG. 4

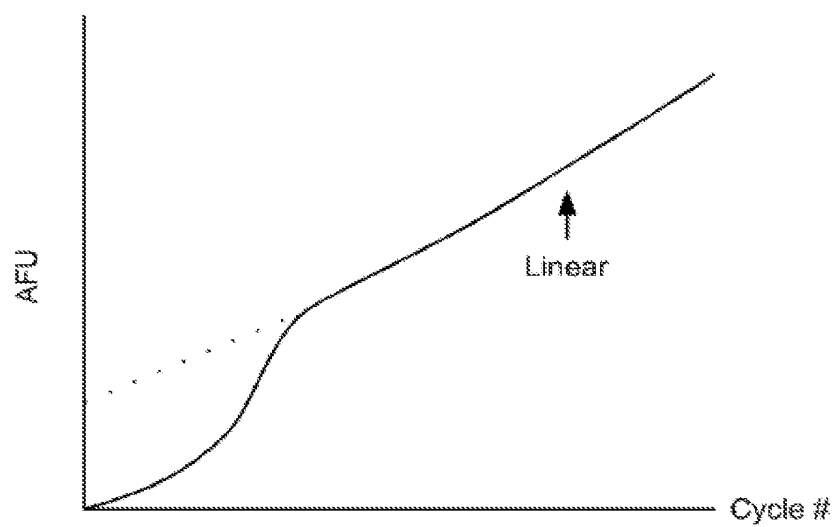


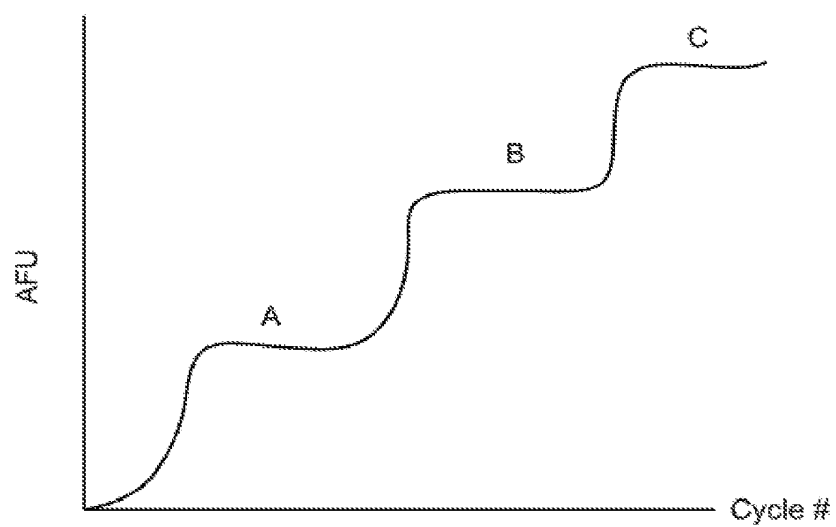
FIG. 5

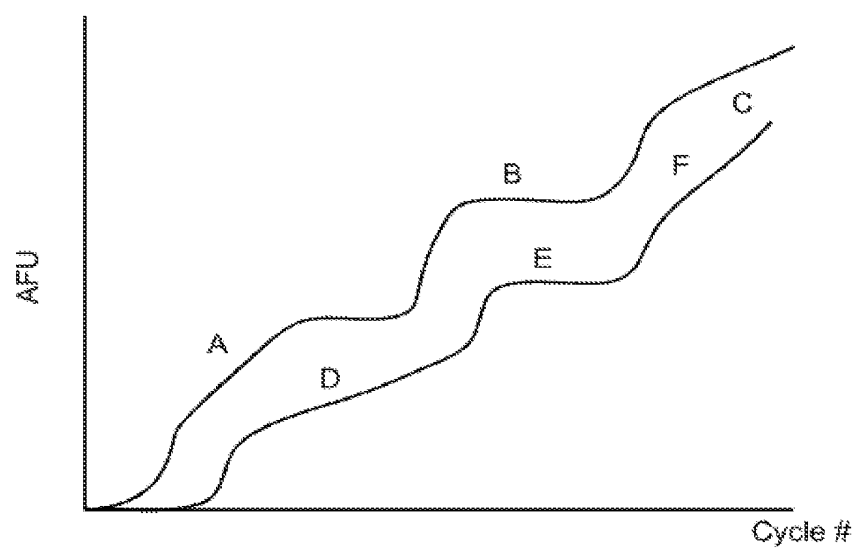
FIG. 6

