**(12) STANDARD PATENT** (11) Application No. **AU 2018374176 B2**
**(19) AUSTRALIAN PATENT OFFICE**

(54) Title
**Detecting breast cancer**

(51) International Patent Classification(s)
***C12N 15/09*** (2006.01) ***C40B 30/04*** (2006.01)
***C12Q 1/68*** (2018.01)

(21) Application No: **2018374176** (22) Date of Filing: **2018.11.28**

(87) WIPO No: **WO19/108626**

(30) Priority Data

| (31) Number | (32) Date | (33) Country |
|---|---|---|
| **62/592,828** | **2017.11.30** | **US** |

(43) Publication Date: **2019.06.06**
(44) Accepted Journal Date: **2024.12.12**

(71) Applicant(s)
**Exact Sciences Corporation;Mayo Foundation for Medical Education and Research**

(72) Inventor(s)
**AHLQUIST, David A.;TAYLOR, William R.;MAHONEY, Douglas W.;YAB, Tracy C.;KISIEL, John B.;ALLAWI, Hatim T.;LIDGARD, Graham P.;KAISER, Michael W.**

(74) Agent / Attorney
**Spruson & Ferguson, GPO Box 3898, Sydney, NSW, 2001, AU**

(56) Related Art
**WO 2016/041010 A1**
**Bediaga et al. SpringerPlus (2016) 5:623 DOI 10.1186/s40064-016-2235-0**
**PLoS ONE, vol. 9, no. 1, 9 January 2014, page e81843, DOI: 10.1371/journal.pone.0081843**
**Clare Stirzaker, et al., NATURE COMMUNICATIONS | 6:5899 | DOI: 10.1038/ncomms6899 | www.nature.com/naturecommunications**
**BEDIAGA NAIARA G ET AL: "DNA methylation epigenotypes in breast cancer molecular subtypes", BREAST CANCER RESEARCH, vol. 12, no. 5, 29 September 2010 (2010-09-29), pages R77, XP021085389, ISSN: 1465-5411, DOI: 10.1186/BCR2721**
**MATAHI MOARII ET AL: "Integrative DNA methylation and gene expression analysis to assess the universality of the CpG island methylator phenotype", HUMAN GENOMICS, vol. 9, no. 1, 2015-10-13, p 26, DOI: 10.1186/S40246-015-0048-9**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
06 June 2019 (06.06.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/108626 A1

(51) International Patent Classification:
*C12N 15/09* (2006.01) *C40B 30/04* (2006.01)
*C12Q 1/68* (2018.01)

(21) International Application Number:
PCT/US2018/062809

(22) International Filing Date:
28 November 2018 (28.11.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/592,828 30 November 2017 (30.11.2017) US

(71) Applicants: **MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH** [US/US]; 200 First Street S.W., Rochester, Minnesota 55905 (US). **EXACT SCIENCES DEVELOPMENT COMPANY, LLC** [US/US]; 441 Charmany Drive, Madison, Wisconsin 53719 (US).

(72) Inventors: **AHLQUIST, David A.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **TAYLOR, William R.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **MAHONEY, Douglas W.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **YAB, Tracy C.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **KISIEL, John B.**; c/o Mayo Foundation for Medical Education and Research, 200 First Street S.W., Rochester, Minnesota 55905 (US). **ALLAWI, Hatim T.**; c/o Exact Sciences Development Company, LLC, 441 Charmany Drive, Madison, Wisconsin 53719 (US). **LIDGARD, Graham P.**; c/o Exact Sciences Development Company, LLC, 441 Charmany Drive, Madison, Wisconsin 53719 (US). **KAISER, Michael W.**; c/o Exact Sciences Development Company, LLC, 441 Charmany Drive, Rochester, Minnesota 55905 (US).

(74) Agent: **GOETZ, Robert, A.**; Casimir Jones, S.C., 2275 Deming Way, Ste 310, Middleton, Wisconsin 53562 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

**Published:**
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: DETECTING BREAST CANCER

(57) Abstract: Provided herein is technology for breast cancer screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of breast cancer.

WO 2019/108626 A1

# DETECTING BREAST CANCER

## CROSS REFERENCE TO RELATED APPLICATION

This application claims priority to and the benefit of U.S. Provisional Application No. 62/592,828, filed November 30, 2017, the content of which is hereby incorporated by reference in its entirety.

## FIELD OF INVENTION

Provided herein is technology for breast cancer screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of breast cancer.

## BACKGROUND

Breast cancer affects approximately 230,000 US women per year and claims about 40,000 lives every year. Although carriers of germline mutations in BRCA1 and BRCA2 genes are known to be at high risk of breast cancer, most women who get breast cancer do not have a mutation in one of these genes and there is limited ability to accurately identify women at increased risk of breast cancer. Effective prevention therapies exist, but current risk prediction models do not accurately identify the majority of women at increased risk of breast cancer (see, e.g., Pankratz VS, et al., J Clin Oncol 2008 Nov 20; 26(33):5374-9).

Improved methods for detecting breast cancer are needed.

The present invention addresses these needs.

