



(51) International Patent Classification:

G01N 33/53 (2006.01) C12Q 1/68 (2006.01)
C12N 15/11 (2006.01) G01N 33/52 (2006.01)

(21) International Application Number:

PCT/US2012/030940

(22) International Filing Date:

28 March 2012 (28.03.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/468,361 28 March 2011 (28.03.2011) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:

US 61/468,361 (CON)
Filed on 28 March 2011 (28.03.2011)

(71) Applicant (for all designated States except US):

NANOSTRING TECHNOLOGIES, INC. [US/US]; 530
Fairview Avenue North, Suite 2000, Seattle, WA 98109
(US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): WEBSTER, Phil-
ippa Jane [US/US]; 4317 Densmore Ave. North, Seattle,
WA 98103 (US). BOYKIN, Richard Kemble [US/US];
16741 10th Avenue NW, Shoreline, WA 98177 (US).
NUSSBAUM, Jeannette [DE/US]; 4202 S. Genesee
Street, Seattle, WA 98118 (US).

(74) Agents: ELRIFI, Ivor, R. et al.; Mintz Levin Cohn Ferris
Glovsky And Popeo, P.C., One Financial Center, Boston,
MA 02111 (US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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.

[Continued on next page]

(54) Title: COMPOSITIONS AND METHODS FOR DIAGNOSING CANCER

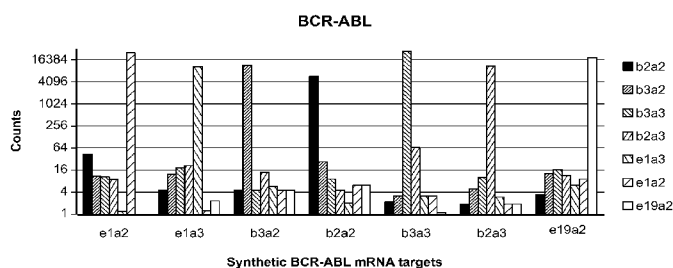


FIG. 1A

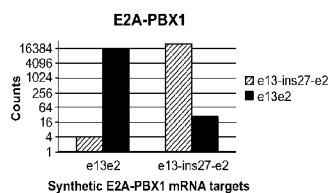


FIG. 1B

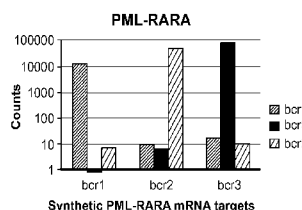


FIG. 1D

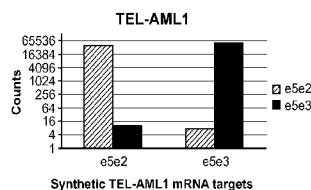


FIG. 1C

(57) Abstract: The application describes methods for dia-
gnosing subjects with leukemia by detecting fusion genes
associated with the onset of leukemia.



(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

COMPOSITIONS AND METHODS FOR DIAGNOSING CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This application claims the benefit of U.S. Provisional Application No. 61/468,361, filed March 28, 2011, the contents of which are incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[02] This disclosure relates generally to the field of cancer biology, and specifically, to the fields of detection and identification of specific cancer cell phenotypes and correlation with relevant patient prognosis.

BACKGROUND OF THE INVENTION

[03] Recurrent chromosomal translocations characterize a substantial proportion of leukemias, resulting in the formation of unique fusion genes. The identification of particular fusion gene abnormalities is of considerable diagnostic and prognostic importance. There is a need in the art for reliable diagnostic compositions for the detection and diagnosis of leukemias and sub-types of leukemias based on gene fusion.

SUMMARY OF THE INVENTION

[04] Leukemia is often accompanied by the formation of fusion genes. These chromosomal abnormalities are caused by the reciprocal translocation of portions of genes at different chromosomal locations forming a new gene characterized by the fusion of a nucleic acid sequence from a first and a second gene. For example, the *BCR-ABL* fusion gene is caused by a reciprocal translocation of parts of chromosome 9 and 22. Specifically, the translocated regions are 9p34 and 22q11. This joins a portion of the *ABL1* gene on chromosome 9 to a part of the *BCR* ("breakpoint cluster region") of chromosome 22. The presence of this and other fusion genes is correlated with the onset and progression of various types of leukemia. This disclosure presents compositions and methods for the diagnosis and prognosis of leukemia based on the detection of fusion genes.

[05] The disclosure provides a composition for the diagnosis of leukemia. In one embodiment, this composition includes one or more target-specific nucleic acid probes. Each of

the one or more target-specific probes includes a target-specific region that specifically binds to one target nucleic acid analyte; and a region to which is attached a plurality of label monomers that create a unique label for each target-specific probe. The label has a detectable signal that distinguishes one target-specific probe which binds to a first target nucleic acid from another target-specific probe that binds to a different second target nucleic acid. One or more targets of the one or more target-specific probes include a fusion gene or mRNA transcribed from said fusion gene. The fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia.

[06] The fusion genes can be caused by the fusion of two or more genes at the onset of leukemia. The fusion gene can be *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*. The fusions can be between varying exons of the upstream and downstream genes. The leukemia can be chronic myelogenous leukemia, acute myeloid leukemia or acute lymphoblastic leukemia.

[07] In certain embodiments, at least one of the one or more target-specific probes has a target that is the sequence of the junction where the two genes are fused. These target-specific probes can also include a target-specific region with one or more mismatches to the target of the target-specific region.

[08] In other specific embodiments, the label monomers can be a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety. The unique label for each target-specific probe can be composed of 4 or 5 label monomers.

[09] The disclosure also provides a composition for the diagnosis of leukemia and different sub-types of leukemia. In one embodiment, the composition comprises a probe pair including a first probe and a second probe. The first probe is made up, at least in part, of a complex including a first molecule. The first molecule includes a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal. The first molecule also includes a second label attachment region, which is non-overlapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal. The first molecule also includes a first target-specific sequence attached to the first molecule. The second probe is made up, at least in part, of a second molecule including a second target-specific sequence and an affinity tag. The first target-specific sequence and the second target-specific sequence bind to different regions of the same target

molecule. The target molecule is a fusion gene or mRNA transcribed from said fusion gene. The fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia. Also, when said probe pair is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitute at least part of a code that identifies the target molecule.

[10] In certain embodiments, the fusion gene results from the fusion of at least a portion of a first gene and at least a portion of a second gene. In this situation, the first target-specific sequence can hybridize to a portion of the first gene and the second target-specific sequence can hybridize to a portion of the second gene. The first target-specific sequence can hybridize to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.

[11] In other embodiments, the first target-specific sequence can also hybridize to a portion of the first gene and a portion of the second gene. In this situation, the first target-specific sequence can hybridize to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene. At least one of the target-specific sequences can include a region with one or more mismatches to the target of the target-specific sequence.

[12] The second target-specific sequence can also hybridize to a portion of the first gene and a portion of the second gene. In this situation, the first target-specific sequence can hybridize to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene. At least one of the target-specific sequences can include a region with one or more mismatches to the target of the target-specific sequence.

[13] The leukemia can be chronic myelogenous leukemia, acute myeloid leukemia or acute lymphoblastic leukemia. The fusion gene can be *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*.

[14] In certain embodiments, each of said label monomers can be a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety. The unique label for each target-specific probe can be composed of 4 or 5 label monomers.

[15] The disclosure also provides a method for diagnosing leukemia, a specific subtype of leukemia or determining a genetic aberration underlying leukemia in a subject. In one embodiment, the method includes providing a sample from the subject and contacting the sample with one or more target-specific probes. Each of the one or more target-specific probes includes

a target-specific region that binds to one target nucleic acid analyte; and a region to which is attached a plurality of label monomers that create a unique label for each target-specific probe. The label has a detectable signal that distinguishes one target-specific probe which binds to a first target nucleic acid from another target-specific probe that binds to a different second target nucleic acid. One or more targets of the one or more target-specific probes include a fusion gene or mRNA transcribed from said fusion gene. The fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia. When the fusion gene is detected the subject is diagnosed with leukemia, a specific subtype of leukemia or the genetic aberration underlying leukemia in a subject is determined.

[16] The fusion gene can be *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*. The leukemia can be chronic myelogenous leukemia, acute myeloid leukemia or acute lymphoblastic leukemia.

[17] In certain embodiments, at least one of the one or more target-specific probes can have a target that is the sequence of the junction of the two genes that are fused. The one or more target-specific probes can include a target-specific region with one or more mismatches to the target of the target-specific region.

[18] In other embodiments, each of the label monomers can be a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety. The unique label for each target-specific probe can be composed of 4 or 5 label monomers.

[19] The disclosure also provides a method for diagnosing leukemia, a specific subtype of leukemia or determining a genetic aberration underlying leukemia in a subject. In one embodiment, the method includes providing a sample from the subject and contacting the sample with a probe pair. The probe pair includes a first probe and a second probe. The first probe is made up, at least in part, of a complex including a first molecule with a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal. The first probe also includes a second label attachment region, which is non-overlapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal. The first probe also includes a first target-specific sequence attached to the first molecule. The second probe is made up, at least in part, of a second molecule, comprising a second target-specific sequence and an affinity tag. The first target-specific sequence and the second target-specific sequence bind to different regions of the

same target molecule. The target molecule is a fusion gene or mRNA transcribed from said fusion gene. The fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia. When said probe pair is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitute at least part of a code that identifies the target molecule. Also, when the fusion gene is detected the subject is diagnosed with leukemia, a specific subtype of leukemia or the genetic aberration underlying leukemia is determined.

[20] In certain embodiments, the fusion gene can include the fusion of at least a portion of a first gene and at least a portion of a second gene. The first target-specific sequence can hybridize to a portion of the first gene and the second target-specific sequence hybridizes to a portion of the second gene. The first target-specific sequence can also hybridize to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.

[21] In other embodiments, the first target-specific sequence can hybridize to a portion of the first gene and a portion of the second gene. The first target-specific sequence can also hybridize to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene. At least one of the target-specific sequences can include a region with one or more mismatches to the target of the target-specific sequence.

[22] In other embodiments, the second target-specific sequence can hybridize to a portion of the first gene and a portion of the second gene. The first target-specific sequence can also hybridize to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene. At least one of the target-specific sequences can include a region with one or more mismatches to the target of the target-specific sequence.

[23] The leukemia can be chronic myelogenous leukemia, acute myeloid leukemia or acute lymphoblastic leukemia. The fusion gene can be *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*.

[24] Each of said label monomers can be a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety. The unique label for each target-specific probe can be composed of 4 or 5 label monomers.

[25] The disclosure also provides a kit. In one embodiment, the kit includes one or more reporter probes and one or more capture probes. The reporter probe and the capture probe each

bind specifically to a fusion gene. The fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia.

[26] The fusion gene can be *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*. The kit can include one or more reporter probes and one or more capture probes from Table 2.

BRIEF DESCRIPTION OF THE DRAWINGS

[27] Figures 1A is a bar graph showing counts for probes for the *BCR-ABL* gene fusion when exposed to gene fragments of the *BCR-ABL* gene fusion and non-specific targets.

[28] Figure 1B is a bar graph showing counts for probes for the *E2A-PBX1* gene fusion when exposed to gene fragments of the *E2A-PBX1* gene fusion and non-specific targets.

[29] Figure 1C is a bar graph showing counts for probes for the *TEL-AML1* gene fusion when exposed to gene fragments of the *TEL-AML1* gene fusion and non-specific targets.

[30] Figure 1D is a bar graph showing counts for probes for the *PML-RARA* gene fusion when exposed to gene fragments of the *PML-RARA* gene fusion and non-specific targets.

[31] Figure 2A is a bar graph showing the relative expression of various gene fusions in the SUP B13 cell line.

[32] Figure 2B is a bar graph showing the relative expression of various gene fusions in the K562 cell line.

[33] Figure 2C is a bar graph showing the relative expression of various gene fusions in the NB4 cell line.

[34] Figure 2D is a bar graph showing the relative expression of various gene fusions in the Kasumi cell line.

[35] Figure 2E is a bar graph showing the relative expression of various gene fusions in the KCL-22 cell line.

[36] Figure 2F is a bar graph showing the relative expression of various gene fusions in the MV4-11 cell line.

DETAILED DESCRIPTION OF THE INVENTION

[37] This disclosure describes compositions and methods for the diagnosis of leukemia, diagnosis of the subtype of leukemia present in a subject and determination of the genetic

aberration underlying leukemia in a subject. The onset and progression of leukemia is associated with the formation of several fusion genes. These fusion genes include *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*. The compositions disclosed herein are nucleic acid probes used to detect the presence of these fusion genes.