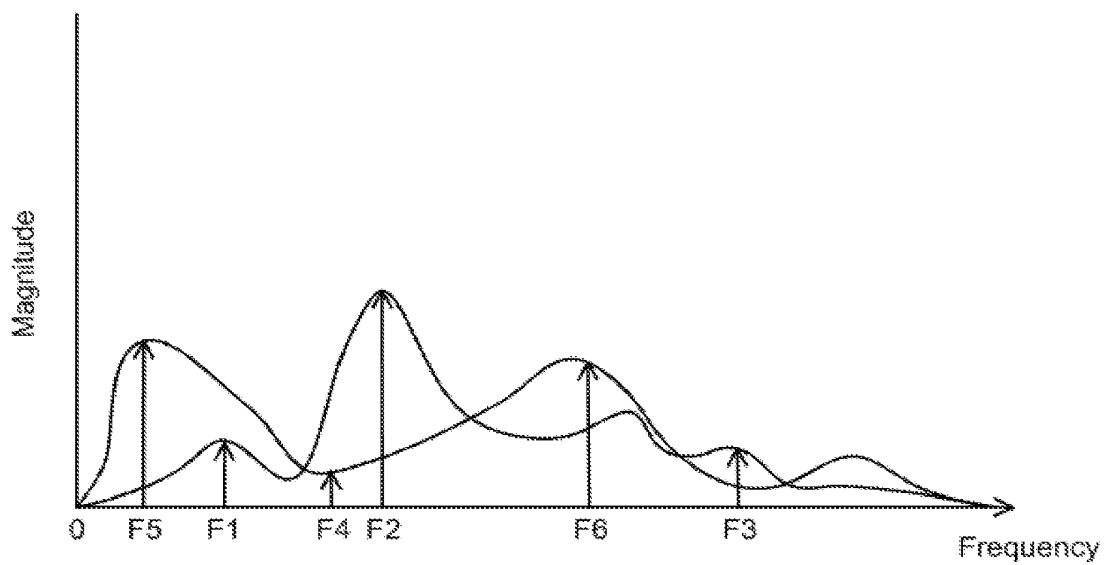
FIG. 7

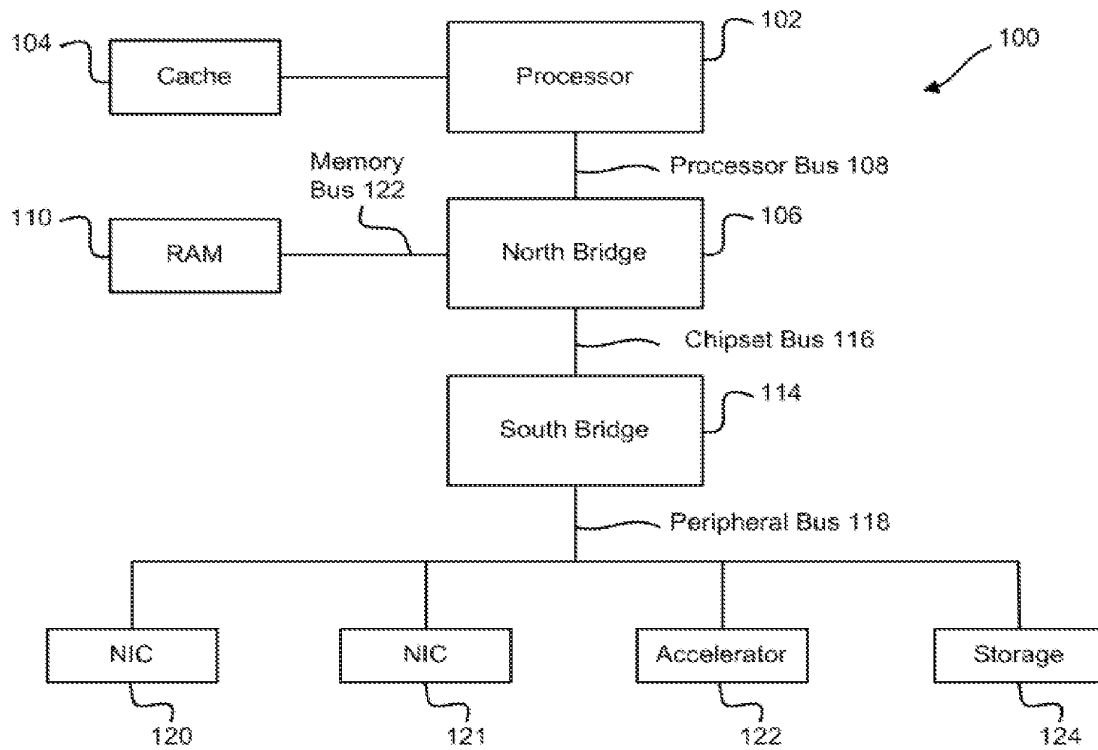
FIG. 8

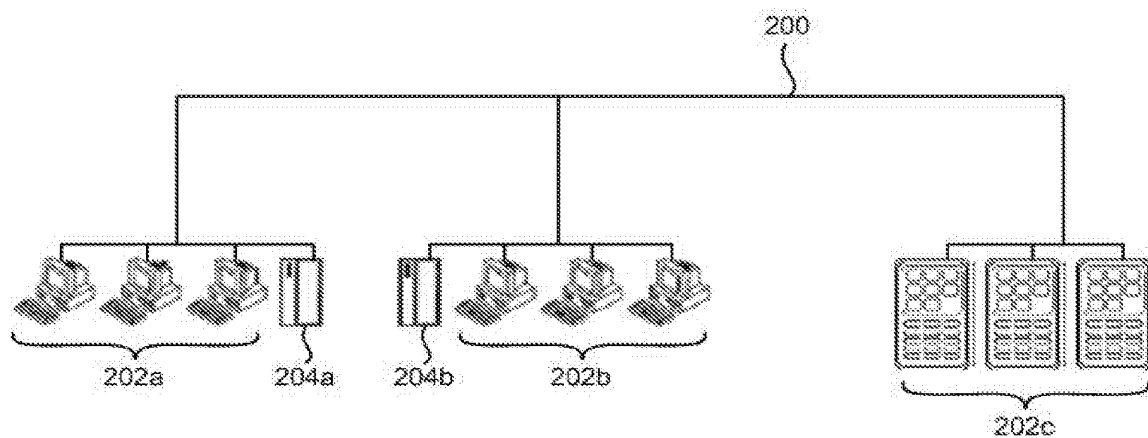
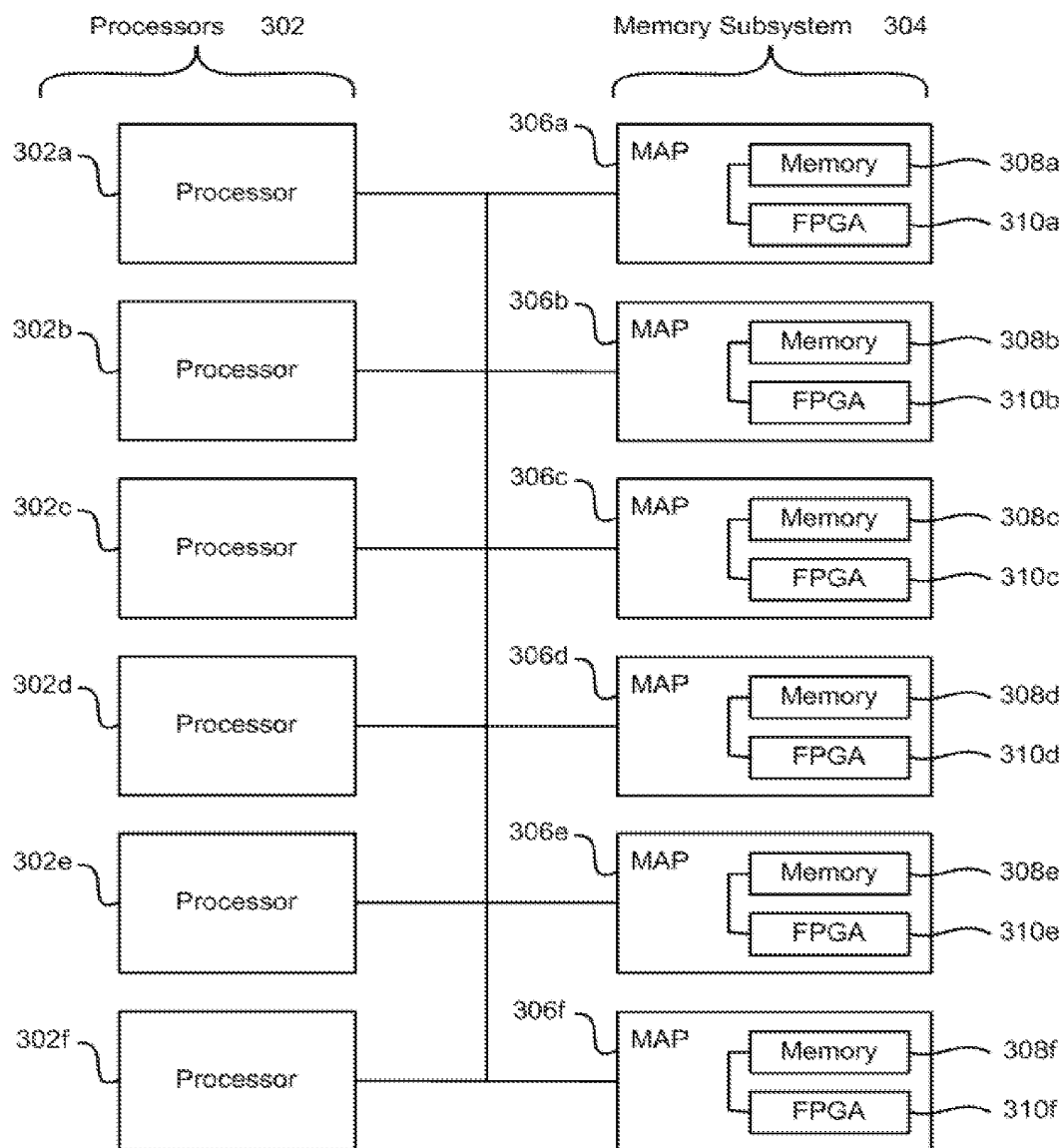
FIG. 9