## SUMMARY

Methylated DNA has been studied as a potential class of biomarkers in the tissues of most tumor types. In many instances, DNA methyltransferases add a methyl group to DNA at cytosine-phosphate-guanine (CpG) island sites as an epigenetic control of gene expression. In a biologically attractive mechanism, acquired methylation events in promoter regions of tumor suppressor genes are thought to silence expression, thus contributing to oncogenesis. DNA methylation may be a more chemically and biologically stable diagnostic tool than RNA or protein expression (Laird (2010) Nat Rev Genet 11: 191–203). Furthermore, in other cancers like sporadic colon cancer, methylation markers offer excellent specificity and are

more broadly informative and sensitive than are individual DNA mutations (Zou et al (2007) Cancer Epidemiol Biomarkers Prev 16: 2686–96).

Analysis of CpG islands has yielded important findings when applied to animal models and human cell lines. For example, Zhang and colleagues found that amplicons from different parts of the same CpG island may have different levels of methylation (Zhang et al. (2009) PLoS Genet 5: e1000438). Further, methylation levels were distributed bi-modally between highly methylated and unmethylated sequences, further supporting the binary switch-like pattern of DNA methyltransferase activity (Zhang et al. (2009) PLoS Genet 5: e1000438). Analysis of murine tissues in vivo and cell lines in vitro demonstrated that only about 0.3% of high CpG density promoters (HCP, defined as having >7% CpG sequence within a 300 base pair region) were methylated, whereas areas of low CpG density (LCP, defined as having <5% CpG sequence within a 300 base pair region) tended to be frequently methylated in a dynamic tissue-specific pattern (Meissner et al. (2008) Nature 454: 766–70). HCPs include promoters for ubiquitous housekeeping genes and highly regulated developmental genes. Among the HCP sites methylated at >50% were several established markers such as Wnt 2, NDRG2, SFRP2, and BMP3 (Meissner et al. (2008) Nature 454: 766–70).

Epigenetic methylation of DNA at cytosine-phosphate-guanine (CpG) island sites by DNA methyltransferases has been studied as a potential class of biomarkers in the tissues of most tumor types. In a biologically attractive mechanism, acquired methylation events in promotor regions of tumor suppressor genes are thought to silence expression, contributing to oncogenesis. DNA methylation may be a more chemically and biologically stable diagnostic tool than RNA or protein expression. Furthermore, in other cancers like sporadic colon cancer, aberrant methylation markers are more broadly informative and sensitive than are individual DNA mutations and offer excellent specificity.

Several methods are available to search for novel methylation markers. While micro-array based interrogation of CpG methylation is a reasonable, high-throughput approach, this strategy is biased towards known regions of interest, mainly established tumor suppressor promotors. Alternative methods for genome-wide analysis of DNA methylation have been developed in the last decade. There are three basic approaches. The first employs digestion of DNA by restriction enzymes which recognize specific methylated sites, followed by several possible analytic techniques which provide methylation data limited to the enzyme recognition site or the primers used to amplify the DNA in quantification steps (such as

methylation-specific PCR; MSP). A second approach enriches methylated fractions of genomic DNA using anti-bodies directed to methyl-cytosine or other methylation-specific binding domains followed by microarray analysis or sequencing to map the fragment to a reference genome. This approach does not provide single nucleotide resolution of all methylated sites within the fragment. A third approach begins with bisulfite treatment of the DNA to convert all unmethylated cytosines to uracil, followed by restriction enzyme digestion and complete sequencing of all fragments after coupling to an adapter ligand. The choice of restriction enzymes can enrich the fragments for CpG dense regions, reducing the number of redundant sequences which may map to multiple gene positions during analysis.

RRBS yields CpG methylation status data at single nucleotide resolution of 80-90% of all CpG islands and a majority of tumor suppressor promoters at medium to high read coverage. In cancer case - control studies, analysis of these reads results in the identification of differentially methylated regions (DMRs). In previous RRBS analysis of pancreatic cancer specimens, hundreds of DMRs were uncovered, many of which had never been associated with carcinogenesis and many of which were unannotated. Further validation studies on independent tissue samples sets confirmed marker CpGs which were 100% sensitive and specific in terms of performance.

Provided herein is technology for breast cancer screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of breast cancer.

Indeed, as described in Examples I, II and III, experiments conducted during the course for identifying embodiments for the present invention identified a novel set of differentially methylated regions (DMRs) for discriminating cancer of the breast derived DNA from non-neoplastic control DNA.

Such experiments list and describe 375 novel DNA methylation markers distinguishing breast cancer tissue from benign breast tissue (see, Tables 2 and 18, Examples I, II and III).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing breast cancer tissue from benign breast tissue:

- ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B,

MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1 (see, Table 16E, Example II); and

- ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B, FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TRH_A, and TRIM67_B (see, Table 22, Example III).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers for detecting breast cancer in blood samples (e.g., plasma samples, whole blood samples, serum samples):

- CD1D, ITPRIPL1, FAM59B, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B (see, Table 27, Example III).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing triple negative breast cancer tissue from benign breast tissue:

- ABLIM1, AJAP1_B, ASCL2, ATP6V1B1, BANK1, CALN1_A, CALN1_B, CLIC6, DSCR6, FOXP4, GAD2, GCGR, GP5, GRASP, HBM, HNF1B_B, KLF16, MAGI2, MAX.chr11.14926602-14927148, MAX.chr12.4273906-4274012, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr2.97193478-97193562, MAX.chr22.42679578-42679917, MAX.chr4.8859253-8859329, MAX.chr4.8859602-8859669, MAX.chr4.8860002-