[38] The detection of these fusion genes can be used to detect the presence of leukemia and also what type of leukemia is present. For example, a *BCR-ABL* fusion gene is associated with chronic myelogenous leukemia. *AML-ETO*, *CBFB-MYH11*, *MLL-AF4*, and *PML-RARA* fusion genes are associated with acute myeloid leukemia. *E2A-PBX1* and *TEL-AML1* fusion genes are associated with acute lymphoblastic leukemia. In addition, specific isoforms of a given fusion may indicate a specific sub-type of disease.

[39] The presence of certain gene fusions in certain types of leukemia is associated with differing prognoses in the outcome of the disease for a subject. Thus, the detection of various gene fusions can be used to provide a prognosis for a subject with leukemia.

[40] Preferably, the nucleic acid probes used according to the methods of the disclosure are nanoreporters. A fully assembled and labeled nanoreporter comprises two main portions, a target-specific sequence that is capable of binding to a target molecule, and a labeled region which emits a “code” of signals (the “nanoreporter code”) associated with the target-specific sequence.

[41] Upon binding of the nanoreporter to the target molecule, the nanoreporter code identifies the target molecule to which the nanoreporter is bound.

Nanoreporters

[42] Many nanoreporters, referred to as singular nanoreporters, are composed of one molecular entity. However, to increase the specificity of a nanoreporter and/or to improve the kinetics of its binding to a target molecule, a preferred nanoreporter is a dual nanoreporter composed of two molecular entities, each containing a different target-specific sequence that binds to a different region of the same target molecule. In a dual nanoreporter, at least one of the two nanoreporter probes is labeled. This labeled nanoreporter probe is referred to herein as a “reporter probe”. The other nanoreporter probe is not necessarily labeled. Such unlabeled components of dual nanoreporters are referred to herein as “capture probes” and often have affinity tags attached, such as biotin, which are useful to immobilize and/or stretch the complex containing the dual nanoreporter and the target molecule to allow visualization and/or imaging of

the complex. When both probes are labeled or both have affinity tags, the probe with more label monomer attachment regions is referred to as the reporter probe and the other probe in the pair is referred to as a capture probe.

[43] For both single and dual nanoreporters, a fully assembled and labeled nanoreporter probe comprises two main portions, a target-specific sequence that is capable of binding to a target molecule, and a labeled portion which provides a "code" of signals associated with the target-specific sequence. Upon binding of the nanoreporter probe to the target molecule, the code identifies the target molecule to which the nanoreporter is bound.

[44] Nanoreporters are modular structures. In some embodiments, the nanoreporter comprises a plurality of different detectable molecules. In some embodiments, a labeled nanoreporter, is a molecular entity containing certain basic elements: (i) a plurality of unique label attachment regions attached in a particular, unique linear combination, and (ii) complementary polynucleotide sequences attached to the label attachment regions of the backbone. In some embodiments, the labeled nanoreporter comprises 2, 3, 4, 5, 6, 7, 8, 9, 10 or more unique label attachment regions attached in a particular, unique linear combination, and complementary polynucleotide sequences attached to the label attachment regions of the backbone. In some embodiments, the labeled nanoreporter comprises 6 or more unique label attachment regions attached in a particular, unique linear combination, and complementary polynucleotide sequences attached to the label attachment regions of the backbone. A nanoreporter probe further comprises a target-specific sequence, also attached to the backbone.

[45] The term label attachment region includes a region of defined polynucleotide sequence within a given backbone that may serve as an individual attachment point for a detectable molecule. In some embodiments, the label attachment regions comprise designed sequences.

[46] In some embodiments, the label nanoreporter also comprises a backbone containing a constant region. The term constant region includes tandemly-repeated sequences of about 10 to about 25 nucleotides that are covalently attached to a nanoreporter. The constant region can be attached at either the 5' region or the 3' region of a nanoreporter, and may be utilized for capture and immobilization of a nanoreporter for imaging or detection, such as by attaching to a solid substrate a sequence that is complementary to the constant region. In certain aspects, the constant region contains 2, 3, 4, 5, 6, 7, 8, 9, 10, or more tandemly-repeated sequences, wherein the repeat sequences each comprise about 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,

26, 27, 28, 29, 30, or more nucleotides, including about 12-18, 13-17, or about 14-16 nucleotides.

[47] The nanoreporters described herein can comprise synthetic, designed sequences. In some embodiments, the sequences contain a fairly regularly-spaced pattern of a nucleotide (e.g. adenine) residue in the backbone. In some embodiments, a nucleotide is spaced at least an average of 8, 9, 10, 12, 15, 16, 20, 30, or 50 bases apart. In some embodiments, a nucleotide is spaced at least an average of 8 to 16 bases apart. In some embodiments, a nucleotide is spaced at least an average of 8 bases apart. This allows for a regularly spaced complementary nucleotide in the complementary polynucleotide sequence having attached thereto a detectable molecule. For example, in some embodiments, when the nanoreporter sequences contain a fairly regularly-spaced pattern of adenine residues in the backbone, whose complement is a regularly-spaced pattern of uridine (U) residues in complementary RNA segments, the in vitro transcription of the segments can be done using an aminoallyl-modified uridine base, which allows the covalent amine coupling of dye molecules at regular intervals along the segment. In some embodiments, the sequences contain about the same number or percentage of a nucleotide (e.g. adenine) that is spaced at least an average of 8, 9, 10, 12, 15, 16, 20, 30, or 50 bases apart in the sequences. This allows for similar number or percentages in the complementary polynucleotide sequence having attached thereto a detectable molecule. Thus, in some embodiments, the sequences contain a nucleotide that is not regularly-spaced but that is spaced at least an average of 8, 9, 10, 12, 15, 16, 20, 30, or 50 bases apart. In some embodiments, 20%, 30%, 50%, 60%, 70%, 80%, 90% or 100% of the complementary nucleotide is coupled to a detectable molecule. For instance, in some embodiments, when the nanoreporter sequences contain a similar percentage of adenine residues in the backbone and the in vitro transcription of the complementary segments is done using an aminoallyl-modified uridine base, 20%, 30%, 50%, 60%, 70%, 80%, 90% or 100% of the aminoallyl-modified uridine base can be coupled to a detectable molecule. Alternatively, the ratio of aminoallyl-modified uridine bases and uridine bases can be changed during the in vitro transcription process to achieve the desired number of sites which can be attached to a detectable molecule. For example, in vitro transcription process can take place in the presence of a mixture with a ratio of 1/1 of uridine to aminoallyl-modified uridine bases, when some or all the aminoallyl-modified uridine bases can be coupled to a detectable molecule.

[48] In some embodiments, the nanoreporters described herein have a fairly consistent melting

temperature (T_m). Without intending to be limited to any theory, the T_m of the nanoreporters described herein provides for strong bonds between the nanoreporter backbone and the complementary polynucleotide sequence having attached thereto a detectable molecule, therefore, preventing dissociation during synthesis and hybridization procedures. In addition, the consistent T_m among a population of nanoreporters allows for the synthesis and hybridization procedures to be tightly optimized, as the optimal conditions are the same for all spots and positions. In some embodiments, the sequences of the nanoreporters have a 50% guanine/cytosine (G/C), with no more than three G's in a row. Thus, in some embodiments, the disclosure provides a population of nanoreporters in which the T_m among the nanoreporters in the population is fairly consistent. In some embodiments, the disclosure provides a population of nanoreporters in which the T_m of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80° C, 85° C, 90° C, 100° C or higher. In some embodiments, the disclosure provides a population of nanoreporters in which the T_m of the complementary polynucleotide sequences when hybridized to its label attachment regions is about 80° C or higher.

[49] In some embodiments, the nanoreporters described herein have minimal or no secondary structures, such as any stable intra-molecular base-pairing interaction (e.g. hairpins). Without intending to be limited to any theory, the minimal secondary structure in the nanoreporters provides for better hybridization between the nanoreporter backbone and the polynucleotide sequence having attached thereto a detectable molecule. In addition, the minimal secondary structure in the nanoreporters provides for better detection of the detectable molecules in the nanoreporters. In some embodiments, the nanoreporters described herein have no significant intra-molecular pairing under annealing conditions of 75° C, 1 X SSPE. Secondary structures can be predicted by programs known in the art such as MFOLD. In some embodiments, the nanoreporters described herein contain less than 1% of inverted repeats in each strand, wherein the inverted repeats are 9 bases or greater. In some embodiments, the nanoreporters described herein contain no inverted repeats in each strand. In some embodiments, the nanoreporters do not contain any inverted repeat of 9 nucleotides or greater across a sequence that is 1100 base pairs in length. In some embodiments, the nanoreporters do not contain any inverted repeat of 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain less than 1% of inverted repeats in each strand, wherein the inverted

repeats are 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the nanoreporters described herein contain less than 1% of inverted repeats in each strand, wherein the inverted repeats are 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain a skewed strand specific content such that one strand is CT-rich and the other is GA-rich.

[50] The disclosure also provides unique nanoreporters. In some embodiments, the nanoreporters described herein contain less than 1% of direct repeats. In some embodiments, the nanoreporters described herein contain no direct repeats. In some embodiments, the nanoreporters do not contain any direct repeat of 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the labeled nanoreporters do not contain any direct repeat of 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain less than 1% of direct repeats in each strand, wherein the direct repeats are 9 nucleotides or greater across a sequence that 1100 base pairs in length. In some embodiments, the nanoreporters described herein contain less than 1% of direct repeats in each strand, wherein the direct repeats are 7 nucleotides or greater across any 100 base pair region. In some embodiments, the nanoreporters described herein contain less than 85, 80, 70, 60, 50, 40, 30, 20, 10, or 5% homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein contain less than 85% homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein contain less than 20, 16, 15, 10, 9, 7, 5, 3, 2 contiguous bases of homology to any other sequence used in the backbones or to any sequence described in the REFSEQ public database. In some embodiments, the nanoreporters described herein have no more than 15 contiguous bases of homology and no more than 85% identity across the entire length of the nanoreporter to any other sequence used in the backbones or to any sequence described in the REFSEQ public database.

[51] In some embodiments, the sequence characteristics of the nanoreporter probes described herein provide sensitive detection of a target molecule. For instance, the binding of the nanoreporter probes to target molecules which results in the identification of the target molecules can be performed by individually detecting the presence of the nanoreporter. This can be performed by individually counting the presence of one or more of the nanoreporter molecules in

a sample.

[52] The complementary polynucleotide sequences attached to a nanoreporter backbone serve to attach detectable molecules, or label monomers, to the nanoreporter backbone. The complementary polynucleotide sequences may be directly labeled, for example, by covalent incorporation of one or more detectable molecules into the complementary polynucleotide sequence. Alternatively, the complementary polynucleotide sequences may be indirectly labeled, such as by incorporation of biotin or other molecule capable of a specific ligand interaction into the complementary polynucleotide sequence. In such instances, the ligand (e.g., streptavidin in the case of biotin incorporation into the complementary polynucleotide sequence) may be covalently attached to the detectable molecule. Where the detectable molecules attached to a label attachment region are not directly incorporated into the complementary polynucleotide sequence, this sequence serves as a bridge between the detectable molecule and the label attachment region, and may be referred to as a bridging molecule, e.g., a bridging nucleic acid.

[53] The nucleic-acid based nanoreporter and nanoreporter-target complexes described herein comprise nucleic acids, which may be affinity-purified or immobilized using a nucleic acid, such as an oligonucleotide, that is complementary to the constant region or the nanoreporter or target nucleic acid. As noted above, in some embodiments the nanoreporters comprise at least one constant region, which may serve as an affinity tag for purification and/or for immobilization (for example to a solid surface). The constant region typically comprises two or more tandemly-repeated regions of repeat nucleotides, such as a series of 15-base repeats. In such exemplary embodiments, the nanoreporter, whether complexed to a target molecule or otherwise, can be purified or immobilized by an affinity reagent coated with a 15-base oligonucleotide which is the reverse complement of the repeat unit.

[54] Nanoreporters, or nanoreporter-target molecule complexes, can be purified in two or more affinity selection steps. For example, in a dual nanoreporter, one probe can comprise a first affinity tag and the other probe can comprise a second (different) affinity tag. The probes are mixed with target molecules, and complexes comprising the two probes of the dual nanoreporter are separated from unbound materials (e.g., the target or the individual probes of the nanoreporter) by affinity purification against one or both individual affinity tags. In the first step, the mixture can be bound to an affinity reagent for the first affinity tag, so that only probes comprising the first affinity tag and the desired complexes are purified. The bound materials are

released from the first affinity reagent and optionally bound to an affinity reagent for the second affinity tag, allowing the separation of complexes from probes comprising the first affinity tag. At this point only full complexes would be bound. The complexes are finally released from the affinity reagent for the second affinity tag and then preferably stretched and imaged. The affinity reagent can be any solid surface coated with a binding partner for the affinity tag, such as a column, bead (e.g., latex or magnetic bead) or slide coated with the binding partner.

Immobilizing and stretching nanoreporters using affinity reagents is fully described in U.S. Publication No. 2010/0161026, which is incorporated by reference herein in its entirety.