FIG. 10



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/37682

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68, C12P 19/34 (2017.01)
CPC - C12Q 1/6844, C12N 15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2007/0072211 A1 (NEWTON et al.) 29 March 2007 (29.03.2007) Claim 1; Claim 3; para [0007]	1-6

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 August 2017

Date of mailing of the international search report

16 NOV 2017

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

IPR2024-01451

California Institute of Technology - Ex. 2003 Page 3579

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/37682

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

-----Please see continuation in first extra sheet-----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-6

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

IPR2024-01451

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/37682

Continuation of Box No III Observations where unity of invention is lacking

Group I: Claims 1-6, drawn to a method of detecting an analyte comprising generating a kinetic signature from a signal generated during an amplification reaction.

Group II: Claims 7-17, 27-31, drawn to methods of detecting target analytes comprising performing an amplification reaction by using a primer set that comprises at least one limiting primer.

Group III: Claims 18-26, drawn to a method of detecting an analyte comprising performing an amplification on a sample by using a primer set that comprises primer pairs with different annealing temperatures.

Group IV: Claims 32-33, drawn to a method of determining an amount of nucleic acid analyte, comprising generating transformed signal by transforming a signal generated during an amplification reaction into a frequency domain.

The inventions listed as Groups I, II, III, and IV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features

Group I requires generating a kinetic signature from a signal generated during an amplification reaction, not required by Groups II, III, and IV.

Group II requires the detection of target analytes by using a set of primers comprising at least one limiting primers, not required by Groups I, III, and IV.

Group III requires the detection of target analytes by using a set of primers by using a primer set that comprises primer pairs with different annealing temperatures, not required by Groups I, II, and IV.

Group IV requires generating transformed signal by transforming a signal generated during an amplification reaction into a frequency domain, not required by Groups I, II, and III.

Common Technical Features

The feature shared by Groups I, II, III, and IV is a method of detecting a nucleic acid analyte in a sample, the method comprising: performing an amplification reaction on the sample; measuring a signal generated during the amplification reaction; and detecting the presence or absence of the target nucleic acid analyte in the sample.

Another feature shared by Groups I and IV is the method step of: comparing a test value to a reference value and determining whether the test value corresponds with the reference value.

However, these shared technical features do not represent a contribution over prior art, because the shared technical features are taught by US 2007/0072211 A1 to Newton et al., (hereinafter 'Newton').

Newton teaches a method of detecting the presence or absence of a target nucleic acid analyte in a sample (para [0011] "the invention comprises methods for identifying one or more nucleic acid targets in a sample"), the method comprising: performing an amplification reaction on the sample (para [0011] "Such methods comprise: performing asymmetric kinetic PCR in a reaction mixture containing one or more labeled 5'-nuclease probes and one or more labeled hybridization probes"); measuring a signal generated during the amplification reaction (para [0011] "monitoring one or more growth curves from the kinetic PCR"); and detecting the presence or absence of the target nucleic acid analyte in the sample (para [0011] "thus, identifying the nucleic acid targets in the sample").

Newton further teaches the method step of: comparing a test value to a reference value and determining whether the test value corresponds with the reference value (para [0011] "correlating the melting curve or annealing curve, thus produced, to standard melting or annealing curves of completely or partially complementary probes of known nucleic acid targets or the same probe tested against a known sample, thus, identifying the nucleic acid targets in the sample").

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

-----Please see continuation in next extra sheet-----

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/37682

Continuation of Box No III Observations where unity of invention is lacking

Another feature shared by Group II and III is the detection of target analytes by using a set of primers.

However, this shared technical feature does not represent a contribution over prior art, because the shared technical feature is taught by US 2010/0233686 A1 to Higuchi et al., (hereinafter 'Higuchi').

Higuchi teaches detection of target analytes by using a set of primers (para [0012] "The present invention is a method for amplification, detection and quantification of one or more target nucleic acids in a single sample container comprising the steps of contacting a sample, suspected of containing one or more target nucleic acids, with at least one set of oligonucleotides, each oligonucleotide within the set labeled with the same one or more reporter moieties"; para [0042] "amplification of several target nucleic acid sequences in the same reaction vessel. Each tube contains one or more sets of probes, each set labeled with the same fluorescent reporter. However, within the set, each probe is characterized by a unique melting temperature with its target nucleic acid. For the maximum multiplex ability, several sets of probes are present in the same reaction vessel"; para [0057] "In some embodiments, the present invention involves asymmetric PCR. In an asymmetric PCR mixture, one of the amplification primers is present in greater amount than the other primer. The primers are referred to as "excess primer" and "limiting primer" respectively. The nucleic acid strands resulting from the extension of these primers are referred to as "excess strand" and "limiting strand" respectively").

As the technical feature was known in the art at the time of the invention, it cannot be considered a special technical feature that would otherwise unify the groups.

Groups I, II, III, and IV therefore lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.



- (51) **International Patent Classification:**
G01N 27/447 (2006.01) *G01N 33/53* (2006.01)
G01N 33/487 (2006.01)
- (21) **International Application Number:**
PCT/US2017/058159
- (22) **International Filing Date:**
24 October 2017 (24.10.2017)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/412,221 24 October 2016 (24.10.2016) US
PCT/US2017/025585
31 March 2017 (31.03.2017) US
- (71) **Applicant:** TWO PORE GUYS, INC. [US/US]; 2161 Delaware Ave, #B, Santa Cruz, CA 95060 (US).
- (72) **Inventors:** ZHAO, Yanan; c/o Two Pore Guys, Inc., 2161 Delaware Ave, #B, Santa Cruz, CA 95060 (US). MCKENNA, William; c/o Two Pore Guys, Inc., 2161 Delaware Ave, #B, Santa Cruz, CA 95060 (US). DUNBAR, William, B.; c/o Two Pore Guys, Inc., 2161 Delaware Ave, #B, Santa Cruz, CA 95060 (US).
- (74) **Agent:** HURT, Nicholas et al.; Fenwick & West LLP, 801 California Street, Mountain View, CA 94041 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, ST, 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).