8860038, MAX.chr5.145725410-145725459, MAX.chr6.157557371-157557657, MPZ, NKX2-6, PDX1, PLXNC1_A, PPARG, PRKCB, PTPRN2, RBFOX_A, SCRT2_A, SLC7A4, STAC2_B, STX16_A, STX16_B, TBX1, TRH_A, VSTM2B_A, ZBTB16, ZNF132, and ZSCAN23 (see, Table 3, Example I);

- CALN1_A, LOC100132891, NACAD, TRIM67, ATP6V1B1, DLX4, GP5, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ST8SIA4, STX16_B ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, KCNK9, SCRT2_B, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, and DSCR6 (see, Table 11, Example I);
- ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4 (see, Table 16A, Example II).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing $HER2^+$ breast cancer tissue from benign breast tissue:

- ABLIM1, AFAP1L1, AKR1B1, ALOX5, AMN, ARL5C, BANK1, BCAT1, BEGAIN, BEST4, BHLHE23_B, BHLHE23_C, C17orf64, C1QL2, C7orf52, CALN1_B, CAV2, CD8A, CDH4_A, CDH4_B, CDH4_C, CDH4_D, CDH4_E, CDH4_F, CHST2_B, CLIP4, CR1, DLK1, DNAJC6, DNM3_A, EMX1_A, ESPN, FABP5, FAM150A, FLJ42875, GLP1R, GNG4, GYPC_A, HAND2, HES5, HNF1B_A, HNF1B_B, HOXA1_A, HOXA1_B, HOXA7_A, HOXA7_B, HOXA7_C, HOXD9, IGF2BP3_A, IGF2BP3_B, IGSF9B_A, IL15RA, INSM1, ITPKA_B, ITPRIPL1, KCNE3, KCNK17_B, LIME1, LOC100132891, LOC283999, LY6H, MAST1, MAX.chr1.158083198-158083476, MAX.chr1.228074764-228074977, MAX.chr1.46913931-46913950, MAX.chr10.130085265-130085312, MAX.chr11.68622869-68622968, MAX.chr14.101176106-101176260, MAX.chr15.96889069-96889128, MAX.chr17.8230197-8230314, MAX.chr19.46379903-46380197, MAX.chr2.97193163-97193287,

MAX.chr2.97193478-97193562, MAX.chr20.1784209-1784461, MAX.chr21.44782441-44782498, MAX.chr22.23908718-23908782, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.180101084-180101094, MAX.chr5.42952185-42952280, MAX.chr5.42994866-42994936, MAX.chr6.27064703-27064783, MAX.chr7.152622607-152622638, MAX.chr8.145104132-145104218, MAX.chr9.136474504-136474527, MCF2L2, MSX2P1, NACAD, NID2_B, NID2_C, ODC1, OSR2_B, PAQR6, PCDH8, PIF1, PPARA, PPP2R5C, PRDM13_A, PRHOXNB, PRKCB, RBFOX3_A, RBFOX3_B, RFX8, SNCA, STAC2_A, STAC2_B, STX16_B SYT5, TIMP2, TMEFF2, TNFRSF10D, TRH_B, TRIM67, TRIM71_C, USP44_A, USP44_B, UTF1, UTS2R, VSTM2B_A, VSTM2B_B, ZFP64, and ZNF132 (see, Table 4, Example I);

- BHLHE23_C, CALN1_A, CD1D, CHST2_A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, NACAD, TRIM67, ATP6V1B1, C17orf64, CHST2_B, DLX4, DNM3_A, EMX1_A, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, COL23A1, KCNK9, LAYN, PLXNC1_A, RIC3, SCRT2_B, ALOX5, CDH4_E, HNF1B_B, TRH_A, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, AJAP1_B, DSCR6, and MAX.chr11.68622869-68622968 (see, Table 11, Example I);
- ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6,

ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, C10orf125 (see, Table 16B, Example II).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing Luminal A breast cancer tissue from benign breast tissue:

- ARL5C, BHLHE23_C, BMP6, C10orf125, C17orf64, C19orf66, CAMKV, CD1D, CDH4_E, CDH4_F, CHST2_A, CRHBP, DLX6, DNM3_A, DNM3_B, DNM3_C, ESYT3, ETS1_A, ETS1_B, FAM126A, FAM189A1, FAM20A, FAM59B, FBN1, FLRT2, FMN2, FOXP4, GAS7, GYPC_A, GYPC_B, HAND2, HES5, HMGA2, HNF1B_B, IGF2BP3_A, IGF2BP3_B, KCNH8, KCNK17_A, KCNQ2, KLHDC7B, LOC100132891, MAX.chr1.46913931-46913950, MAX.chr11.68622869-68622968, MAX.chr12.4273906-4274012, MAX.chr12.59990591-59990895, MAX.chr17.73073682-73073814, MAX.chr20.1783841-1784054, MAX.chr21.47063802-47063851, MAX.chr4.8860002-8860038, MAX.chr5.172234248-172234494, MAX.chr5.178957564-178957598, MAX.chr6.130686865-130686985, MAX.chr8.687688-687736, MAX.chr8.688863-688924, MAX.chr9.114010-114207, MPZ, NID2_A, NKX2-6, ODC1, OSR2_A, POU4F1, PRDM13_B, PRKCB, RASGRF2, RIPPLY2, SLC30A10, ST8SIA4, SYN2, TRIM71_A, TRIM71_B, TRIM71_C, UBTF, ULBP1, USP44_B, and VSTM2B_A (see, Table 5, Example I);
- BHLHE23_C, CD1D, CHST2_A, FAM126A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, SLC30A10, TRIM67, ATP6V1B1, BANK1, C10orf125, C17orf64, CHST2_B, DNM3_A, EMX1_A, GP5, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, ST8SIA4, STX16_B UBTF, LOC100132891, ITPRIPL1, MAX.chr12.4273906-4274012, MAX.chr12.59990671-59990859, BHLHE23_D, COL23A1, KCNK9, OTX1, PLXNC1_A, HNF1B_B, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, DSCR6, MAX.chr11.68622869-68622968 (see, Table 11, Example I);

- ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125 (see, Table 16C, Example II).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing Luminal B breast cancer tissue from benign breast tissue:

- ACCN1, AJAP1_A, AJAP1_B, BEST4, CALN1_B, CBLN1_B, CDH4_E, DLX4, FOXP4, IGSF9B_B, ITPRIPL1, KCNA1, KLF16, LMX1B_A, MAST1, MAX.chr11.14926602-14927148, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr22.42679578-42679917, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr8.124173128-124173268, MPZ, PPARA, PRMT1, RBFOX3_B, RYR2_A, SALL3, SCRT2_A, SPHK2, STX16_B SYNJ2, TMEM176A, TSHZ3, and VIPR2 (see, Table 6, Example I);
- CALN1_A, LOC100132891, MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, DLX4, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, BHLHE23_D, HNF1B_B, TRH_A, ASCL2, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, AJAP1_B, and DSCR6 (see, Table 11, Example I);

- ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D (see, Table 16D, Example II).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing BRCA1 breast cancer tissue from benign breast tissue:

- C10orf93, C20orf195_A, C20orf195_B, CALN1_B, CBLN1_A, CBLN1_B, CCDC61, CCND2_A, CCND2_B, CCND2_C, EMX1_B, FAM150B, GRASP, HBM, ITPRIPL1, KCNK17_A, KIAA1949, LOC100131176, MAST1, MAX.chr1.8277285-8277316, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr18.5629721-5629791, MAX.chr19.30719261-30719354, MAX.chr22.42679767-42679917, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr6.157556793-157556856, MAX.chr8.124173030-124173395, MN1, MPZ, NR2F6, PDXK_A, PDXK_B, PTPRM, RYR2_B, SERPINB9_A, SERPINB9_B, SLC8A3, STX16_B TEPP, TOX, VIPR2, VSTM2B_A, ZNF486, ZNF626, and ZNF671 (see, Table 7, Example I);
- BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461,

MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6 (see, Table 11, Example I);

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing BRCA2 breast cancer tissue from benign breast tissue:

- ANTXR2, B3GNT5, BHLHE23_C, BMP4, CHRNA7, EPHA4, FAM171A1, FAM20A, FMNL2, FSCN1, GSTP1, HBM, IGFBP5, IL17REL, ITGA9, ITPRIPL1, KIRREL2, LRRC34, MAX.chr1.239549742-239549886, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr2.238864674-238864735, MAX.chr5.81148300-81148332, MAX.chr7.151145632-151145743, MAX.chr8.124173030-124173395, MAX.chr8.143533298-143533558, MERTK, MPZ, NID2_C, NTRK3, OLIG3_A, OLIG3_B, OSR2_C, PROM1, RGS17, SBNO2, STX16_B TBKBP1, TLX1NB, VIPR2, VN1R2, VSNL1, and ZFP64 (see, Table 8, Example I);
- MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968 (see, Table 11, Example I).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing invasive breast cancer tissue from benign breast tissue:

- CDH4_E, FLJ42875, GAD2, GRASP, ITPRIPL1, KCNA1, MAX.chr12.4273906-4274012, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MPZ, NKX2-6, PRKCB, RBFOX3_B, SALL3, and VSTM2B_A (see, Table 2, Example I).

From these 375 novel DNA methylation markers, further experiments identified the following markers and/or panels of markers capable of distinguishing ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue:

- SCRT2_B, MPZ, MAX.chr8.124173030-124173395, ITPRIPL1, ITPRIPL1, DLX4, CALN1_A, and IGF2BP3_B (see, Table 15, Example I);
- SCRT2_B, ITPRIPL1, and MAX.chr8.124173030-12417339 (100% sensitive at 91% specificity) (see, Table 15, Example I),
- DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, and ITPRIPL1 (see, Table 17, Example II).

As described herein, the technology provides a number of methylated DNA markers and subsets thereof (e.g., sets of 2, 3, 4, 5, 6, 7, or 8 markers) with high discrimination for breast cancer overall and various types of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer). Experiments applied a selection filter to candidate markers to identify markers that provide a high signal to noise ratio and a low background level to provide high specificity for purposes of breast cancer screening or diagnosis.