[55] The sequence of signals provided by the label monomers associated with the various label attachment regions of the backbone of a given nanoreporter allows for the unique identification of the nanoreporter. For example, when using fluorescent labels, a nanoreporter having a unique identity or unique spectral signature is associated with a target-specific sequence that recognizes a specific target molecule or a portion thereof. When a nanoreporter is exposed to a mixture containing the target molecule under conditions that permit binding of the target-specific sequence(s) of the nanoreporter to the target molecule, the target-specific sequence(s) preferentially bind(s) to the target molecule. Detection of the nanoreporter signal, such as the spectral code of a fluorescently labeled nanoreporter, associated with the nanoreporter allows detection of the presence of the target molecule in the mixture (qualitative analysis). Counting all the label monomers associated with a given spectral code or signature allows the counting of all the molecules in the mixture associated with the target-specific sequence coupled to the nanoreporter (quantitative analysis). Nanoreporters are thus useful for the diagnosis or prognosis of different biological states (e.g., disease vs. healthy) by quantitative analysis of known biological markers. Moreover, the exquisite sensitivity of single molecule detection and quantification provided by the nanoreporters described herein allows for the identification of new diagnostic and prognostic markers, including those whose fluctuations among the different biological states is too slight to detect a correlation with a particular biological state using traditional molecular methods. The sensitivity of nanoreporter-based molecular detection permits detailed pharmacokinetic analysis of therapeutic and diagnostic agents in small biological samples.

[56] Many nanoreporters, referred to as singular nanoreporters, are composed of one molecular entity. However, to increase the specificity of a nanoreporter and/or to improve the

kinetics of its binding to a target molecule, a nanoreporter can be a dual nanoreporter composed of two molecular entities, each containing a different target-specific sequence that binds to a different region of the same target molecule. In a dual nanoreporter, at least one of the two molecular entities is labeled. The other molecular entity need not necessarily be labeled. Such unlabeled components of dual nanoreporters may be used as capture probes and optionally have affinity tags attached, such as biotin, which are useful to immobilize and/or stretch the complex containing the dual nanoreporter and the target molecule to allow visualization and/or imaging of the complex. For instance, in some embodiments, a dual nanoreporter with a 6-position nanoreporter code uses one 6-position coded nanoreporter (also referred to herein as a reporter probe) and a capture probe. In some embodiments, a dual nanoreporter with a 6-position nanoreporter code can be used, using one capture probe with an affinity tag and one 6-position nanoreporter component. In some embodiments an affinity tag is optionally included and can be used to purify the nanoreporter or to immobilize the nanoreporter (or nanoreporter-target molecule complex) for the purpose of imaging.

[57] In some embodiments, the nucleotide sequences of the individual label attachment regions within each nanoreporter are different from the nucleotide sequences of the other label attachment regions within that nanoreporter, preventing rearrangements, such as recombination, sharing or swapping of the label polynucleotide sequences. The number of label attachment regions to be formed on a backbone is based on the length and nature of the backbone, the means of labeling the nanoreporter, as well as the type of label monomers providing a signal to be attached to the label attachment regions of the backbone. In some embodiments, the complementary nucleotide sequence of each label attachment region is assigned a specific detectable molecule.

[58] The disclosure also provides labeled nanoreporters wherein one or more label attachment regions are attached to a corresponding detectable molecule, each detectable molecule providing a signal. For example, in some embodiments, a labeled nanoreporter according to the disclosure is obtained when at least three detectable molecules are attached to three corresponding label attachment regions of the backbone such that these labeled label attachment regions, or spots, are distinguishable based on their unique linear arrangement. A "spot," in the context of nanoreporter detection, is the aggregate signal detected from the label monomers attached to a single label attachment site on a nanoreporter, and which, depending on the size of the label attachment

region and the nature (e.g., primary emission wavelength) of the label monomer, may appear as a single point source of light when visualized under a microscope. Spots from a nanoreporter may be overlapping or non-overlapping. The nanoreporter code that identifies that target molecule can comprise any permutation of the length of a spot, its position relative to other spots, and/or the nature (e.g., primary emission wavelength(s)) of its signal. Generally, for each probe or probe pair described herein, adjacent label attachment regions are non-overlapping, and/or the spots from adjacent label attachment regions are spatially and/or spectrally distinguishable, at least under the detection conditions (e.g., when the nanoreporter is immobilized, stretched and observed under a microscope, as described in U.S. Publication No. 2010/0112710, incorporated herein by reference).

[59] Occasionally, reference is made to a spot size as a certain number of bases or nucleotides. As would be readily understood by one of skill in the art, this refers to the number of bases or nucleotides in the corresponding label attachment region.

[60] The order and nature (e.g., primary emission wavelength(s), optionally also length) of spots from a nanoreporter serve as a nanoreporter code that identifies the target molecule capable of being bound by the nanoreporter through the nanoreporter's target specific sequence(s). When the nanoreporter is bound to a target molecule, the nanoreporter code also identifies the target molecule. Optionally, the length of a spot can be a component of the nanoreporter code.

[61] Detectable molecules providing a signal associated with different label attachment regions of the backbone can provide signals that are indistinguishable under the detection conditions ("like" signals), or can provide signals that are distinguishable, at least under the detection conditions (e.g., when the nanoreporter is immobilized, stretched and observed under a microscope).

[62] The disclosure also provides a nanoreporter wherein two or more detectable molecules are attached to a label attachment region. The signal provided by the detectable molecules associated with said label attachment region produces an aggregate signal that is detected. The aggregate signal produced may be made up of like signals or made up of at least two distinguishable signals (e.g., spectrally distinguishable signals).

[63] In one embodiment, a nanoreporter includes at least three detectable molecules providing like signals attached to three corresponding label attachment regions of the backbone and said three detectable molecules are spatially distinguishable. In another embodiment, a nanoreporter

includes at least three detectable molecules providing three distinguishable signals attached to three neighboring label attachment regions, for example three adjacent label attachment regions, whereby said at least three label monomers are spectrally distinguishable.

[64] In other embodiments, a nanoreporter includes spots providing like or unlike signals separated by a spacer region, whereby interposing the spacer region allows the generation of dark spots, which expand the possible combination of uniquely detectable signals. The term "dark spot" refers to a lack of signal from a label attachment site on a nanoreporter. Dark spots can be incorporated into the nanoreporter code to add more coding permutations and generate greater nanoreporter diversity in a nanoreporter population. In one embodiment, the spacer regions have a length determined by the resolution of an instrument employed in detecting the nanoreporter.

[65] In other embodiments, a nanoreporter includes one or more "double spots." Each double spot contains two or more (e.g., three, four or five) adjacent spots that provide like signals without being separated by a spacer region. Double spots can be identified by their sizes.

[66] A detectable molecule providing a signal described herein may be attached covalently or non-covalently (e.g., via hybridization) to a complementary polynucleotide sequence that is attached to the label attachment region. The label monomers may also be attached indirectly to the complementary polynucleotide sequence, such as by being covalently attached to a ligand molecule (e.g., streptavidin) that is attached through its interaction with a molecule incorporated into the complementary polynucleotide sequence (e.g., biotin incorporated into the complementary polynucleotide sequence), which is in turn attached via hybridization to the backbone.

[67] A nanoreporter can also be associated with a uniquely detectable signal, such as a spectral code, determined by the sequence of signals provided by the label monomers attached (e.g., indirectly) to label attachment regions on the backbone of the nanoreporter, whereby detection of the signal allows identification of the nanoreporter.

[68] In other embodiments, a nanoreporter also includes an affinity tag attached to the reporter probe backbone, such that attachment of the affinity tag to a support allows backbone stretching and resolution of signals provided by label monomers corresponding to different label attachment regions on the backbone. Nanoreporter stretching may involve any stretching means known in the art including but not limited to, means involving physical, hydrodynamic or

electrical means. The affinity tag may comprise a constant region.

[69] In other embodiments, a nanoreporter also includes a target-specific sequence coupled to the backbone. The target-specific sequence is selected to allow the nanoreporter to recognize, bind or attach to a target molecule. The nanoreporters described herein are suitable for identification of target molecules of all types. For example, appropriate target-specific sequences can be coupled to the backbone of the nanoreporter to allow detection of a target molecule. Preferably the target molecule is DNA (including cDNA), RNA (including mRNA and cRNA), a peptide, a polypeptide, or a protein.

[70] One embodiment of the disclosure provides increased flexibility in target molecule detection with label monomers described herein. In this embodiment, a dual nanoreporter comprising two different molecular entities, each with a separate target-specific region, at least one of which is labeled, bind to the same target molecule. Thus, the target-specific sequences of the two components of the dual nanoreporter bind to different portions of a selected target molecule, whereby detection of the spectral code associated with the dual nanoreporter provides detection of the selected target molecule in a biomolecular sample contacted with said dual nanoreporter.

[71] The disclosure also provides a method of detecting the presence of a specific target molecule in a biomolecular sample comprising: (i) contacting said sample with a nanoreporter as described herein (e.g., a singular or dual nanoreporter) under conditions that allow binding of the target-specific sequences in the dual nanoreporter to the target molecule and (ii) detecting the spectral code associated with the dual nanoreporter. Depending on the nanoreporter architecture, the dual nanoreporter may be labeled before or after binding to the target molecule.

[72] The uniqueness of each nanoreporter probe in a population of probe allows for the multiplexed analysis of a plurality of target molecules. For example, in some embodiments, each nanoreporter probe contains six label attachment regions, where each label attachment region of each backbone is different from the other label attachment regions in that same backbone. If the label attachment regions are going to be labeled with one of four colors and there are 24 possible unique sequences for the label attachment regions and each label attachment region is assigned a specific color, each label attachment region in each backbone will consist of one of four sequences. There will be 4096 possible nanoreporters in this example. The number of possible nanoreporters can be increased, for example, by increasing the number of colors, increasing the

number of unique sequences for the label attachment regions and/or increasing the number of label attachment regions per backbone. Likewise the number of possible nanoreporters can be decreased by decreasing the number of colors, decreasing the number of unique sequences for the label attachment regions and/or decreasing the number of label attachment regions per backbone.

[73] In certain embodiments, the methods of detection are performed in multiplex assays, whereby a plurality of target molecules are detected in the same assay (a single reaction mixture). In a preferred embodiment, the assay is a hybridization assay in which the plurality of target molecules are detected simultaneously. In certain embodiments, the plurality of target molecules detected in the same assay is, at least 2, at least 5 different target molecules, at least 10 different target molecules, at least 20 different target molecules, at least 50 different target molecules, at least 75 different target molecules, at least 100 different target molecules, at least 200 different target molecules, at least 500 different target molecules, or at least 750 different target molecules, or at least 1000 different target molecules. In other embodiments, the plurality of target molecules detected in the same assay is up to 50 different target molecules, up to 100 different target molecules, up to 150 different target molecules, up to 200 different target molecules, up to 300 different target molecules, up to 500 different target molecules, up to 750 different target molecules, up to 1000 different target molecules, up to 2000 different target molecules, or up to 5000 different target molecules. In yet other embodiments, the plurality of target molecules detected is any range in between the foregoing numbers of different target molecules, such as, but not limited to, from 20 to 50 different target molecules, from 50 to 200 different target molecules, from 100 to 1000 different target molecules, from 500 to 5000 different target molecules, and so on and so forth.

[74] In certain embodiments, nanoreporters for detecting the presence of fusion genes contain both a reporter probe and a capture probe that hybridize to the fusion gene at different positions. In one embodiment, the target-specific regions of the capture and reporter probe hybridize to sequences on either side of the border of the fusion between the two fused genes. Optionally, the target-specific regions hybridize to regions adjacent to each other on either side of the border of fusion. The sequences to which the target-specific regions of the reporter and capture probes hybridize can also be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more base pairs apart.

[75] In another embodiment, the reporter probe's target-specific region hybridizes to a sequence that spans the fusion junction while the target-specific region of the capture probe hybridizes to a sequence upstream or downstream of the reporter probe's target. In another embodiment, the capture probe's target-specific region hybridizes to a sequence that spans the fusion junction while the target-specific region of the reporter probe hybridizes to a sequence upstream or downstream of the reporter probe's target. Optionally, the target sequences of the reporter and capture probes are adjacent. The sequences to which the target-specific regions of the reporter and capture probes hybridize can also be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more base pairs apart. In certain embodiments, the target-specific region of a probe has 1, 2, 3, 4, 5 or more mismatches introduced into the target-specific sequence. The sequences which come together to form the junction of a fusion gene can be GC rich, resulting in a high melting temperature for the probe on both sides of the junction. Thus, in certain situations, the probe has an increased propensity to cross-hybridize with isoforms containing the same sequence as only one side of the junction, rather than requiring the presence of both sequences for successful hybridization. Introduction of certain mismatches into the probe produces a decreased melting temperature and higher specificity. In preferred embodiments, probes that hybridize to a sequence that encompasses the junction of the fusion genes have 1, 2, 3, 4, 5, or more mismatches. Additional disclosure regarding nanoreporters can be found in International Publication Nos. WO 07/076129 and WO 07/076132, the contents of which are incorporated herein in their entireties. Further, the term nucleic acid probes and nanoreporters can include the rationally designed (*e.g.* synthetic sequences) described in International Publication No. WO 2010/019826, incorporated herein by reference in its entirety.