(54) **Title:** FRACTIONAL ABUNDANCE OF POLYNUCLEOTIDE SEQUENCES IN A SAMPLE

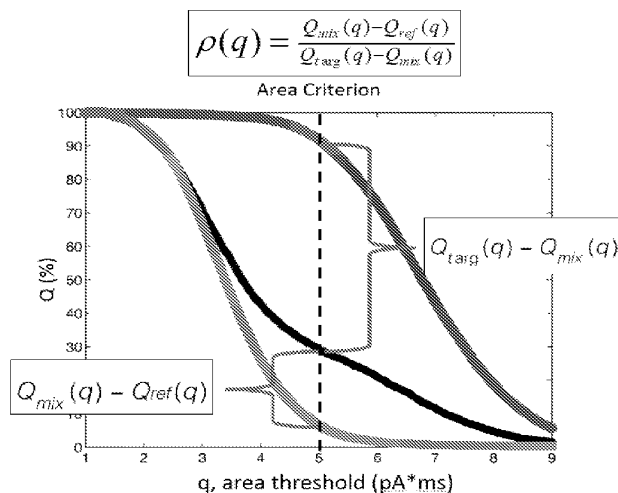


Figure 4C

(57) **Abstract:** Disclosed herein are methods and compositions for determining an improved estimate of the true fractional abundance of target analytes (e.g., specific polynucleotide sequences) in a sample using a nanopore sensor, e.g., by correcting errors inherent to identifying and correlating electrical signals to amounts of a target analyte or reference analyte in a sample.

[Continued on next page]

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

TITLE

[0001] Fractional Abundance of Polynucleotide Sequences in a Sample

CROSS-REFERENCE TO RELATED APPLICATIONS

[0002] This application claims the benefit of US Provisional Application No. 62/412,221, filed October 24, 2016, and International Application No. PCT/US2017/025585, filed March 31, 2017, the contents of which are each incorporated by reference in their entirety.

FIELD OF THE INVENTION

[0003] A method to determine the fractional abundance of specific polynucleotide sequences from a sample using solid state nanopores and mathematical methods for precise and accurate quantitation.

BACKGROUND OF THE INVENTION

[0004] Characterization of a liquid sample by determining a relative abundance of components present in the sample can provide valuable information for many scientific fields and applications. For example, a relative abundance of a point mutation in circulating cell free DNA can be used to diagnose or monitor progression of cancer in a patient. As another example, determining the fractional amount of a transgenic sequence of a genetically modified organism (GMO) to the non-GMO reference sequence within genomic DNA, obtained from a collection of seeds for example, is important for regulatory and economic reasons.

[0005] Some methods exist for sensitive detection of a fractional amount of a target analyte in a sample, however, these methods are usually expensive and time-consuming, or have other limitations. For example, quantitative real-time PCR (qPCR) assays remain the standard method used to determine the relative quantity of target nucleic acid sequences to an invariant reference sequence within a test sample. The quantitative performance of qPCR is, however, limited by variability in amplification efficiencies per sample, and per amplicon. Factors that affect amplification efficiency include inhibitors and carryover contaminants from the sample substrate and the extraction reagents themselves. These factors vary by sample and prep, but also in the degree to which they affect the amplification efficiency of one sequence versus another. Slight, variable differences in the amplification efficiency of target versus reference amplicons limit qPCR to resolving quantity differences $> 1.5\times$ -fold.

Furthermore, amplification reactions require specialized reagent sets and must be properly stored, and can be time consuming and sensitive to reaction conditions.

[0006] The use of nanopore devices have emerged as a sensitive tool for single molecule identification, wherein individual molecules are identified upon translocation through the nanopore under an applied voltage. Nanopore devices are amenable to point of use applications, and be sufficiently inexpensive and efficient for routine daily use cases, in human health, agriculture, or anywhere else. However, the use of data from a nanopore can be subject to errors that can impact a determination of quantitative estimates of analytes in a sample, such that the reliable use of this data is not feasible.

[0007] What is needed, therefore, are improved methods of determining fractional abundance of a target analyte compared to a reference analyte in a sample that are versatile, inexpensive and easy to use.

SUMMARY OF THE INVENTION

[0008] According to some embodiments, provided herein is a method of determining an improved estimate of a true relative abundance of a target analyte in a mixed unknown sample using a nanopore device, comprising applying a voltage across a nanopore in a nanopore device to generate a detectable electronic signature and to induce translocation of charged analytes through said nanopore separately for each of: a control sample comprising a known relative abundance of target analytes to reference analytes, and a mixed unknown sample comprising said target analytes and said reference analytes, wherein the relative abundance of said target analytes in said sample is to be determined; generating a plurality of event signatures generated by translocation of said target analytes or said reference analytes through said nanopore for each sample; identifying a quantity of first event signatures associated with said target analyte and a quantity of second event signatures associated with said reference analyte from said plurality of event signatures to determine a detected relative abundance of first and second event signatures for each sample; and adjusting a detected relative abundance of said first and second event signatures in said mixed unknown sample using the detected relative abundance of said first and second event signatures in said control sample to correct for an error in the detected relative abundance, thereby determining an improved estimate of the true relative abundance of said target analyte in said mixed unknown sample. In some embodiments the sample is a liquid sample.

[0009] In some embodiments, the control sample is a target control sample comprising said target analytes, but not said reference analytes. In some embodiments, the control sample is a reference control sample comprising said reference analytes, but not said target analytes.

[0010] In some embodiments, the method of determining an improved estimate of a true relative abundance of a target analyte in a mixed unknown sample using a nanopore device further comprises applying a voltage to a nanopore device to induce translocation of charged analytes through a nanopore sensor for a target control sample comprising said target analytes, but not said reference analytes.

[0011] In some embodiments, the adjustment of said detected relative abundance of said first and second event signatures in said unknown sample comprises using the detected relative abundance of said first and second event signatures in said target control sample and in said reference control sample to correct for said error in the detected relative abundance. In some embodiments, the error comprises a false positive or a false negative detection error of said target analyte.

[0012] In some embodiments, the method of determining an improved estimate of a true relative abundance of a target analyte in a mixed unknown sample using a nanopore device further comprises applying a voltage to a nanopore device to induce translocation of charged analytes through a nanopore sensor for a mixed control sample comprising said target analytes and said reference analytes, wherein the relative abundance of said target analytes and said reference analytes is known.