In some embodiments, the technology is related to assessing the presence of and methylation state of one or more of the markers identified herein in a biological sample (e.g., breast tissue, plasma sample). These markers comprise one or more differentially methylated regions (DMR) as discussed herein, e.g., as provided in Tables 2 and 18. Methylation state is assessed in embodiments of the technology. As such, the technology provided herein is not restricted in the method by which a gene's methylation state is measured. For example, in some embodiments the methylation state is measured by a genome scanning method. For example, one method involves restriction landmark genomic scanning (Kawai et al. (1994) *Mol. Cell. Biol.* 14: 7421–7427) and another example involves methylation-sensitive arbitrarily primed PCR (Gonzalgo et al. (1997) *Cancer Res.* 57: 594–599). In some embodiments, changes in methylation patterns at specific CpG sites are monitored by digestion of genomic DNA with methylation-sensitive restriction enzymes followed by Southern analysis of the regions of interest (digestion-Southern method). In some embodiments, analyzing changes in methylation patterns involves a PCR-based process that

involves digestion of genomic DNA with methylation-sensitive restriction enzymes or methylation-dependent restriction enzymes prior to PCR amplification (Singer-Sam et al. (1990) *Nucl. Acids Res.* 18: 687). In addition, other techniques have been reported that utilize bisulfite treatment of DNA as a starting point for methylation analysis. These include methylation-specific PCR (MSP) (Herman et al. (1992) *Proc. Natl. Acad. Sci. USA* 93: 9821–9826) and restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA (Sadri and Hornsby (1996) *Nucl. Acids Res.* 24: 5058–5059; and Xiong and Laird (1997) *Nucl. Acids Res.* 25: 2532–2534). PCR techniques have been developed for detection of gene mutations (Kuppuswamy et al. (1991) *Proc. Natl. Acad. Sci. USA* 88: 1143–1147) and quantification of allelic-specific expression (Szabo and Mann (1995) *Genes Dev.* 9: 3097–3108; and Singer-Sam et al. (1992) *PCR Methods Appl.* 1: 160–163). Such techniques use internal primers, which anneal to a PCR-generated template and terminate immediately 5' of the single nucleotide to be assayed. Methods using a "quantitative Ms-SNuPE assay" as described in U.S. Pat. No. 7,037,650 are used in some embodiments.

Upon evaluating a methylation state, the methylation state is often expressed as the fraction or percentage of individual strands of DNA that is methylated at a particular site (e.g., at a single nucleotide, at a particular region or locus, at a longer sequence of interest, e.g., up to a ~100-bp, 200-bp, 500-bp, 1000-bp subsequence of a DNA or longer) relative to the total population of DNA in the sample comprising that particular site. Traditionally, the amount of the unmethylated nucleic acid is determined by PCR using calibrators. Then, a known amount of DNA is bisulfite treated and the resulting methylation-specific sequence is determined using either a real-time PCR or other exponential amplification, e.g., a QuARTS assay (e.g., as provided by U.S. Pat. No. 8,361,720; and U.S. Pat. Appl. Pub. Nos. 2012/0122088 and 2012/0122106, incorporated herein by reference).

For example, in some embodiments methods comprise generating a standard curve for the unmethylated target by using external standards. The standard curve is constructed from at least two points and relates the real-time Ct value for unmethylated DNA to known quantitative standards. Then, a second standard curve for the methylated target is constructed from at least two points and external standards. This second standard curve relates the Ct for methylated DNA to known quantitative standards. Next, the test sample Ct values are determined for the methylated and unmethylated populations and the genomic equivalents of DNA are calculated from the standard curves produced by the first two steps. The percentage of methylation at the site of interest is calculated from the amount of methylated DNAs

relative to the total amount of DNAs in the population, e.g., (number of methylated DNAs) / (the number of methylated DNAs + number of unmethylated DNAs) × 100.

Also provided herein are compositions and kits for practicing the methods. For example, in some embodiments, reagents (e.g., primers, probes) specific for one or more markers are provided alone or in sets (e.g., sets of primers pairs for amplifying a plurality of markers). Additional reagents for conducting a detection assay may also be provided (e.g., enzymes, buffers, positive and negative controls for conducting QuARTS, PCR, sequencing, bisulfite, or other assays). In some embodiments, the kits contain a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent). In some embodiments, the kits containing one or more reagent necessary, sufficient, or useful for conducting a method are provided. Also provided are reactions mixtures containing the reagents. Further provided are master mix reagent sets containing a plurality of reagents that may be added to each other and/or to a test sample to complete a reaction mixture.

In some embodiments, the technology described herein is associated with a programmable machine designed to perform a sequence of arithmetic or logical operations as provided by the methods described herein. For example, some embodiments of the technology are associated with (e.g., implemented in) computer software and/or computer hardware. In one aspect, the technology relates to a computer comprising a form of memory, an element for performing arithmetic and logical operations, and a processing element (e.g., a microprocessor) for executing a series of instructions (e.g., a method as provided herein) to read, manipulate, and store data. In some embodiments, a microprocessor is part of a system for determining a methylation state (e.g., of one or more DMR, e.g., DMR 1-375 as provided in Tables 2 and 18); comparing methylation states (e.g., of one or more DMR, e.g., DMR 1-375 as provided in Tables 2 and 18); generating standard curves; determining a Ct value; calculating a fraction, frequency, or percentage of methylation (e.g., of one or more DMR, e.g., DMR 1-375 as provided in Tables 2 and 18); identifying a CpG island; determining a specificity and/or sensitivity of an assay or marker; calculating an ROC curve and an associated AUC; sequence analysis; all as described herein or is known in the art.

In some embodiments, a microprocessor or computer uses methylation state data in an algorithm to predict a site of a cancer.

In some embodiments, a software or hardware component receives the results of multiple assays and determines a single value result to report to a user that indicates a cancer

risk based on the results of the multiple assays (e.g., determining the methylation state of multiple DMR, e.g., as provided in Tables 2 and 18). Related embodiments calculate a risk factor based on a mathematical combination (e.g., a weighted combination, a linear combination) of the results from multiple assays, e.g., determining the methylation states of multiple markers (such as multiple DMR, e.g., as provided in Tables 2 and 18). In some embodiments, the methylation state of a DMR defines a dimension and may have values in a multidimensional space and the coordinate defined by the methylation states of multiple DMR is a result, e.g., to report to a user, e.g., related to a cancer risk.