[76] Examples of various probes and their targets are provided below in Table 1.

Table 1

Gene Fusion	Isoform	Reporter Probe	SEQ ID NO:
BCR-ABL	b2a2	TGGCCGCTGAAGGGCTT CTCCTTATTGATGGTCA	1
BCR-ABL	b3a2	TGGCCGCTGAAGGGCTT TTGAACCTCTGCTTAAATC	2
BCR-ABL	e1a2	TGGCCGCTGAAGGGCTT CTGCATCTCCATGGAAGG (*)	3
BCR-ABL	e19a2	TGGCCGCTGAAGGGCTT TGACGTCGAAAGCTGCCT (*)	4
BCR-ABL	b3a3	GACCCGGAGCTTTTCAC TTGAACCTCTGCTTAAATC	5
BCR-ABL	b2a3	GACCCGGAGCTTTTCAC CTCCTTATTGATGGTCA	6
BCR-ABL	e1a3	GACCCGGAGCTTTTCAC CTGCATCTCCATGGAAGG (*)	7
AML-ETO		CTTCACATCCACAGGTGAGTCTGGCATTGTGGAGTGCTTCTCAGTACGAT (+)	8
CBFb-MYH11	e5e12	TCCAGGGCCCGCTTGGACTTCTCCAGCTCATGGAC (+)	9
CBFb-MYH11	e5e8	CGGCCTCGTTAAGCATCCCTGTGACGCTCTCAACTTCATT (+)	10
CBFb-MYH11	e5e7	AGCGTCTGCTTATTCTTGTCTAGGTTGCGCTTGGC (+)	11
E2A-PBX1	e13-insertion-n27-e2	AAAAGGTTCCAGGTGACCGAACACTTTCAT CACTG	12
E2A-PBX1	e13e2	TCCTCGGATACTCAAAA CACTGTAGGAGTCGGGAG	13
MLL-AF4	e8e7	AGATGTATCATATTGTT CTGATTCTGGTGGTGGAG	14
MLL-AF4	e9e5	CATGAATGGGTCATTC CTTTTCTTTTGGTTTTTG	15
MLL-AF4	e9e4	TCATTGGAGTAGGTCTG CTTTTCTTTTGGTTTTTG	16
MLL-AF4	e10e5	CATGAATGGGTCATTC CTAAAGTCCACTCTGAT	17
MLL-AF4	e10e4	TCATTGGAGTAGGTCTG CTAAAGTCCACTCTGAT	18
MLL-AF4	e11e5	CATGAATGGGTCATTC CTATACATGCCCACTACT (*)	19
MLL-AF4	e11e4	TCATTGGAGTAGGTCTG CTATACATGCCCACTACT (*)	20
PML-RARA	bcr1 (e6e3)	TCTGGGTCTCAATGG CTGACTCCCAAGCGCCACTG (*)	21
PML-RARA	bcr3 (e3e3)	TCTGGGTCTCAATGG CTTTCCACTGGGTGATGCAA (*)	22
PML-RARA	bcr2 (e6partial_e3)	GCTCTGGGTCTCAATGG TTCCTATGACGGGACTCC (*)	23
TEL-AML1	e5e2	TCCCCTAGGGCCACCAC CTGCTATTCTCAATGG (*)	24
TEL-AML1	e5e3	CAGAGTGAAGCTTTTCC CTGCTATTCTACCAATGG (*)	25
Gene Fusion	Isoform	Capture Probe	SEQ ID NO:
BCR-ABL	b2a2	GCGGAATGCTGTGGACAGTCTGGAGTTTCACACACGAGTT	26
BCR-ABL	b3a2	CAGTGGCTGAGTGGACGATGACATTCAGAAACCCATAGAGCCCCGGAGAC	27
BCR-ABL	e1a2	CGCCCTCGCCATCGTTGGGCCAGATCTGCCCGGTCTTGCG	28
BCR-ABL	e19a2	TCAGTGCCTGGATGTCCGTGGCCACACCGGACACG	29
BCR-ABL	b3a3	CAGTGGCTGAGTGGACGATGACATTCAGAAACCCA	30
BCR-ABL	b2a3	GCGGAATGCTGTGGACAGTCTGGAGTTTCACACAC	31
BCR-ABL	e1a3	CGCCCTCGCCATCGTTGGGCCAGATCTGCCCGGTCTTGCG	32
AML-ETO		TTCGAGGTTCTCGGGGCCATCCACTGTGATTTTGATGGCTCTGTGGTAG	33
CBFb-MYH11	e5e12	CAGTTACTGCCAGCAGCTGTGAAACTCTCACCTCCATTTT	34
CBFb-MYH11	e5e8	CAGTTACTGCCAGCAGCTGTGAAACTCTCACCTCCATTTT	35

CBFb-MYH11	e5e7	CAGTTACTGCCAGCAGCTGTGAAACTCTCACCTCCATTTC	36
E2A-PBX1	e13-insertion-n27-e2	TAGGAGTCGGGAGGCCGAGACAGGTCAGGGAGGGT	37
E2A-PBX1	e13e2	GCCGAGACAGGTCAGGGAGGGTGCCTGGCTGGCTG	38
MLL-AF4	e8e7	GCTGCTTTTTCTTGGGCTCACTAGGAGTGGTTTTG	39
MLL-AF4	e9e5	TTTACAGGGATACTTGGGCGGGGAGCCACTTTTT	40
MLL-AF4	e9e4	TTTACAGGGATACTTGGGCGGGGAGCCACTTTTT	41
MLL-AF4	e10e5	CCTGTGGACTCCATCTGCTGGAATTTTTTGCTTAG	42
MLL-AF4	e10e4	CCTGTGGACTCCATCTGCTGGAATTTTTTGCTTAG	43
MLL-AF4	e11e5	GGCACAGAGAAAGCAAACCACCCTGGGTGTTATAG	44
MLL-AF4	e11e4	GGCACAGAGAAAGCAAACCACCCTGGGTGTTATAG	45
PML-RARA	bcr1 (e6e3)	GCCACGTGGTTGCTGTTGGGCAGGAAGACCTCACTTCCTAT	46
PML-RARA	bcr3 (e3e3)	GAGCTGAGGTCCTGCAGGCGCACCTTGAACTCGTCGAAGC	47
PML-RARA	bcr2 (e6partial_e3)	TGGGGCTAGGCGGTCCATCCAGGTGGGGTGGTGAG	48
TEL-AML1	e5e2	GCATGGCGTGCTCTTCAGGCGGGGAGACAGAGACC	49
TEL-AML1	e5e3	GCATGGCGTGCTCTTCAGGCGGGGAGACAGAGACC	50

[77] Probes in Table 1 marked with (*) are mismatch junction probes that hybridize to the sequence across the gene fusion junction. Each has one, two or three mismatches shown underlined. These mismatches stop the probe from inappropriately hybridizing to a transcript which has only half of the relevant target sequence (for example, the wild-type transcript of one of the fusion genes), by lowering the amount of partial sequence complementarity so that stable hybridization can only occur with the true target. Such mismatches are only necessary in cases where the T_m of the sequence of one exon at the junction is high enough to be stable under standard hybridization conditions, and the sequence cannot simply be shortened because the shorter probe will not retain enough sequence specificity.

[78] Probes in Table 1 marked with (+) are border probes. These probes do not hybridize to a sequence that crosses the gene fusion junction. The target sequence of these reporter probes is adjacent to the target sequence of the capture probes. Each probe only hybridizes to a sequence on one side of the gene fusion junction. The probe pair (capture and reporter) meets at the border of the fusion, with the capture probe sitting in the upstream exon and the reporter probe sitting fully in the downstream exon.

[79] The unmarked sequences in Table 1 are probes with target sequences that cross the gene fusion junction, but do not have mismatches. The sequences of this class of probes are the exact reverse complement of the target sequence. Gaps in the sequence represent the fusion border;

the probe sequence is actually contiguous across this gap. In these probe pairs, the reporter probe spans the fusion junction, and the capture probe sits adjacent in the upstream exon.

[80] The disclosure also describes a panel of probes. These probes include probes for the detection of fusion genes from several gene families from various leukemia subtypes. The panel can contain probes for detection of 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more fusion genes. The genes can be from 1, 2, 3, 4, 5, 6, 7 or more gene families. The fusion genes can be present in any leukemia subtype including CML, AML or ALL.

[81] The fusion events, or translocations, usually involve only one of the pair of each chromosome, leaving one intact copy, or “wild-type” copy, of each gene also present in the cell. The panel of probes can also include probes specific for mRNAs produced by the wild-type copies of the relevant genes, allowing the level of expression of the intact version of each gene involved in the translocation to be determined. The panel can include probes specific for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more wild-type mRNAs.

[82] The panel of probes can also include probes specific for mRNAs that are clinically relevant for the detection and diagnosis of leukemia. These markers for leukemia can include WT1, FLT3, BAALC, PRAME, MLLT11, ERG, MN1, RHAMM, HOX11L2, EVI1, N-RAS, RB1 and TP53.

[83] The panel of probes can also include probes specific for housekeeper mRNAs. The results produced from these probes can be used as an internal control to estimate the amount of sample present in the assay. Probes for 1, 2, 3, 4 or more housekeeper mRNAs can be included in the panel.

[84] An example of a panel of probes for the detection and/or diagnosis of leukemia or specific subtypes of leukemia is shown below in Table 2.

Table 2.