[0013] In some embodiments, the adjustment of said detected relative abundance of said first and second event signatures in said unknown sample comprises using the detected relative abundance of said first and second event signatures in said target control sample, said reference control sample, and said mixed control sample to correct for said error in the detected relative abundance.

[0014] In some embodiments, the error comprises a false positive target analyte detection error, a false negative target analyte detection error, a capture rate constant differential between said target analyte and said reference analyte, or any combination thereof.

[0015] In some embodiments, the control sample is a mixed control sample comprising said target analytes and said reference analytes, wherein the relative abundance of said target analytes and said reference analytes is known. In some embodiments, the error comprises a capture rate constant differential between said target analyte and said reference analyte.

[0016] In some embodiments, the mixed control sample comprises a relative abundance of said target analytes to said reference analytes that differs by no more than a factor of 1.2, a

factor of 1.5, a factor of 2, a factor of 5, or a factor of 10 relative to said mixed unknown sample.

[0017] In some embodiments, the estimate of the true relative abundance is an estimate of the true ratio of said target analyte to said reference analyte in said mixed unknown sample. In some embodiments, the estimate (R_{mix}^*) of the true ratio is determined by $R_{mix}^* = \rho\alpha$, wherein the parameter ρ is an estimate for the ratio that can compensate for a false positive detection error, a false negative detection error, or both, and wherein the parameter α can be used to compensate for a capture rate constant differential between said target analyte and said reference analyte. In some embodiments, the parameter α is an estimate of the ratio of the reference analyte capture rate divided by the target analyte capture rate.

[0018] In some embodiments, the estimate of the true relative abundance is an estimate of the true fraction of said target analytes in a population of said reference analytes and said target analytes in said mixed unknown sample. In some embodiments, the estimate of the true fraction (F_{mix}^*) is determined by $F_{mix}^* = \frac{\rho\alpha}{\rho\alpha+1}$, wherein the parameter ρ is an estimate for the ratio that can compensate for a false positive detection error, a false negative detection error, or both, and wherein the parameter α can be used to compensate for a capture rate constant differential between said target analyte and said reference analyte. In some embodiments, the parameter α is an estimate of the ratio of the reference analyte capture rate divided by the target analyte capture rate

[0019] In some embodiments, the parameter $\rho = \left(\frac{Q_{mix}-Q_{ref}}{Q_{targ}-Q_{mix}} \right)$, and $\alpha = \left(\frac{Q_{targ}-Q_{X:Y}}{Q_{X:Y}-Q_{ref}} \right) \times \frac{X}{Y}$.

In some embodiments, the parameter Q_{targ} is the fraction of said first event signature observed in said target control sample if said control sample is used, or $Q_{targ} = 1$ if no target control sample is used. In some embodiments, the parameter Q_{ref} is the fraction of said first event signature observed in said reference control sample if said reference control sample is used, or $Q_{ref} = 0$ if no reference control sample is used. In some embodiments, the parameter $Q_{X:Y}$ is the fraction of said first event signature observed in said mixed control sample and wherein $\frac{X}{Y}$ is the known ratio of target analytes (X) to reference analytes (Y) in the mixed control sample if said control sample is used, or $\alpha = 1$ if no mixed control sample is used. In some embodiments, the parameter Q_{mix} is the fraction of said first event signature observed in said mixed unknown sample.

[0020] In some embodiments, the unknown or control sample is prepared by nucleic acid amplification. In some embodiments, the unknown or control sample is not prepared by

nucleic acid amplification. In some embodiments, the sample is purified to substantially consist of reference and target molecules. In some embodiments, the sample is not purified.

[0021] In some embodiments, the quantity or concentration of said reference analytes in said mixed unknown sample are known. In some embodiments, the method of determining an improved estimate of a true relative abundance of a target analyte in a mixed unknown sample using a nanopore device further comprises determining an estimate of the absolute quantity or concentration of said target analytes in said mixed unknown sample using said estimate of the true relative abundance of said target analytes to said reference analytes in said mixed unknown sample and said known quantity or concentration of said reference analytes in said mixed unknown sample.

[0022] In some embodiments, the quantity of first event signatures associated with said target analyte and said quantity of second event signatures associated with said reference analyte are identified according to a defined threshold. In some embodiments, the method of determining an improved estimate of a true relative abundance of a target analyte in a mixed unknown sample using a nanopore device further comprises optimizing said threshold to increase accuracy of detection of said reference analytes and/or said target analytes using a Q-test, a support vector machine, or an expectation maximization algorithm. In some embodiments, the support vector machine is trained using electronic signatures from control samples comprising known quantities of target analytes and reference analytes.

[0023] In some embodiments, the defined threshold is a function of one or more features of an event signature selected from the group consisting of: an event duration, a maximum δG , a median δG , an average δG , a standard deviation of the event signature, a mean or median of the noise power of the event below 50 Hz, a unique pattern in said event signature, an area of an event, or any combination thereof.

[0024] In some embodiments, the adjustment of said detected relative abundance of said first and second event signatures in said mixed unknown sample to correct for said error in the detected relative abundance is performed using a Q-test, a support vector machine, or an expectation maximization algorithm.

[0025] In some embodiments, the target analyte and said reference analyte each comprise a polynucleotide. In some embodiments, the target analyte polynucleotide and said reference analyte polynucleotide are of different lengths. In some embodiments, the lengths are different by at least 10 nucleotides, at least 20 nucleotides, at least 50 nucleotides, at least 100 nucleotides, at least 150 nucleotides or at least 200 nucleotides.

[0026] In some embodiments, the method of determining an improved estimate of a true relative abundance of a target analyte in a mixed unknown sample using a nanopore device further comprises contacting said control or unknown samples with a first probe bound to a first payload, wherein said first probe is configured to bind specifically to said first analyte. In some embodiments, the method of determining an improved estimate of a true relative abundance of a target analyte in a mixed unknown sample using a nanopore device further comprises contacting said control or unknown samples with a second probe bound to a second payload, wherein said second probe is configured to bind specifically to said second analyte.

[0027] In some embodiments, the target analyte is correlated with a genetically modified organism. In some embodiments, the target analyte comprises a marker associated with the presence or absence of cancer in a patient.

[0028] Also provided herein is a method of determining a relative quantity of a target analyte in a sample comprising running separately in a nanopore system each of: a first control sample comprising reference analytes and no target analytes, a second control sample comprising target analytes and no reference analytes, a third control sample comprising a known relative abundance of said target analytes and said reference analytes, and an experimental sample comprising an unknown relative abundance of said target analytes and said reference analytes; detecting a quantity of first event signatures associated with a reference analyte and a quantity of second event signatures associated with a target analyte for each sample; and comparing a relative abundance of said quantity of first and second event signatures from said experimental sample with a relative abundance of said quantity of first and second event signatures from each of said first control sample, said second control sample, and said third control sample to determine an estimate of the true relative abundance of said reference analyte and said target analyte in said experimental sample.