Some embodiments comprise a storage medium and memory components. Memory components (e.g., volatile and/or nonvolatile memory) find use in storing instructions (e.g., an embodiment of a process as provided herein) and/or data (e.g., a work piece such as methylation measurements, sequences, and statistical descriptions associated therewith). Some embodiments relate to systems also comprising one or more of a CPU, a graphics card, and a user interface (e.g., comprising an output device such as display and an input device such as a keyboard).

Programmable machines associated with the technology comprise conventional extant technologies and technologies in development or yet to be developed (e.g., a quantum computer, a chemical computer, a DNA computer, an optical computer, a spintronics based computer, etc.).

In some embodiments, the technology comprises a wired (e.g., metallic cable, fiber optic) or wireless transmission medium for transmitting data. For example, some embodiments relate to data transmission over a network (e.g., a local area network (LAN), a wide area network (WAN), an ad-hoc network, the internet, etc.). In some embodiments, programmable machines are present on such a network as peers and in some embodiments the programmable machines have a client/server relationship.

In some embodiments, data are stored on a computer-readable storage medium such as a hard disk, flash memory, optical media, a floppy disk, etc.

In some embodiments, the technology provided herein is associated with a plurality of programmable devices that operate in concert to perform a method as described herein. For example, in some embodiments, a plurality of computers (e.g., connected by a network) may work in parallel to collect and process data, e.g., in an implementation of cluster computing or grid computing or some other distributed computer architecture that relies on complete computers (with onboard CPUs, storage, power supplies, network interfaces, etc.) connected

to a network (private, public, or the internet) by a conventional network interface, such as Ethernet, fiber optic, or by a wireless network technology.

For example, some embodiments provide a computer that includes a computer-readable medium. The embodiment includes a random access memory (RAM) coupled to a processor. The processor executes computer-executable program instructions stored in memory. Such processors may include a microprocessor, an ASIC, a state machine, or other processor, and can be any of a number of computer processors, such as processors from Intel Corporation of Santa Clara, California and Motorola Corporation of Schaumburg, Illinois. Such processors include, or may be in communication with, media, for example computer-readable media, which stores instructions that, when executed by the processor, cause the processor to perform the steps described herein.

Embodiments of computer-readable media include, but are not limited to, an electronic, optical, magnetic, or other storage or transmission device capable of providing a processor with computer-readable instructions. Other examples of suitable media include, but are not limited to, a floppy disk, CD-ROM, DVD, magnetic disk, memory chip, ROM, RAM, an ASIC, a configured processor, all optical media, all magnetic tape or other magnetic media, or any other medium from which a computer processor can read instructions. Also, various other forms of computer-readable media may transmit or carry instructions to a computer, including a router, private or public network, or other transmission device or channel, both wired and wireless. The instructions may comprise code from any suitable computer-programming language, including, for example, C, C++, C#, Visual Basic, Java, Python, Perl, and JavaScript.

Computers are connected in some embodiments to a network. Computers may also include a number of external or internal devices such as a mouse, a CD-ROM, DVD, a keyboard, a display, or other input or output devices. Examples of computers are personal computers, digital assistants, personal digital assistants, cellular phones, mobile phones, smart phones, pagers, digital tablets, laptop computers, internet appliances, and other processor-based devices. In general, the computers related to aspects of the technology provided herein may be any type of processor-based platform that operates on any operating system, such as Microsoft Windows, Linux, UNIX, Mac OS X, etc., capable of supporting one or more programs comprising the technology provided herein. Some embodiments comprise a personal computer executing other application programs (e.g., applications). The applications can be contained in memory and can include, for example, a word processing

application, a spreadsheet application, an email application, an instant messenger application, a presentation application, an Internet browser application, a calendar/organizer application, and any other application capable of being executed by a client device.

All such components, computers, and systems described herein as associated with the technology may be logical or virtual.

Accordingly, provided herein is technology related to a method of screening for breast cancer and/or various forms of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer) in a sample obtained from a subject, the method comprising assaying a methylation state of a marker in a sample obtained from a subject (e.g., breast tissue) (e.g., plasma sample) and identifying the subject as having breast cancer and/or a specific form of breast cancer when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have breast cancer, wherein the marker comprises a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1–375 as provided in Tables 2 and 18.

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has breast cancer: ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1 (see, Table 16E, Example II).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has breast cancer: ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B,

FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TRH_A, and TRIM67_B (see, Table 22, Example III).

In some embodiments wherein the sample obtained from the subject is a blood sample (e.g., plasma, serum, whole blood) and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has breast cancer: CD1D, ITPRIPL1, FAM59B, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B (see, Table 27, Example III).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has triple negative breast cancer: ABLIM1, AJAP1_B, ASCL2, ATP6V1B1, BANK1, CALN1_A, CALN1_B, CLIC6, DSCR6, FOXP4, GAD2, GCGR, GP5, GRASP, HBM, HNF1B_B, KLF16, MAGI2, MAX.chr11.14926602-14927148, MAX.chr12.4273906-4274012, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr2.97193478-97193562, MAX.chr22.42679578-42679917, MAX.chr4.8859253-8859329, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr6.157557371-157557657, MPZ, NKX2-6, PDX1, PLXNC1_A, PPARG, PRKCB, PTPRN2, RBFOX_A, SCRT2_A, SLC7A4, STAC2_B, STX16_A, STX16_B, TBX1, TRH_A, VSTM2B_A, ZBTB16, ZNF132, and ZSCAN23 (see, Table 3, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has triple negative breast cancer: CALN1_A, LOC100132891, NACAD, TRIM67, ATP6V1B1, DLX4, GP5, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ST8SIA4, STX16_B ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, KCNK9, SCRT2_B, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461,

MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, and DSCR6 (see, Table 11, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has triple negative breast cancer: ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4 (see, Table 16A, Example II).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has $HER2^+$ breast cancer: ABLIM1, AFAP1L1, AKR1B1, ALOX5, AMN, ARL5C, BANK1, BCAT1, BEGAIN, BEST4, BHLHE23_B, BHLHE23_C, C17orf64, C1QL2, C7orf52, CALN1_B, CAV2, CD8A, CDH4_A, CDH4_B, CDH4_C, CDH4_D, CDH4_E, CDH4_F, CHST2_B, CLIP4, CR1, DLK1, DNAJC6, DNM3_A, EMX1_A, ESPN, FABP5, FAM150A, FLJ42875, GLP1R, GNG4, GYPC_A, HAND2, HES5, HNF1B_A, HNF1B_B, HOXA1_A, HOXA1_B, HOXA7_A, HOXA7_B, HOXA7_C, HOXD9, IGF2BP3_A, IGF2BP3_B, IGSF9B_A, IL15RA, INSM1, ITPKA_B, ITPRIPL1, KCNE3, KCNK17_B, LIME1, LOC100132891, LOC283999, LY6H, MAST1, MAX.chr1.158083198-158083476, MAX.chr1.228074764-228074977, MAX.chr1.46913931-46913950, MAX.chr10.130085265-130085312, MAX.chr11.68622869-68622968, MAX.chr14.101176106-101176260, MAX.chr15.96889069-96889128, MAX.chr17.8230197-8230314, MAX.chr19.46379903-46380197, MAX.chr2.97193163-97193287, MAX.chr2.97193478-97193562, MAX.chr20.1784209-1784461, MAX.chr21.44782441-44782498, MAX.chr22.23908718-23908782, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.180101084-180101094, MAX.chr5.42952185-42952280, MAX.chr5.42994866-42994936, MAX.chr6.27064703-27064783, MAX.chr7.152622607-152622638, MAX.chr8.145104132-145104218, MAX.chr9.136474504-136474527, MCF2L2, MSX2P1, NACAD, NID2_B, NID2_C, ODC1, OSR2_B, PAQR6, PCDH8, PIF1, PPARA, PPP2R5C,

PRDM13_A, PRHOXNB, PRKCB, RBFOX3_A, RBFOX3_B, RFX8, SNCA, STAC2_A, STAC2_B, STX16_B SYT5, TIMP2, TMEFF2, TNFRSF10D, TRH_B, TRIM67, TRIM71_C, USP44_A, USP44_B, UTF1, UTS2R, VSTM2B_A, VSTM2B_B, ZFP64, and ZNF132 (see, Table 4, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has $HER2^+$ breast cancer: BHLHE23_C, CALN1_A, CD1D, CHST2_A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, NACAD, TRIM67, ATP6V1B1, C17orf64, CHST2_B, DLX4, DNM3_A, EMX1_A, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, COL23A1, KCNK9, LAYN, PLXNC1_A, RIC3, SCRT2_B, ALOX5, CDH4_E, HNF1B_B, TRH_A, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, AJAP1_B, DSCR6, and MAX.chr11.68622869-68622968 (see, Table 11, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has $HER2^+$ breast cancer: ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, C10orf125 (see, Table 16B, Example II).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has Luminal A breast cancer: ARL5C, BHLHE23_C, BMP6, C10orf125, C17orf64, C19orf66, CAMKV, CD1D, CDH4_E, CDH4_F, CHST2_A, CRHBP, DLX6, DNM3_A, DNM3_B, DNM3_C, ESYT3, ETS1_A, ETS1_B, FAM126A, FAM189A1, FAM20A, FAM59B, FBN1, FLRT2, FMN2, FOXP4, GAS7, GYPC_A, GYPC_B, HAND2, HES5, HMGA2, HNF1B_B, IGF2BP3_A, IGF2BP3_B, KCNH8, KCNK17_A, KCNQ2, KLHDC7B, LOC100132891, MAX.chr1.46913931-46913950, MAX.chr11.68622869-68622968, MAX.chr12.4273906-4274012, MAX.chr12.59990591-59990895, MAX.chr17.73073682-73073814, MAX.chr20.1783841-1784054, MAX.chr21.47063802-47063851, MAX.chr4.8860002-8860038, MAX.chr5.172234248-172234494, MAX.chr5.178957564-178957598, MAX.chr6.130686865-130686985, MAX.chr8.687688-687736, MAX.chr8.688863-688924, MAX.chr9.114010-114207, MPZ, NID2_A, NKX2-6, ODC1, OSR2_A, POU4F1, PRDM13_B, PRKCB, RASGRF2, RIPPLY2, SLC30A10, ST8SIA4, SYN2, TRIM71_A, TRIM71_B, TRIM71_C, UBTF, ULBP1, USP44_B, and VSTM2B_A (see, Table 5, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has Luminal A breast cancer: BHLHE23_C, CD1D, CHST2_A, FAM126A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, SLC30A10, TRIM67, ATP6V1B1, BANK1, C10orf125, C17orf64, CHST2_B, DNM3_A, EMX1_A, GP5, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, ST8SIA4, STX16_B UBTF, LOC100132891, ITPRIPL1, MAX.chr12.4273906-4274012, MAX.chr12.59990671-59990859, BHLHE23_D, COL23A1, KCNK9, OTX1, PLXNC1_A, HNF1B_B, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, DSCR6, MAX.chr11.68622869-68622968 (see, Table 11, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has Luminal A breast cancer: ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125 (see, Table 16C, Example II).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has Luminal B breast cancer: ACCN1, AJAP1_A, AJAP1_B, BEST4, CALN1_B, CBLN1_B, CDH4_E, DLX4, FOXP4, IGSF9B_B, ITPRIPL1, KCNA1, KLF16, LMX1B_A, MAST1, MAX.chr11.14926602-14927148, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr22.42679578-42679917, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr8.124173128-124173268, MPZ, PPARA, PRMT1, RBFOX3_B, RYR2_A, SALL3, SCRT2_A, SPHK2, STX16_B SYNJ2, TMEM176A, TSHZ3, and VIPR2 (see, Table 6, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has Luminal B breast cancer: CALN1_A, LOC100132891, MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, DLX4, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, BHLHE23_D, HNF1B_B, TRH_A, ASCL2, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-