Gene Type/Fusion Gene Family	Fusion Gene Isoform	Target Sequence (5'-3')	SEQ ID NO:	Reporter Probe (5'-3')	SEQ ID NO:	Capture Probe (5'-3')	SEQ ID NO:
Housekeeping Genes	GUSB	UCUGUGGCCAACAGAGCCUGCGUCCACCACCUAGAAU CUGCUGGCUACUACUUGAAGAUUGGUAUCGCUCA CACCAAUCCUUGGA (SEQ ID NO:)	51	TCCAAGGATTGGT GTGAGCGATCACCA TCTTCAAGTAGTAG CCAGC	52	AGATTCTAGGTGGG ACGCAGGCTCGTTG GCCACAGA	53
		CGCCGGCUGUUUAAUUCGCUUCCGCGUGGCCAU AGUGAUCUUUGCAGUGACCCAGCAGCAUCACUGU UUCUUGGCGUGUGAAGAUAAACCAAGGAUUUG	54	CAATTCCTTGGGT ATCTTCACACGCCA AGAAACAGTGATGC TGCTGGGT	55	CACTGCAAAAGATCA CTATGGGCCAGCGG AAGCGAAGTTAAA CAGCCGGCG	56
	TBP	CUGUGAUUGACUUCUAAUUGCUGACUUGUGGAGAU GCAGCGAAUGUGAAAUCCACGUAUUGCCAUUU CCUUCUACGUCGCGUGACCGUUCUGGGAAGAU	57	GATCTCCAGAACG GTCAGCGAGCGTAG AGGAAATGGCATA TACGTGGG	58	ATTTCACATTGCT GCATCTCCACAAGT CAGCAATTGAAGTC AATCACAG	59
	ABL1	CUGGGAGACAGGGUGAAAGGGAGUGGUUUUUUAUGA ACUUAACUAGAGUCUAAAAGAUUUUACUGGAU CACUUGUCAAGAUUGCGCCUUCUGGGGAGAA	60	TTCTCCCAGAGAG GGCGCATCTTGACA AGTGATCCAGTAGA AATCTTTT	61	AGACTCTAAGTTAA GTTCATAAAAACCA CTCCCTTCACCCCTG TCTCCAG	62
	BCR	GGAACGGCACUGUUCACCUUUUUCUGACCAAAACC GCUCUACACGUCAGCACCAUCUCUGCAGCAUCUC UGUAGGCUACCAUUAACAUAUGUACCGGUGC	63	GCACCGGTACATTT GTTAATGTTGAGCC TACAGAGATGCTGC AGAGATGG	64	TGCTGACGTGTAGA GCGGTTTGGTCAGA TAAAGGTGAACAG TGCCGTTC	65
	SOCS2	UCAGAGAGCAAGGCAUCCGGGGGUGAAUUCUUGUCU AACAUUCCCGAGGUCAGCCAGGCGUCUAACCUUG AAAGCAGGAUGUAGUUCUGCCAGGCAACUUUU	66	AAAAGTTGCCCTGGC AGAACTACATCCTG CTTCCAGGTTAGC AGCCTGGC	67	TGACCTCGGGAATG TTAGACAAGATTCA CCCCCGATGCCTTG CTCTCTGA	68
	WT1	AAUUGGAAAACCAGGACCGCCUUGGUCUGCAUAUC UGAGAGCGUUCCAGAGCCGUAUCGUGGAAUUGGUG CUUUGCGAUUCACAGGGGGAAGCUGUAAAGA	69	TCCTTACAGCTTTC CCCTGTGAATCGCA AAGCACCCATTCCA CGATCGG	70	CTCTGGAACGCTCT CAGATATGCAGACC AGGGCGTCTGTTT TTCCATT	71
	FLT3	AAGAAUCACAAAAGAACUGUGUCAAACUAGCAGAGA GUCCAAAGCAGAAAGGCGAGAUUGACUUUCUUCAGUG UCCUUCACGGCACUGGAUCCCAUCAAGAAACC	72	GGTCTTTGATGGG ATCCAGTGCCGTGA AGGACACTGAAGAA GTCCATCT	73	GCCCTTCTGCTGG ACTCTCTGCTAGTT GACACAGTCTTTG TGATTCTT	74
	BAALC	GGUCUUUGAUGAGUGUGGGAUACCGGAUGAUCAG CUCCUUGCCCUCCUGCCUUCUCCUGAGCCACUGCUC	75	TTAAGGTTGTAAGC TGGGAGCAGTGGCT	76	GGCAAGGAGCTGA TCATCCGTGATCCC	77
	PRAME						
Leukemia Biomarkers	WT1	UCAGAGAGCAAGGCAUCCGGGGGUGAAUUCUUGUCU AACAUUCCCGAGGUCAGCCAGGCGUCUAACCUUG AAAGCAGGAUGUAGUUCUGCCAGGCAACUUUU	66	AAAAGTTGCCCTGGC AGAACTACATCCTG CTTCCAGGTTAGC AGCCTGGC	67	TGACCTCGGGAATG TTAGACAAGATTCA CCCCCGATGCCTTG CTCTCTGA	68
		AAUUGGAAAACCAGGACCGCCUUGGUCUGCAUAUC UGAGAGCGUUCCAGAGCCGUAUCGUGGAAUUGGUG CUUUGCGAUUCACAGGGGGAAGCUGUAAAGA	69	TCCTTACAGCTTTC CCCTGTGAATCGCA AAGCACCCATTCCA CGATCGG	70	CTCTGGAACGCTCT CAGATATGCAGACC AGGGCGTCTGTTT TTCCATT	71
	FLT3	AAGAAUCACAAAAGAACUGUGUCAAACUAGCAGAGA GUCCAAAGCAGAAAGGCGAGAUUGACUUUCUUCAGUG UCCUUCACGGCACUGGAUCCCAUCAAGAAACC	72	GGTCTTTGATGGG ATCCAGTGCCGTGA AGGACACTGAAGAA GTCCATCT	73	GCCCTTCTGCTGG ACTCTCTGCTAGTT GACACAGTCTTTG TGATTCTT	74
	BAALC	GGUCUUUGAUGAGUGUGGGAUACCGGAUGAUCAG CUCCUUGCCCUCCUGCCUUCUCCUGAGCCACUGCUC	75	TTAAGGTTGTAAGC TGGGAGCAGTGGCT	76	GGCAAGGAGCTGA TCATCCGTGATCCC	77
	PRAME						

	CCAGCUUACAACCUUAA		CAGGGAAGGCAGG AG		ACACTCATCAAAGA CC	
MLLT1 1	GGAAUAGGACUAGGUUUAUUUACCCAUUUGAGG GUAGAGAGGCGAGUCUGGAGGAGCAGGGAUUGGG AGAAAGGGUGGAAAAUACUCUGAUUUCUAAA	78	TTTAAGAAATCAGAG TATTTTCCACCCCT TCTCCAATCCCTG CTCCTCC	79	AGACTCGCCTCTCT ACCCTCACAATGGG TAAATAAACCTAGT CCTATTCC	80
ERG	GACCUCAUCAUUAUUGUGGGGCUUUGUUCUCCAC AGGUUCAGGUUAGAGAUUGGCCUUCUUGGUGCCA CAUUCAGAAUUCACGCGAGCAUUGUGGUAGG	81	CCTACCCAAATGC CTGCGTGAATTTCTG ATTGTGCGAGCCAA GAAGGCCA	82	TCCTTAACCTGACC CTGTGGAGAAACA AGCCCCACATAAT GATGAGGTC	83
MN1	AGGUGACGAACCAAGGAGCCGUCGACUCGCGUGG AAUACAUAUACCCGGCGAGCGCCCTCCGGGACA UUUUGACAUGUUUUCGCCCCU	84	AGGCGAAAAACATG TCAAAAATGTCCCGA GGCGCCTCGCCCG G	85	GTAATTGTATTCCA GCGAGTCGACGGCT CCTTGTTTCGTAC CCT	86
RHAM M	UUUAAACCGGUAGGGAGUGAUAAUCCGCAUUCAG UUUCGAGGAGUGCCAGUCACCUUCAGUUUCUGG AGCUGGCCGUAACAUGUCCUUCUUAAGGCG	87	CGCCTTAGGAAAGG ACATGTGACGGCC AGCTCCAGAAACTG AAGGTGAC	88	TGGCACTCCTCGAC AACTGAATGCGGAT TATCACTCCCTACC GGTTCAA	89
HOX11	ACCACAUCCAGCCCAUCCAGGUACGCACAGA CAGGUUUUCACAUAAUAGCAGCCCAUUCUCCAG AACCCAUUUGAGGGUGGGGGUGUUAAUU	90	AATTAACACCCCCC CACCCCTCAAAATGG GTTCTGGAGAAATG GGCTGCAT	91	TTATGTGAAAAACCT GTCTGTGCGTACCT GGATTGGGCTGGG ATGTGTGGT	92
HOX11 L2	GGGGGCGCGCCAAAGUCCCUCAAAUAGACGGACG CGCAGGUCAAAGACCUUGGUUCCAAAACCGGAGGAC CAAGUGGCGG	93	CCGCCACTTGGTCC TCCGGTTTGGAAAC CAGGTCTTGACC	94	TGCGCGTCCGTCTAT TTTGAGGGGACTTGG CGAGCGCCGC	95
EV11	GGGCAGGACUAGGAUUAUGGACCAAAAGGAAGA UCGAAUAGGUGAAAAAGUUUGGGCCUUAUGUGGG AGAGCAGAGGUCAAAACCUAGAAAGACCCCAU	96	AACTGGGGTCTTTC AGTTTGAACCTCTG CTCTCCACATAAG GCCCCAAC	97	TTTTCACTACTTC GATCTTCCTTTTGG TCCATATTCTCTAGT CCTGCCCC	98
N-RAS	CCUUGUCCUGACUCCUCCUGGAGGAGAAUUAUUC CUGUUGCUGUCUCAGUCUCACAGAGAAAGCUCCU GCUACUCCCCAGCUCUCAGUAGUUUAGUAC	99	GTAATAAACTACTG AGAGCTGGGAAAGT AGCAGGAGCTTCTC TGTGAGAC	100	TGAAGACAGCAAC AGGAATACTCTTCC TCCAGGGAAGTCA GGACCAGG	101
RB1	CCUUCUCCGGGUAAAUACACUUAUUGUAAACGCCU UCUGUCUGAGCACCCAGAAUUAAGAACAUUAUCAUC UGGACCCUUUCCAGCACACCCUAGCAGAAUGA	102	TCATTCTGACGGGT GTGCTGGAAAAAGG TCCAGATGATATGT TCTAATTC	103	TGGTGTCTCAGACA GAAGCGTTTACA AAGTGATTAGCC GGAGATAGG	104
TP53	CAGGGCUCACUCCAGCCACCUAGAGUCCAAAAAG GGUCAGUCUACCUCCCGCCAUAAAAACUCAUGU UCAAGACAGAGGGGCCUCAGAC	105	GTCTGAGTCAGGCC CTTCTGTCTTGAAC ATGAGTTTTTATG	106	TAGACTGACCCCTT TTGGACTTCAGGTG GCTGGAGTGAGCCC	107

				GC GGGAGG	TG	
Wild Type Transcripts	PML	108	GAGUGGGGCGUCAGCAGCUUCCUCAGCUUUCUGA GCUCCAUUGCGCGCCCUAUUCUUGGCCUGCUACAA GCUUG	CACAGCTTGTAGCA GGCCAAGATAGGGC GGCGCATGGAG	109	CTCAGAAAGCTGA GGAAGTGTGCAG CCCCACCTC
	RARA	111	CUGGGGCCCAUCUAGGAGUGGCAUCUUUUUUGGU GCCCCGAAAGGCCAGCUCUGGACCUUCCAGGAAA AGUGCCAG	CTGGCACTTTTCTCTG GGAAGGTCCAGAGC TGGCCTT	112	CAGGGCACCAAAA AAGATGCCACTCCT AGATGGGCCCCAG
	ETO (RUNX 1)	114	CAGCCAUGAAGAACAGGUUGCAAGAUUUAAUGA CCUCAGGUUUGCGUCGAAAGUGGAAGAGGGAAA AGCUACUCUGACCAUCACUGUCUUA	TGAAGACAGTGATG GTCAGAGTGAAGCT TTTCCCTCTTCCACT TCG	115	ACCGACAAACCTG AGGTCAATTAATCT TGCAACCTGGTTCT TCATGGCTG
	ETV6 (TEL)	117	CGCCACUACUACAAACUAAACAUUAUCAGGAAAG AGCCAGGACAAAGCGCUUUUGUUCAGGUUUUAUGAA AACCCAGAUAGAAUCAUGAGUG	CACTCATGATTTCAT TCTGGGGTTTTCAT AAACCTGAACAAAA GCCTTTGT	118	CTGGCTCCTTCCT GATAATGTTTAGTT TGTAAGTAGGGC
	TCF3 (E2A)	120	CUUGGAGCAGCAAGUGCGGAGAGCGGAACCUUGAAU CCCAAAGCAGCCUGUUUGAAACGGCGGAGAAAGAG AAAAG	CTTTTCTCTTCTCG CCGTTTCAACAGG CTGCTTTGG	121	GATTCAGGTTCCGC TCTCGCACTTGCTG CTCCAAG
	PBX1	123	CGGGAGGAAGCAGGACAUUGGAGACAUUUUACAG CAAAUUAUGACCAUCACAGACAGAGUUUGGAUG AGGCGCAGGCCAGAAACAUGCUUUAACUGC	GCAGTTTAAAGCAT GTTTCTGGCTGCG GCCTCATCCAAACT CTGGTCTG	124	TGATGGTCATAATT TGCTGTAAAATGTC TCCAATGTCTGCT TCCTCCCC
	CBFB	126	CAGUUGCCCCAGGCGUGUCUGCAACUCCUGGCA UCAAGCGAUCCUCCUGCCUUAAGCCUCCAGAGUA CUG	CAGTACTCTGGAG GCTAAGCGAGGAGG ATCGCTTG	127	ATGCCAGGAGTTCCG AGACCAGCCTGGG CAACACTG
	MYH11	129	CUGCUAGAAAAAUACGGGCAAUUCGCCAAGCCA GAGACGAGAGGACAUUCCACAUUUUUACUACAU GAUUGCUGAGCCAGGAGAGAGAGAGAGAGUG	CACTTCTCATCTTCT CCTTGGCTCCAGCA ATCATGTAGTAAAA GATGTGG	130	AATGTCTCTCTCGTC TCTGGCTTGGCGAA TTGCCCGTGATTTT TCTAGCAG
	MLL	132	CAAAUAGUGGACACAUUAUUUGUCUUUUGU GGUCAGCGGAAGUGUUUAGAGAUAGACGGAU ACUAAAGAAUUGUCAUUAUGGCUUGAUCA	CTGATCACAGCCAT ATGCACATCTTTA GTGATCCGTCATCA TCTTCA	133	AACACTTCCGCTGA CCACAAAGCACAA TTTACATGTGTCCA CTCATTTTG
	AFB1 (AF4)	135	UGAACUGAAACCAACUGCCGGAGGACUUAUCGACAG CAGACCUUUUAGAAAAACAGACUUGAAAGUGCCUG CCAAAGCCAAAGCUCACCAACAAACUGAAGAGGCCU	AGGCATCTTCAGTT TGCTGAGCTTGGCT TTGGCAGGCACTTT CAAGTCTG	136	TTTTTCAAAGGTC TGCTGTGATAGTC CTCCGGCAGTGGTT TCAGTTCA
	b2a2	138	AACUCGUGUGAACAACUCCAGACUUGCCACAGCA UUCCGUGACCAUCAAAUAGGAAGAAGCCCUCA	TGGCCGCTGAAGGG CTTCTTCTTATTGA	139	GCGGAATGCTGTGG ACAGTCTGGAGTTT
BCR-ABL						