[0029] In some embodiments, the event signature comprises an electrical signal induced by translocation of said reference analyte through said nanopore.

[0030] In some embodiments, the target analyte and said reference analyte each comprise a polynucleotide. In some embodiments, the reference analyte and said target analyte are discriminated by length

[0031] In some embodiments, the reference analyte and said target analyte are each bound to a sequence-specific probe comprising a payload to facilitate discrimination between said reference analyte and said target analyte in said nanopore device.

[0032] In some embodiments, the relative abundance is a fractional amount of said target analyte compared to the total population of the target analyte and the reference analyte in said sample.

[0033] Also provided herein is a method of determining a relative abundance of a target analyte in an unknown sample, comprising providing an unknown sample comprising a plurality of reference analytes and a plurality of target analytes; loading said unknown sample into a first chamber of a nanopore device comprising a nanopore disposed between said first chamber and a second chamber; applying a voltage across said nanopore to pass said reference analytes and said target analytes through said nanopore from said first chamber to said second chamber; detecting a number of first electrical signals each associated with the translocation of said reference analyte through the nanopore; detecting a number of second electrical signals each associated with the translocation of said target analyte through the nanopore; and converting a relative abundance of the number of detected first electrical signals and the number of detected second electrical signals to an estimate of the true relative abundance of said target analyte in said unknown sample using a reference value that accounts for at least one error associated with said electrical signal relative abundance.

[0034] In some embodiments, the reference value is determined from a fractional abundance of said first electrical signal determined from a mixed control sample comprising a known amount of target analytes and reference analytes. In some embodiments, the reference value is determined from a fractional abundance of said first electrical signal determined from a mixed control sample comprising a known amount of target analytes and reference analytes. In some embodiments, the reference value is determined from a fractional abundance of said first electrical signal determined from a mixed control sample comprising a known amount of target analytes and reference analytes.

[0035] In some embodiments, the mixed control sample, said target control sample, or said reference control sample is run in said nanopore device under conditions substantially identical to conditions in said nanopore device during said detection of said first and second electrical signals from said unknown sample.

[0036] In some embodiments, the nanopore device comprises a membrane that separates an interior space of the device into a first chamber and a second chamber, wherein said membrane comprises said nanopore, wherein said first chamber and said second chamber are in fluid communication through said nanopore, and wherein said device comprises an electrode in each chamber for applying a voltage across said nanopore. In some

embodiments, the electrodes are configured to monitor electrical current through said nanopore. In some embodiments, the electrodes are connected to a power supply.

[0037] In some embodiments, the methods provided herein improve the accuracy of an estimate of fractional abundance of a target analyte in a mixed unknown sample by accounting for false positive or false negative detection errors, or a capture rate constant differential between said target analyte and said reference analyte. In some embodiments, a series of controls is run to improve the accuracy of the estimate of fractional abundance, including a reference-only control to account for false positive target analyte detection errors, a target-only control to account for false negative target analyte detection errors, and one or more mixed control samples to account for a capture rate constant differential between the target analyte and the reference analyte.

[0038] In some embodiments, the capture rates between the target analyte and the reference analyte in the mixed unknown sample are relatively consistent, such that the mixed control does not need to be used to improve the estimate of the relative abundance. In some embodiments, the relative capture rates between the target analyte and the reference analyte in a mixed sample are known such that a correction term can be applied to data from a mixed unknown sample to compensate for this difference to improve the estimate of fractional abundance without running a mixed control sample. In some embodiments, data from a mixed control sample run under substantially identical nanopore conditions using the same target analyte and reference analyte species as in the mixed unknown sample is used to improve the estimate of the fractional abundance without actually running the mixed control sample as part of the method.

[0039] In some embodiments, a threshold value is determined such that a false positive value from the mixed unknown sample is negligible, and a reference-only control does not need to be used to improve the estimate of the relative abundance. In some embodiments, the false positive value from a mixed sample is known such that a correction term can be applied to data from a mixed unknown sample to compensate for a false positive error to improve the estimate of fractional abundance without running a reference-only control sample. In some embodiments, data from a reference-only control sample run under substantially identical nanopore conditions using the same reference analyte species as in the mixed unknown sample is used to improve the estimate of the fractional abundance without actually running the reference-only control as part of the method.

[0040] In some embodiments, a threshold value is determined such that a false negative value from the mixed unknown sample is negligible, and a target-only control does not need

to be used to improve the estimate of the relative abundance. In some embodiments, the false negative value from a mixed sample is known such that a correction term can be applied to data from a mixed unknown sample to compensate for a false negative error to improve the estimate of fractional abundance without running a target-only control sample. In some embodiments, data from a target-only control sample run under substantially identical nanopore conditions using the same target analyte species as in the mixed unknown sample is used to improve the estimate of the fractional abundance without actually running the target-only control as part of the method.

[0041] In some embodiments, provided herein is a method of a determining an estimate of a relative abundance of a target analyte to a reference analyte in a mixed sample, comprising applying a voltage to a nanopore device to induce translocation of charged analytes through a nanopore sensor separately for each of: a mixed control sample comprising a known relative abundance of target analytes to reference analytes, and a mixed unknown sample comprising said target analytes and said reference analytes, wherein the relative abundance of said target analytes to said reference analytes is unknown; detecting a quantity of first event signatures associated with said reference analyte and a quantity of second event signatures associated with a target analyte for each sample; and determining an estimate of the true relative abundance of said target analytes to said reference analytes in said mixed unknown sample by adjusting a detected relative abundance of said first and second event signatures from said mixed unknown sample using the detected relative abundance of said first and second event signatures in said mixed control sample and the true relative abundance of said target analytes to said reference analytes in said mixed control sample.

[0042] In some embodiments, provided herein is a method of a determining an estimate of a relative abundance of a target analyte to a reference analyte in a mixed sample, comprising applying a voltage to a nanopore device to induce translocation of charged analytes through a nanopore sensor separately for each of: a target control sample comprising target analytes, but not reference analytes, a reference control sample comprising reference analytes, but not target analytes, and a mixed unknown sample comprising said target analytes and said reference analytes, wherein the relative abundance of said target analytes to said reference analytes is unknown; detecting a quantity of first event signatures associated with said reference analyte and a quantity of second event signatures associated with a target analyte for each sample; and determining an estimate of the true relative abundance of said target analytes to said reference analytes in said mixed unknown sample by adjusting a detected relative abundance of said first and second event signatures in said mixed unknown sample

using the detected relative abundance of said first and second event signatures in said target control sample and said reference control sample. In some embodiments, the target control sample provides a correction term for false negative detection of target analytes from said mixed unknown sample. In some embodiments, the reference control sample provides a correction term for false positive detection of target analytes in said mixed unknown sample.

[0043] In some embodiments, provided herein is a method of determining an estimate of a relative abundance of a target analyte to a reference analyte in a mixed sample, comprising applying a voltage to a nanopore device to induce translocation of charged analytes through a nanopore sensor separately for each of: a mixed control sample comprising a known relative abundance of target analytes to reference analytes, a target control sample comprising target analytes, but not reference analytes, a reference control sample comprising reference analytes, but not target analytes, and a mixed unknown sample comprising said target analytes and said reference analytes, wherein the relative abundance of said target analytes to said reference analytes is unknown; detecting a quantity of first event signatures associated with said reference analyte and a quantity of second event signatures associated with a target analyte for each sample; and determining an estimate of the true relative abundance of said target analytes to said reference analytes in said mixed unknown sample by adjusting a detected relative abundance of said first and second event signatures from said mixed unknown sample using the detected relative abundance of said first and second event signatures in said target control sample and said reference control sample, and the detected relative abundance of said first and second event signatures in said mixed control sample and the true relative abundance of said target analytes to said reference analytes in said mixed control sample.

[0044] In some embodiments, the method of determining an estimate of a relative abundance of a target analyte to a reference analyte in a mixed sample further comprises applying a voltage to a nanopore device to induce translocation of charged analytes through a nanopore sensor for a target control sample comprising said target analytes, but not said reference analytes.

[0045] In some embodiments, the method of determining an estimate of a relative abundance of a target analyte to a reference analyte in a mixed sample further comprises applying a voltage to a nanopore device to induce translocation of charged analytes through a nanopore sensor for a reference control sample comprising said reference analytes, but not said target analytes. In some embodiments, determining said estimate of the true relative abundance of said target analytes to said reference analytes in said mixed unknown sample

comprises adjusting said detected relative abundance of said first and second event signatures in said mixed unknown sample using the detected relative abundance of said first and second event signatures in said target control sample, said reference control sample, and said mixed control sample and the true relative abundance of said target analytes to said reference analytes in said mixed control sample.

[0046] In some embodiments, the mixed control sample comprises a relative abundance of said target analytes to said reference analytes that differs by no more than a factor of 1.2, a factor of 1.5, a factor of 2, a factor of 5, or a factor of 10 relative to said mixed unknown sample.

[0047] In some embodiments, the relative abundance comprises the ratio of target analyte:reference analyte. In some embodiments, the estimate of the true ratio (R_{mix}^*) of said target analyte to said reference analytes in said mixed unknown sample is determined by $R_{mix}^* = \rho\alpha$, wherein the parameter ρ is an estimate for the ratio that can compensate for a false positive detection error, a false negative detection error, or both, and wherein the parameter α can be used to compensate for a capture rate constant differential between said target analyte and said reference analyte. In some embodiments, the parameter α is an estimate of the ratio of the reference analyte capture rate divided by the target analyte capture rate.

[0048] In some embodiments, the relative abundance comprises the fraction of said target analyte in a population of said target analytes and said reference analytes. In some embodiments, the estimate of the true fraction (F_{mix}^*) of said target analyte in a population of said reference analytes and said target analytes in said mixed unknown sample is determined by $F_{mix}^* = \frac{\rho\alpha}{\rho\alpha+1}$, wherein the parameter ρ is an estimate for the ratio that can compensate for a false positive detection error, a false negative detection error, or both, and wherein the parameter α can be used to compensate for a capture rate constant differential between said target analyte and said reference analyte. In some embodiments, the parameter α is an estimate of the ratio of the reference analyte capture rate divided by the target analyte capture rate.

[0049] In some embodiments, provided herein is a kit comprising a control sample comprising a target analyte and a reference analyte at a known relative abundance; and instructions for use to run said control sample and an unknown sample comprising said reference analyte and said target analyte in a nanopore device to determine a relative abundance of said reference analyte and said target analyte in said unknown sample.

[0050] In some embodiments, provided herein is a kit comprising a first control sample comprising a target analyte, wherein said first control sample does not contain a reference analyte; a second control sample comprising said reference analyte, wherein said second control sample does not contain said target analyte; a third control sample comprising said target analyte and said reference analyte at a known relative abundance; and instructions for use to run said first control sample, said second control sample, said third control sample and an unknown sample comprising said reference analyte and said target analyte separately in a nanopore device to determine a relative abundance of said reference analyte and said target analyte in said unknown sample.

[0051] In some embodiments, provided herein is a computer-implemented method of determining an estimate of a true fractional abundance of a target analyte in a sample, comprising: obtaining data from a nanopore sensor from at least one of a reference analyte control or a target analyte control, wherein said data comprises a plurality of event signatures from target analytes or reference analytes translocating through said nanopore; identifying one or more features of event signatures to differentiate those correlated with target analytes and those correlated with reference analytes; training said support vector machine to identify an optimized threshold to distinguish said first events from said second events and to generate an estimate of the true relative abundance of said reference analytes and said target analytes in a sample, wherein said training comprises use of a control selected from the group consisting of a reference control sample, a target control sample, and a mixed control sample, and wherein training comprises validation using known mixed samples; and using said trained support vector to determine a fractional abundance of target analytes in a sample from events recorded on a nanopore device from a mixed sample.

[0052] In some embodiments, provided herein is a computer-implemented method of determining an estimate of a true fractional abundance of a target analyte in a sample, comprising: obtaining a set of data from a nanopore device, said data comprising event signatures from at least one control sample and at least one unknown sample; identifying a set of features to use for generating a threshold to discriminate first event signatures correlated with said target analytes from second event signatures correlated with said reference analytes; and estimating a true value of a fractional abundance in said unknown sample using a trained support vector machine.

BRIEF DESCRIPTION OF THE DRAWINGS

[0053] The foregoing and other objects, features and advantages will be apparent from the following description of particular embodiments of the invention, as illustrated in the accompanying drawings in which like reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead placed upon illustrating the principles of various embodiments of the invention.

[0054] Figure 1A shows a typical electronic signature of a single-molecule event caused by a dsDNA passing through a nanopore, which a characteristic duration of translocation and decrease in current during translocation.

[0055] Figure 1B shows an all-event scatter plot of max δG versus duration for 5.6 kb dsDNA recorded in a 22 nm diameter nanopore.

[0056] Figure 2A shows a typical event when a 727 bp DNA goes through a 25 nm diameter solid-state nanopore at 100 mV in 1M LiCl. The vent area is shaded.

[0057] Figure 2B illustrates an increase in event duration with increased dsDNA length, while event depth is conserved.

[0058] Figure 2C shows a plot of the distribution of the $\log_{10}$ of the area of all events recorded for dsDNA at each length shown.

[0059] Figure 3A depicts an example of a threshold generated between events from type 1 analytes (squares) and type 2 analytes (circles).

[0060] Figure 3B shows an example of the results of transformation of input features to a higher dimensional space to increase the accuracy of a linear threshold between events from type 1 analytes (squares) and type 2 analytes (circles).