		GCGGCCA		TGGTCA		CACACACGAGTT	
			141		142		143
	<i>b3a2</i>	GUCUCCGGGGCUCUAUGGGUUUCUGAAUGUCAUC GUCCACUCAGCCACUCUGGAUUUAGCAGAGUUCAA AAGCCUUCAGCGCCA		TGGCCGCTGAAGGG CTTTGAACTCTGCT TAAATC		CAGTGGCTGAGTGG ACGATGACATTTCAG AAACCCATAGAGC CCCGGAGAC	
	<i>e1a2</i>	CGCAAGACCGGGCAGAUUCUGGCCCAACGAUGGGC AGGCGCCUUCUCCAUUGGAGACGCAGAACGCCUUCAGCGGC GCGGCCA	144	TGGCCGCTGAAGGG CTTCTGCACTCTCCAT GGAAGG	145	CGCCCTCGCCATCG TTGGGCCAGATCTG CCCGTCTTGCG	146
	<i>e19a2</i>	CGUUCGGUGUGGCCACGGACAUCCAGGCACUG AAGCAGCCUUCGACGUCUCAAAGCCUUCAGCGGC CA	147	TGGCCGCTGAAGGG CTTTGACGTCGAA GCTGCCT	148	TCAGTGCCTGGATG TCCGTGGCCACACC GGACACG	149
	<i>b3a3</i>	UUGGUUUCUGAAUGUCAUCGUGCCACUCAGCCACU GGAUUUAGCAGAGAGUUCUCAAAGUGAAAAGCUCGCCGG UC	150	GACCCGGAGCTTTT CACTTGAACCTGTC TTAAATC	151	CAGTGGCTGAGTGG ACGATGACATTTCAG AAACCCA	152
	<i>b2a3</i>	GUGUGUAAAACUCCAGACUCUGUCCACAGCAUUCG CUGACCAUCAAUUAGGAAGGUGUAAAAGCUCGCCGG UC	153	GACCCGGAGCTTTT CACCTTCTTATTGA TGGTCA	154	GCGGAATGCTGTGG ACAGTCTGGAGTTT CACACAC	155
	<i>e1a3</i>	CGCAAGACCGGGCAGAUUCUGGCCCAACGAUGGGC AGGCGCCUUCUCCAUUGGAGACGCAGGUGAAAAGCU CCGGGUC	156	GACCCGGAGCTTTT CACCTGCACTTCCA TGGAAGG	157	CGCCCTCGCCATCG TTGGGCCAGATCTG CCCGTCTTGCG	158
AML-ETO		CUACCACAGAGCCAUCAAAUACACAGUGGAUGGG CCCCGAGAACCCUGAAUUCGUACUGAGAAGCACU CCACAUGCCAGACUACCCUGUGGAUGUGAAG	159	CTTACATCCACAG GTGAGTCTGGCAAT GTGAGTGTCTCTC AGTACGAT	160	TTGAGGTTCTCGG GGCCCATCCACTGT GATTTGATGGCTC TGTGGTAG	161
	<i>e5e12</i>	GAAUUGGAGGUGAGAGUUUCACAGCUGCUGGCAG UACUGGUCCAUGAGCUGGAGAGUCCAAAGCAGCGG CCUUGGA	162	TCCAGGGCCCGCTT GGACTTCTCCAGCT CATGGAC	163	CAGTTACTGCCAGC AGCTGTGAAACTCT CACCTCCATTTC	164
CBFb- MYH11	<i>e5e8</i>	GAAUUGGAGGUGAGAGUUUCACAGCUGCUGGCAG UACUGGCCAAGCGCAACCUAGACAAGAAUAAAGC AGACGCU	165	CGGCTCGTTAAGC ATCCCTGTGACGCT CTCAACTTCAT	166	CAGTTACTGCCAGC AGCTGTGAAACTCT CACCTCCATTTC	167
	<i>e5e7</i>	GAAUUGGAGGUGAGAGUUUCACAGCUGCUGGCAG UACUGGCCAAGCGCAACCUAGACAAGAAUAAAGC AGACGCU	168	AGCGTCTGCTTATT CTTGCTAGGTTGCG CCTTGGC	169	CAGTTACTGCCAGC AGCTGTGAAACTCT CACCTCCATTTC	170
<i>e13- insertio n-n27- e2</i>		ACCUCCUGACCCUGUCUGGGCCUCCCGACUCCUA CAGUGAUGAAAGUGUUCGGUACCCUGGAACCUUU U	171	AAAAGGTTCCAGGT GACCGAACACTTTC ATCACTG	172	TAGGAGTCGGGAG GCCGAGACAGGTC AGGGAGGT	173
E2A-PBX1	<i>e13e2</i>	AGCCAGCCAGGCACCCUCCUGACCCUUGUCUGGCC UCCCGACUCCUACAGUGUUUUUGAGUAUCCGAGGA	174	TCCTCGGATACTCA AAACTGTAGGAG	175	GCCGAGACAGGTC AGGAGGCTGCCT	176

MLL-AF4	<i>e8e7</i>	CAAAACCAUCCUAGUGAGCCCAAGAAAAAGCAG CCUCCACCACAGAAUCAGAACAAUUAUGAUACA CU	177	AGATGTATCATATT GTTCTGATTCTGGT GGTGGAG	178	GGCTGGCTG	179		
	<i>e9e5</i>	AAAAAGUGGCUCCCCGCCCAAGUAUCCCUGUAAA ACAAAAACCAAAAGAAAAAGAAUUGACCCAUUCA UG	180	CATGAATGGGTCAT TTCCTTTTCTTTTGG TTTTTG	181	TTTACAGGGATAC TTGGCGGGGAGC CACTTTT	182		
	<i>e9e4</i>	AAAAAGUGGCUCCCCGCCCAAGUAUCCCUGUAAA ACAAAAACCAAAAGAAAAAGCAGACCUACUCCAAU GA	183	TCATTGGAGTAGGT CTGCTTTTCTTTTGG TTTTTG	184	TTTACAGGGATAC TTGGCGGGGAGC CACTTTT	185		
	<i>e10e5</i>	CUAAGCAAAAAUUUCCAGCAGAUUGGAGUCCACAG GAUCAGAGUGGACUUUAAGGAAUUGACCCAUUCA UG	186	CATGAATGGGTCAT TTCCTTAAAGTCCA CTCTGAT	187	CCTGTGGACTCCAT CTGCTGGAATTTT TGCTTAGC	188		
	<i>e10e4</i>	CUAAGCAAAAAUUUCCAGCAGAUUGGAGUCCACAG GAUCAGAGUGGACUUUAAGCAGACCUACUCCAAU GA	189	TCATTGGAGTAGGT CTGCTTAAAGTCCA CTCTGAT	190	CCTGTGGACTCCAT CTGCTGGAATTTT TGCTTAG	191		
	<i>e11e5</i>	CUAUAACACCCAGGUGGUUUGCUUUCUCUGUGC CAGUAGUGGGCAUGUAGAGGAAUUGACCCAUUCA UG	192	CATGAATGGGTCAT TTCCTATACATGCC CACTACT	193	GGCACAGAGAAAG CAAAACCACCTGGG TGTTATAG	194		
	<i>e11e4</i>	CUAUAACACCCAGGUGGUUUGCUUUCUCUGUGC CAGUAGUGGGCAUGUAGAGCAGACCUACUCCAAU GA	195	TCATTGGAGTAGGT CTGCTATACATGCC CACTACT	196	GGCACAGAGAAAG CAAAACCACCTGGG TGTTATAG	197		
	<i>bcr1</i> <i>(e6e3)</i>	AUAGGAAGUGAGGUCUUCUGGCCCAACAGCAACC ACGUGGCCAGUGGCGCCGGGAGGCAGCCAAUUGA GACCCAGA	198	TCTGGGTCTCAATG GCTGACTCCCAAGC GCCACTG	199	GCCACGTGGTTGCT GTTGGCAGGAAG ACCTCACTTCCTAT	200		
	<i>bcr3</i> <i>(e3e3)</i>	GGCUUCGACGAGUUCAAAGGUGGCCUGCAGGACC UCAGCUCUUGCAUACCCAGGGGAAAGCCAAUUGA GACCCAGA	201	TCTGGGTCTCAATG GCTTCCACTGGGT GATGCAA	202	GAGCTGAGGTCTCTG CAGGCGCACCTTGA ACTCGTCGAAGC	203		
	<i>bcr2</i> <i>(e6parti</i> <i>al_e3)</i>	CUCACACCCCAUUGGAUGGACCGCCUAGCCCCA GGAGCCCCGUCUAUAGGAACCAUUGAGACCCAGAG C	204	GCTCTGGGTCTCAA TGTTCTCTATGACG GGACTCC	205	TGGGGCTAGGCGGT CCATCCAGGTGGGG TGGTGAG	206		
TEL-AML1	<i>e5e2</i>	GGUCUCUGUCUCCCCGCCUGAAGAGCACGCCAUG CCCAUUGGGAGAAUAGCAGGUGGUGGCCCUAGGG GA	207	TCCCTAGGGCCAC CACCTGCTATTCTC ACAATGG	208	GCATGGCGTGTCTCT TCAGGCGGGGAGA CAGAGACC	209		
	<i>e5e3</i>	GGUCUCUGUCUCCCCGCCUGAAGAGCACGCCAUG CCCAUUGGGAGAAUAGCAGGGGAAAGCUUCACUC UG	210	CAGAGTGAAGCTTT TCCCTGCTATTCTAC CAATGG	211	GCATGGCGTGTCTCT TCAGGCGGGGAGA CAGAGACC	212		

nCounter® Analysis System overview

[85] The basis of the nCounter® Analysis system is the unique code assigned to each gene to be assayed (International Patent Application No. PCT/US2008/059959 and Geiss *et al.* Nature Biotechnology. 2008. 26(3): 317-325; the contents of which are each incorporated herein by reference in their entireties). The code is composed of an ordered series of colored fluorescent spots which create a unique barcode for each target to be assayed. A pair of probes is designed for each target, a biotinylated capture probe and a reporter probe carrying the fluorescent barcode.

[86] Specific reporter and capture probes are synthesized for each target. Briefly, sequence-specific DNA oligonucleotide probes are attached to code-specific reporter molecules. Capture probes are made by ligating a second sequence-specific DNA oligonucleotide for each target to a universal oligonucleotide containing biotin. Reporter and capture probes are all pooled into a single hybridization mixture, the “probe library”.

[87] The expression levels of all targets are measured in a single multiplexed hybridization reaction. The sample is combined with the probe library, and specific hybridization of the reporter and the capture probes to their target molecules occurs in solution. After hybridization, the tripartite hybridized complexes are purified in a two-step procedure using magnetic beads linked to oligonucleotides complementary to universal sequences present on the capture and reporter probes. This dual purification process allows the hybridization reaction to be driven to completion with a large excess of target-specific probes, as they are ultimately removed, and, thus, do not interfere with binding and imaging of the sample. All post hybridization steps are handled robotically on a custom liquid-handling robot (Prep Station, NanoString Technologies).

[88] Purified reactions are deposited by the Prep Station into individual flow cells of a sample cartridge, bound to a streptavidin-coated surface via the capture probe, electrophoresed to elongate the reporter probes, and immobilized. After processing, the sample cartridge is transferred to a fully automated imaging and data collection device (Digital Analyzer, NanoString Technologies). The expression level of a target is measured by imaging each sample and counting the number of times the code for that target is detected. For each sample, typically 600 fields-of-view (FOV) are imaged (1376 X 1024 pixels) representing approximately 10 mm² of the binding surface. Typical imaging density is 100-1200 counted reporters per field of view depending on the degree of multiplexing, the amount of RNA, and overall target expression

levels. Data is output in simple spreadsheet format listing the number of counts per target, per sample.

[89] In certain embodiments the analysis system used herein also used a reference sample as a control. The reference samples are known quantities of nucleic acid fragments with sequences that correspond to one or more fusion genes. For example, a reference sample with a sequence corresponding to a specific fusion gene is used with a nanoreporter with a target-specific probe that specifically binds to a sequence on the same fusion gene. When multiple nanoreporters are used for multiple fusion genes, the reference sample may contain multiple target nucleic acids which correspond to the sequences that the target specific probes of the nanoreporters specifically bind to.

Kits

[90] The present disclosure also describes kits useful for detecting and diagnosing leukemia. These kits comprise a set of probes described above. For example, the kit may include any number of probes described in Table 2, above. The kit may further comprise a computer readable medium.

[91] In another embodiment, the kit includes probes that may be provided in a lyophilized or reconstituted form, or may be provided as a set of nucleotide sequences. In one embodiment, the probes are provided in a microplate format, where each probe set occupies a well (or multiple wells, as in the case of replicates) in the microplate. The kit may further comprise reagents and instructions sufficient for the detection of the various target sequences described above.