[0061] Figure 4A shows a probability histogram for all events from a reference analyte sample, a target analyte sample, and a mixed sample according to event area.

[0062] Figure 4B depicts a graph of the percentage of events that are below an area threshold from reference analyte only (Q_{ref}), target analyte only (Q_{targ}), and a mixed sample of target analytes and reference analytes (Q_{mix}).

[0063] Figure 4C shows how fractional amount parameter $\rho(q)$ appears graphically at a q value. The $q = 5 \text{ pA}\cdot\text{ms}$ threshold (vertical dashed line) corresponds to a false positive of 0.05 (i.e., $Q_{ref} = 0.05$) and a false negative of 0.1 (i.e., $Q_{targ} = 0.9$).

[0064] Figure 5A shows the results of a determination of an estimate of a relative abundance of target gene (GMO (%)) (R_{mix}^*) vs. the true relative abundance of a target gene

(GMO (%)). 10% error margins above and below the zero-error line (slope = 1) are shown for comparison.

[0065] Figure 5B shows the results of a determination of an estimate of the true relative abundance of genetically modified organisms in a sample using two isolated controls and six known mixtures. Values of predicted target abundance percentage were plotted against true target abundance percentage. 10% error margins above and below the zero-error line (slope = 1) are shown for comparison.

[0066] Figure 6 shows the results of estimate of target analyte abundance (GMO (%)) over a range of thresholds for discriminating target analyte from reference analyte according to area of an event.

[0067] Figure 7 shows a prediction of accuracy across a set of testing data from a trained support vector machine with optimal parameters for discriminating event signatures from target and reference analytes.

[0068] Figure 8 shows an event plot for two molecule types (94bp target dsDNA bound to a probe/payload and 74bp reference dsDNA bound to a probe/payload) that were run as isolated controls sequentially on the same pore.

[0069] Figure 9A shows a representative event plot of mean δG vs. duration for the 100% target analyte control sample (closed circles) and the 100% reference analyte control sample (open squares) overlaid. The target analyte is 89bp DNA with G12D-bound probe linked to a 3-branch PEG (denoted G12D-3bPEG). The reference analyte is 89bp DNA with wild-type (c.35G)-bound probe linked to an 8-arm PEG (denoted WT-8armPEG). The thresholds for identifying an event signature as from a target analyte passing through the nanopore ($q_1 = 1$ msec, $q_2 = 0.4$ nS and $q_3 = 0.65$ nS) create the target tagging box (dashed line).

[0070] Figure 9B shows the plot from Figure 9A, with data from unknown sample A (triangle) and sample B (star) comprising target analytes and reference analytes overlaid onto the plot.

[0071] Figure 10 shows a representative event plot of mean δG vs. duration for the 100% target analyte control sample (closed circles) and the 100% reference analyte control sample (open squares) overlaid. Also plotted is the support vector machine-identified decision boundary (i.e. threshold) for discriminating target analytes from reference analytes.

[0072] Figure 11 shows events from a 50% target / 50% reference mixture sample plotted on an all-event scatter plot of max δG versus duration. The target domain box encompasses events associated with a probe-bound mutant targets.

[0073] Figure 12 shows the results of application of Expectation Maximization Algorithm for Gaussian Mixtures (EMGM) using a 3-Gaussian mixture model to the data from a 50% target / 50% reference mixture sample shown in Figure 11 for identification of target (mutant) and reference (wild-type) populations.

[0074] Figure 13 shows the results of application of EMGM using a 3-Gaussian mixture model to data from a reference-only control sample to establish a false positive fraction.

[0075] Figure 14 shows the results of application of EMGM using a 3-Gaussian mixture model to data from a mixed unknown sample to identify a relative abundance of mutant (target) molecules in the unknown sample.

DETAILED DESCRIPTION

[0076] The details of various embodiments of the invention are set forth in the description below. Other features, objects, and advantages of the invention will be apparent from the description and the drawings, and from the claims.

Definitions

[0077] Throughout this application, the text refers to various embodiments of the present nutrients, compositions, and methods. The various embodiments described are meant to provide a variety of illustrative examples and should not be construed as descriptions of alternative species. Rather it should be noted that the descriptions of various embodiments provided herein may be of overlapping scope. The embodiments discussed herein are merely illustrative and are not meant to limit the scope of the present invention.

[0078] Also throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

[0079] As used in the specification and claims, the singular form “a,” “an” and “the” include plural references unless the context clearly dictates otherwise. For example, the term “an electrode” includes a plurality of electrodes, including mixtures thereof.

[0080] As used herein, the term “comprising” is intended to mean that the devices and methods include the recited components or steps, but not excluding others. “Consisting essentially of” when used to define devices and methods, shall mean excluding other components or steps of any essential significance to the combination. “Consisting of” shall

mean excluding other components or steps. Embodiments defined by each of these transition terms are within the scope of this invention.

[0081] All numerical designations, *e.g.*, distance, size, temperature, time, voltage and concentration, including ranges, are approximations which are intended to encompass ordinary experimental variation in measurement of the parameters, and that variations are intended to be within the scope of the described embodiment. It is to be understood, although not always explicitly stated that all numerical designations are preceded by the term “about”. It also is to be understood, although not always explicitly stated, that the components described herein are merely exemplary and that equivalents of such are known in the art.

[0082] As used herein the term “analyte” refers to any molecule, compound, complex, or other entity whose presence can be detected using a nanopore sensor to facilitate determination of a relative abundance of an analyte in the pore. When referring to target or reference analytes, the term target or reference molecule may be used interchangeably.

[0083] As used herein, the term “target analyte” refers to a molecule or complex of interest in a sample. In some embodiments, the target analyte comprises portion of a polynucleotide having a sequence of nucleic acids of interest. The target analyte can be specifically targeted for binding by a probe to facilitate detection of the target analyte in a nanopore sensor, as described herein.

[0084] As used herein, the term “reference analyte” refers to a molecule or complex of interest in a sample, whose abundance is used as a relative measure of quantification for the target analyte. In some embodiments, the reference analyte comprises portion of a polynucleotide having a sequence of nucleic acids of interest. The reference analyte can be specifically targeted for binding by a probe to facilitate detection of the target analyte in a nanopore sensor, as described herein.

[0085] As used herein, the term “specific binding” or “bind specifically” refers to the targeted binding of a probe to a target analyte or a reference analyte.

[0086] As used herein, the term “probe” refers to a molecule that binds specifically to a target analyte or to a fragment thereof. In some embodiments, the probe comprises a payload molecule configured to affect the electronic signature generated upon translocation of a complex comprising a target or reference analyte bound to a probe-payload molecule or complex. In some embodiments, the probe comprises a payload molecule binding moiety adapted to bind to a payload molecule.

[0087] As used herein, the term “payload molecule” refers to a molecule with physical dimensions that facilitate generation of a unique electrical signal when captured in a