EXAMPLE

[92] **Example 1. Functional performance of fusion probes showing specificity for fusion gene isoforms on synthetic gene fragments.**

[93] A multiplexed collection of nanoreporters containing probes to all the isoforms of BCR-ABL, E2A-PBX1, TEL-AML1 and PML-RARA listed in Table 1 was used to detect synthetic transcripts of sequences corresponding with each of these gene fusions. An nCounter® leukemia fusion gene expression assay was performed as described below, and some of the resulting counts for each probe are shown in log₂ scale in Figure 1. Cross-hybridization for each individual probe to non-specific targets was below 1%.

[94] The leukemia fusion gene expression assay was run on the nCounter® Analysis system (NanoString). Each sample was hybridized to the multiplexed nanoreporters in solution, followed by post-hybridization sample processing on the nCounter® Prep Station and data collection on the nCounter® Digital Analyzer.

[95] Hybridization Protocol

[96] Each hybridization reaction contained the following components: 10 µL Reporter CodeSet, 10 µL hybridization buffer, 5 µL synthetic RNA target to a final concentration of 1 pM, and 5 µL Capture ProbeSet.

[97] Aliquots of both the Reporter CodeSet and Capture ProbeSet reagent were removed from the freezer and thawed on ice. The aliquots were inverted several times to mix well. The reagent was briefly spun down at <1000 rpm.

[98] A master mix was created containing 130 µL of the Reporter CodeSet and 130 µL of hybridization buffer by adding the hybridization buffer to the tube containing the Reporter CodeSet. The master mix was inverted to mix and spun down, and 20 µL of master mix was added to each of 12 tubes.

[99] Samples of synthetic target transcripts were denatured 85°C for 5 minutes and quick-cooled on ice before being added to each tube and mixed. Targets were added to a final concentration of 1 pM.

[100] The thermocycler was pre-heated to 65°C. 5 µL of Capture ProbeSet were added to each tube and mixed well before placing immediately at 65°C. Hybridization assays were incubated at 65°C overnight.

[101] Hybridizations were left at 65°C until ready for processing. Once removed from 65°C, the assay proceeded immediately to post-hybridization processing on the nCounter® Prep Station.

[102] Post-Hybridization Processing and Data Collection

[103] After hybridization, an automated protocol on the nCounter® Prep Station was used to wash away excess probes and bind the captured probe/target complexes onto a cartridge, and then stretch and immobilize the captured nanoreporters in preparation for imaging. The purification procedure involved a two-step magnetic beadbased affinity purification. Magnetic beads derivatized with short nucleic acid sequences that were complementary to the Capture Probes and the Reporter Probes were used sequentially. First, the hybridization mixture

containing target/probe complexes was allowed to bind to magnetic beads complementary to sequences on the Capture Probe. Wash steps were performed to remove excess Reporter Probes. After washing, the Capture Probes and target/probe complexes were eluted off the beads and hybridized to magnetic beads complementary to sequences on the Reporter Probe. An additional wash was performed to remove excess Capture Probes. Finally, the purified Capture Probe/target/Reporter Probe complexes were eluted off the beads and the captured nanoreporters were bound, stretched and immobilized on a cartridge for data collection.

[104] Automated data collection was carried out in the nCounter® Digital Analyzer. Digital images of 600 fields of view of the bound nanoreporters were processed and the counts for each nanoreporter barcode were tabulated in a comma separated value (CSV) format.

[105] Example 2. Leukemia Fusion Gene Panel Tested on total RNA isolated from cell lines.

[106] Figures 2A-F show results from a Leukemia Fusion Gene Panel tested on total RNA isolated from 5 cell lines. The assay was performed as described in Example 1 with 100ng of input RNA. Probes included in the CodeSet identified their respective targets expressed in the different cell lines: the BCR-ABL e1a2 isoform expressed in cell line SUP B13 (A), the BCR-ABL b3a2 isoform expressed in cell line K562 (B), the PML-RARA bcr1 isoform expressed in cell line NB4 (C), the AML-ETO fusion gene transcript expressed in the Kasumi cell line (D), the BCR-ABL b2a2 isoform expressed in cell line KCL-22 (E) and the MLL-AF4 e9e5 isoform expressed in cell line MV4-11 (F).

OTHER EMBODIMENTS

[107] While the disclosure has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the disclosure, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

[108] The patent and scientific literature referred to herein establishes the knowledge that is available to those with skill in the art. All United States patents and published or unpublished United States patent applications cited herein are incorporated by reference. All published foreign patents and patent applications cited herein are hereby incorporated by reference. Genbank and NCBI submissions indicated by accession number cited herein are hereby incorporated by reference. All other published references, documents, manuscripts and scientific literature cited

herein are hereby incorporated by reference.

[109] While this disclosure has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the disclosure encompassed by the appended claims.

CLAIMS

What is claimed is:

1. A composition for the diagnosis of leukemia or the identification of a subtype of leukemia comprising one or more target-specific nucleic acid probes wherein each of the one or more target-specific probes comprises
 - (i) a target-specific region that specifically binds to one target nucleic acid analyte; and
 - (ii) a region to which is attached a plurality of label monomers that create a unique label for each target-specific probe, said label having a detectable signal that distinguishes one target-specific probe which binds to a first target nucleic acid from another target-specific probe that binds to a different second target nucleic acid,wherein one or more targets of the one or more target-specific probes comprise a fusion gene or mRNA transcribed from said fusion gene and wherein the fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia.
2. The composition of claim 1, wherein the fusion gene is selected from the group consisting of *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*.
3. The composition of claim 1, wherein the leukemia is selected from the group consisting of chronic myelogenous leukemia, acute myeloid leukemia and acute lymphoblastic leukemia.
4. The composition of claim 1, wherein at least one of the one or more target-specific probes has a target that is the sequence of the junction of the two genes that are fused.
5. The composition of claim 1, wherein at least one of the one or more target-specific probes comprises a target-specific region with one or more mismatches to the target of the target-specific region.
6. The composition of claim 1, wherein each of said label monomers are selected from the group consisting of a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety.

7. The composition of claim 6, wherein each of said label monomers comprise a fluorescent moiety.

8. The composition of claim 1, wherein the unique label for each target-specific probe is composed of 4 label monomers.

9. The composition of claim 1, wherein the unique label for each target-specific probe is composed of 5 label monomers.

10. A composition for the diagnosis of leukemia comprising a probe pair comprising a first probe and a second probe, said first probe comprising a complex comprising:

(a) a first molecule, comprising:

(i) a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal;

(ii) a second label attachment region, which is non-overlapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and

(b) a first target-specific sequence attached to the first molecule, said second probe comprising a second molecule, comprising

(i) a second target-specific sequence; and

(ii) an affinity tag

wherein the first target-specific sequence and the second target-specific sequence bind to different regions of the same target molecule, wherein the target molecule is a fusion gene or mRNA transcribed from said fusion gene, wherein the fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia and wherein when said probe pair is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitute at least part of a code that identifies the target molecule.

11. The composition of claim 10, wherein the fusion gene comprises the fusion of at least a portion of a first gene and at least a portion of a second gene.
12. The composition of claim 11, wherein the first target-specific sequence hybridizes to a portion of the first gene and the second target-specific sequence hybridizes to a portion of the second gene.
13. The composition of claim 12, wherein the first target-specific sequence hybridizes to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.
14. The composition of claim 11, wherein the first target-specific sequence hybridizes to a portion of the first gene and a portion of the second gene.
15. The composition of claim 14, wherein the first target-specific sequence hybridizes to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.
16. The composition of claim 14, wherein at least one of the target-specific sequences comprises a region with one or more mismatches to the target of the target-specific sequence.
17. The composition of claim 11, wherein the second target-specific sequence hybridizes to a portion of the first gene and a portion of the second gene.
18. The composition of claim 17, wherein the first target-specific sequence hybridizes to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.
19. The composition of claim 17, wherein at least one of the target-specific sequences comprises a region with one or more mismatches to the target of the target-specific sequence.

20. The composition of claim 10, wherein the leukemia is selected from the group consisting of chronic myelogenous leukemia, acute myeloid leukemia and acute lymphoblastic leukemia.
21. The composition of claim 10, wherein the fusion gene is selected from the group consisting of *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*.
22. The composition of claim 10, wherein each of said label monomers are selected from the group consisting of a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety.
23. The composition of claim 22, wherein each of said label monomers comprise a fluorescent moiety.
24. The composition of claim 22, wherein the unique label for each target-specific probe is composed of 4 label monomers.
25. The composition of claim 22, wherein the unique label for each target-specific probe is composed of 5 label monomers.
26. A method for diagnosing leukemia, a specific subtype of leukemia or determining a genetic aberration underlying leukemia in a subject comprising
- (a) providing a sample from the subject; and
 - (b) contacting the sample with one or more target-specific probes wherein each of the one or more target-specific probes comprise
 - (i) a target-specific region that binds to one target nucleic acid analyte; and
 - (ii) a region to which is attached a plurality of label monomers that create a unique label for each target-specific probe, said label having a detectable signal that distinguishes one target-specific probe which binds to a first target nucleic acid from another target-specific probe that binds to a different second target nucleic acid,

wherein one or more targets of the one or more target-specific probes comprise a fusion gene or mRNA transcribed from said fusion gene that is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia and wherein when the fusion gene is detected the subject is diagnosed with leukemia, a specific subtype of leukemia or the genetic aberration underlying leukemia is determined.

27. The method of claim 26, wherein fusion gene is selected from the group consisting of *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*.

28. The method of claim 26, wherein the leukemia is selected from the group consisting of chronic myelogenous leukemia, acute myeloid leukemia and acute lymphoblastic leukemia.

29. The method of claim 26, wherein at least one of the one or more target-specific probes has a target that is the sequence of the junction of the two genes that are fused.

30. The method of claim 26, wherein at least one of the one or more target-specific probes comprises a target-specific region with one or more mismatches to the target of the target-specific region.

31. The method of claim 26, wherein each of said label monomers are selected from the group consisting of a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety.

32. The method of claim 31, wherein each of said label monomers comprise a fluorescent moiety.

33. The method of claim 26, wherein the unique label for each target-specific probe is composed of 4 label monomers.

34. The method of claim 26, wherein the unique label for each target-specific probe is composed of 5 label monomers.

35. A method for diagnosing leukemia, a specific subtype of leukemia or determining a genetic aberration underlying leukemia in a subject comprising:

- (a) providing a sample from the subject; and
- (b) contacting the sample with a probe pair comprising a first probe and a second probe, said first probe being a complex comprising:

- (i) a first molecule, comprising:

- (1) a first label attachment region to which are attached one or more label monomers that emit light constituting a first signal;

- (2) a second label attachment region, which is non-overlapping with the first label attachment region, to which are attached one or more label monomers that emit light constituting a second signal; and

- (ii) a first target-specific sequence attached to the first molecule,
- said second probe comprising a second molecule, comprising

- (1) a second target-specific sequence; and

- (2) an affinity tag

wherein the first target-specific sequence and the second target-specific sequence bind to different regions of the same target molecule, wherein the target molecule is a fusion gene or mRNA transcribed from said fusion gene, wherein the fusion gene is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia and wherein when said probe pair is bound to its target molecule, the identity of the first and second signals and their locations relative to each other constitute at least part of a code that identifies the target molecule and wherein when the fusion gene is detected the subject is diagnosed with leukemia, a specific subtype of leukemia or the genetic aberration underlying leukemia is determined.

36. The method of claim 35, wherein the fusion gene comprises the fusion of at least a portion of a first gene and at least a portion of a second gene.

37. The method of claim 35, wherein the first target-specific sequence hybridizes to a portion of the first gene and the second target-specific sequence hybridizes to a portion of the second gene.

38. The method of claim 37, wherein the first target-specific sequence hybridizes to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.

39. The composition of claim 35, wherein the first target-specific sequence hybridizes to a portion of the first gene and a portion of the second gene.

40. The method of claim 39, wherein the first target-specific sequence hybridizes to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.

41. The method of claim 39, wherein at least one of the target-specific sequences comprises a region with one or more mismatches to the target of the target-specific sequence.

42. The method of claim 35, wherein the second target-specific sequence hybridizes to a portion of the first gene and a portion of the second gene.

43. The method of claim 42, wherein the first target-specific sequence hybridizes to a sequence adjacent to a sequence that the second target-specific sequence hybridizes to on the fusion gene.

44. The method of claim 42, wherein at least one of the target-specific sequences comprises a region with one or more mismatches to the target of the target-specific sequence.

45. The method of claim 35, wherein the leukemia is selected from the group consisting of chronic myelogenous leukemia, acute myeloid leukemia and acute lymphoblastic leukemia.

46. The method of claim 35, wherein the fusion gene is selected from the group consisting of *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*.

47. The method of claim 35, wherein each of said label monomers are selected from the group consisting of a fluorochrome moiety, a fluorescent moiety, a dye moiety or a chemiluminescent moiety.
48. The method of claim 47, wherein each of said label monomers comprise a fluorescent moiety.
49. The method of claim 35, wherein the unique label for each target-specific probe is composed of 4 label monomers.
50. The method of claim 35, wherein the unique label for each target-specific probe is composed of 5 label monomers.
51. A kit comprising one or more reporter probes and one or more capture probes, wherein the reporter probe and the capture probe each bind specifically to a fusion gene that is the result of the fusion of two or more genes in a patient at the onset of or during progression of leukemia.
52. The kit of claim 51, wherein the fusion gene is selected from the group consisting of *BCR-ABL*, *AML-ETO*, *CBFB-MYH11*, *E2A-PBX1*, *MLL-AF4*, *PML-RARA*, or *TEL-AML1*.
53. The kit of claim 51, comprising one or more reporter probes and one or more capture probes from Table 2.

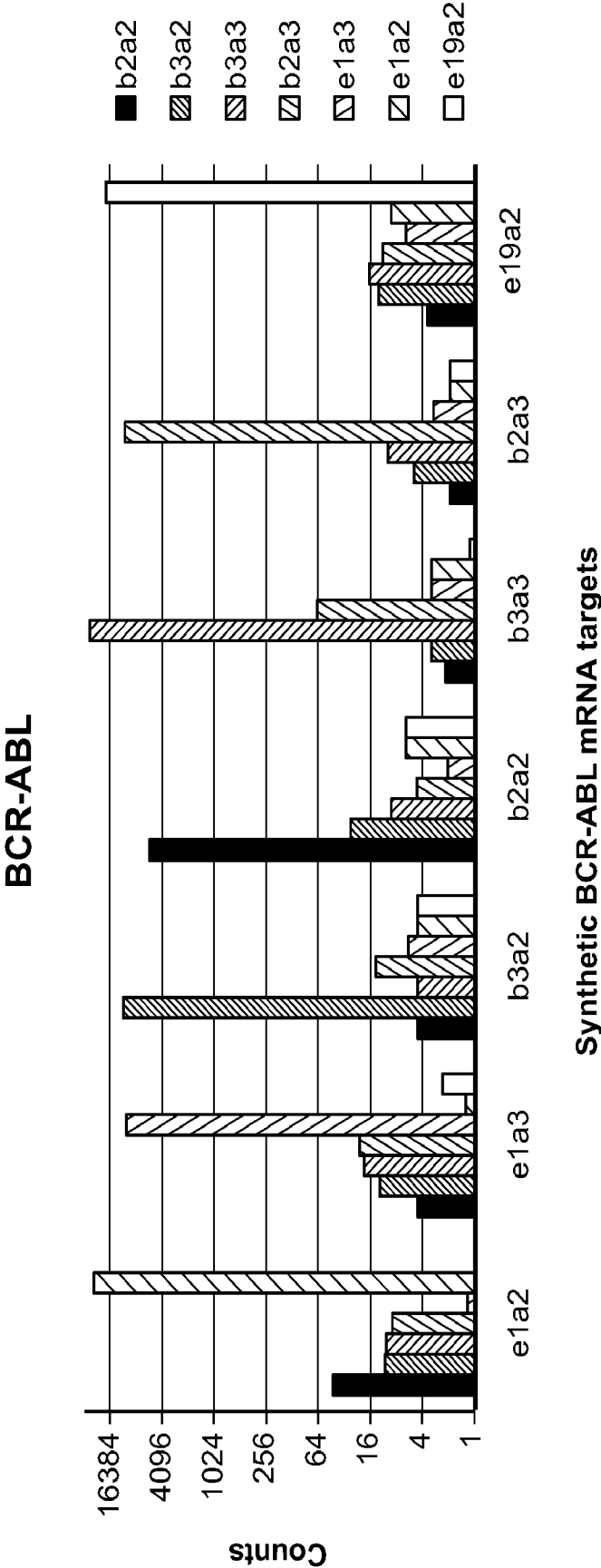


FIG. 1A

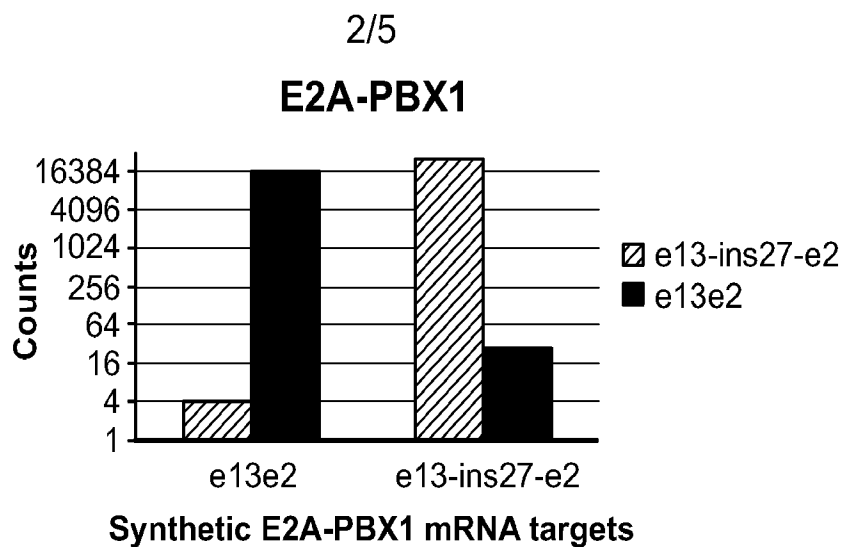


FIG. 1B

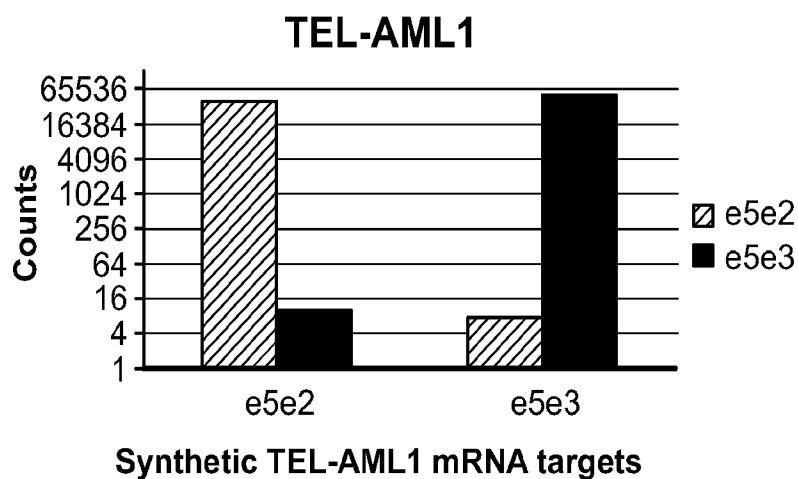


FIG. 1C

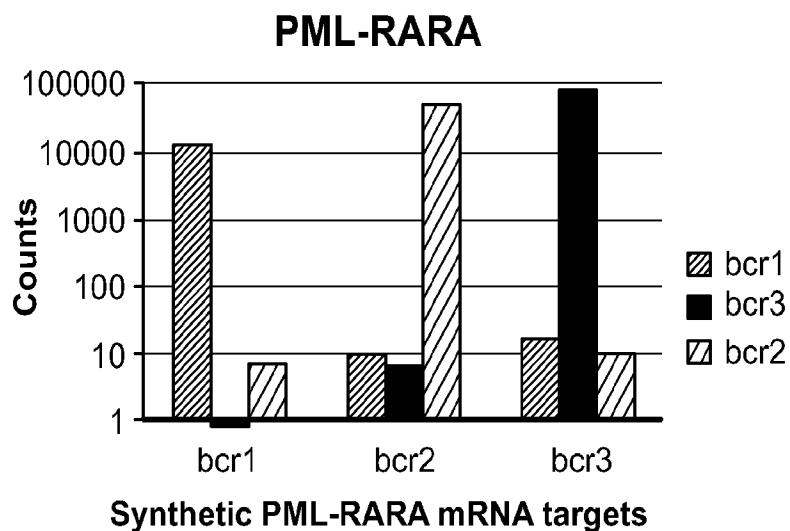


FIG. 1D

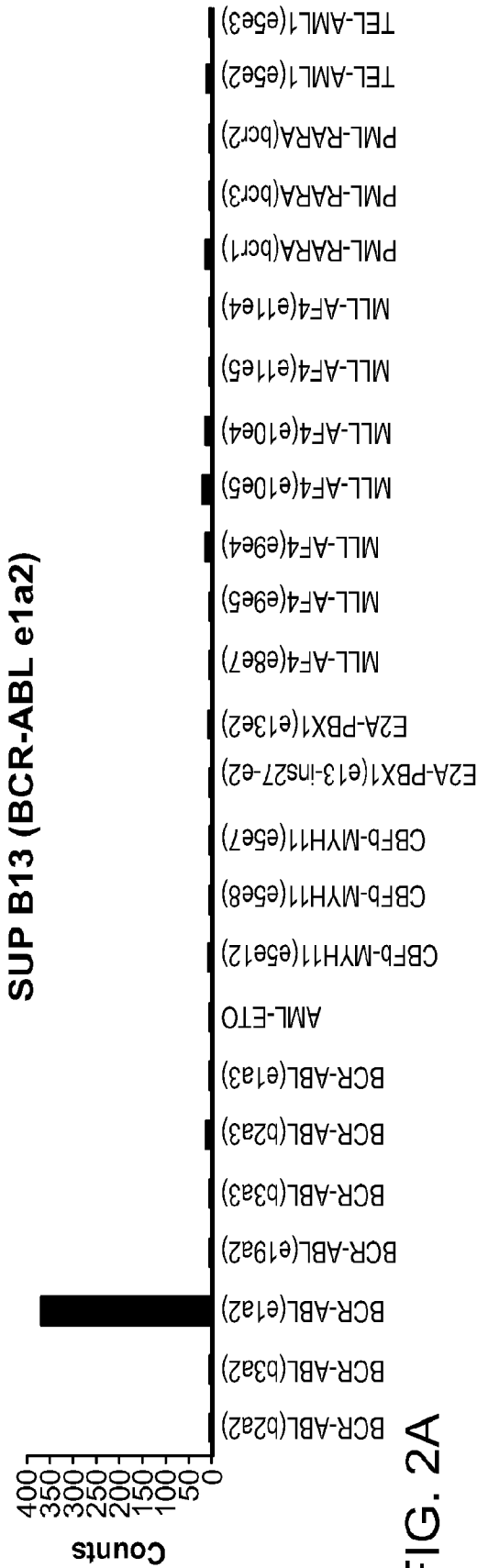


FIG. 2A

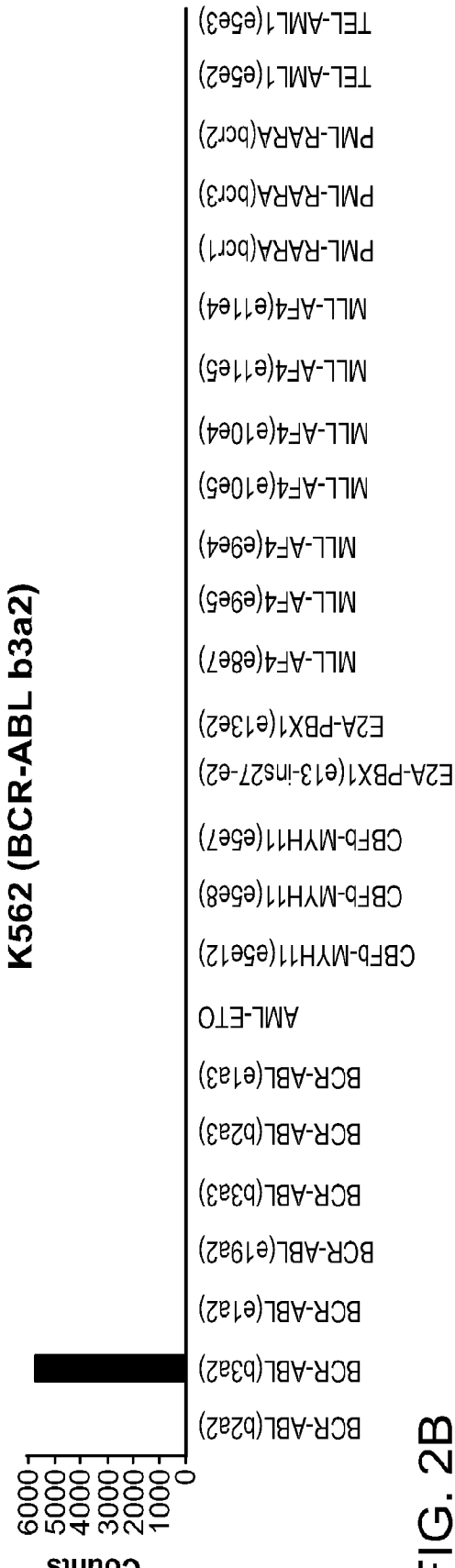


FIG. 2B