145725459, MAX.chr5.77268672-77268725, BEST4, AJAP1_B, and DSCR6 (see, Table 11, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has Luminal B breast cancer: ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D (see, Table 16D, Example II).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has BRCA1 breast cancer: C10orf93, C20orf195_A, C20orf195_B, CALN1_B, CBLN1_A, CBLN1_B, CCDC61, CCND2_A, CCND2_B, CCND2_C, EMX1_B, FAM150B, GRASP, HBM, ITPRIPL1, KCNK17_A, KIAA1949, LOC100131176, MAST1, MAX.chr1.8277285-8277316, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr18.5629721-5629791, MAX.chr19.30719261-30719354, MAX.chr22.42679767-42679917, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr6.157556793-157556856, MAX.chr8.124173030-124173395, MN1, MPZ, NR2F6, PDXK_A, PDXK_B, PTPRM, RYR2_B, SERPINB9_A, SERPINB9_B, SLC8A3, STX16_B TEPP, TOX, VIPR2, VSTM2B_A, ZNF486, ZNF626, and ZNF671 (see, Table 7, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has BRCA1 breast cancer: BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-

96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6 (see, Table 11, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has BRCA2 breast cancer: ANTXR2, B3GNT5, BHLHE23_C, BMP4, CHRNA7, EPHA4, FAM171A1, FAM20A, FMNL2, FSCN1, GSTP1, HBM, IGFBP5, IL17REL, ITGA9, ITPRIPL1, KIRREL2, LRRC34, MAX.chr1.239549742-239549886, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr2.238864674-238864735, MAX.chr5.81148300-81148332, MAX.chr7.151145632-151145743, MAX.chr8.124173030-124173395, MAX.chr8.143533298-143533558, MERTK, MPZ, NID2_C, NTRK3, OLIG3_A, OLIG3_B, OSR2_C, PROM1, RGS17, SBNO2, STX16_B TBKBP1, TLX1NB, VIPR2, VN1R2, VSNL1, and ZFP64 (see, Table 8, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has BRCA2 breast cancer: MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968 (see, Table 11, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer indicates the subject has invasive breast cancer: CDH4_E, FLJ42875, GAD2, GRASP, ITPRIPL1, KCNA1, MAX.chr12.4273906-4274012, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr4.8859602-

8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MPZ, NKX2-6, PRKCB, RBFOX3_B, SALL3, and VSTM2B_A (see, Table 9, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer distinguishes between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue: SCRT2_B, MPZ, MAX.chr8.124173030-124173395, ITPRIPL1, ITPRIPL1, DLX4, CALN1_A, and IGF2BP3_B (see, Table 15, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer distinguishes between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue: SCRT2_B, ITPRIPL1, and MAX.chr8.124173030-12417339 (100% sensitive at 91% specificity) (see, Table 15, Example I).

In some embodiments wherein the sample obtained from the subject is breast tissue and the methylation state of one or more of the following markers is different than a methylation state of the one or more markers assayed in a subject that does not have breast cancer distinguishes between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue: DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, and ITPRIPL1 (see, Table 17, Example II).

The technology is related to identifying and discriminating breast cancer and/or various forms of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer). Some embodiments provide methods comprising assaying a plurality of markers, e.g., comprising assaying 2 to 11 to 100 or 120 or 375 markers.

The technology is not limited in the methylation state assessed. In some embodiments assessing the methylation state of the marker in the sample comprises determining the methylation state of one base. In some embodiments, assaying the methylation state of the

marker in the sample comprises determining the extent of methylation at a plurality of bases. Moreover, in some embodiments the methylation state of the marker comprises an increased methylation of the marker relative to a normal methylation state of the marker. In some embodiments, the methylation state of the marker comprises a decreased methylation of the marker relative to a normal methylation state of the marker. In some embodiments the methylation state of the marker comprises a different pattern of methylation of the marker relative to a normal methylation state of the marker.

Furthermore, in some embodiments the marker is a region of 100 or fewer bases, the marker is a region of 500 or fewer bases, the marker is a region of 1000 or fewer bases, the marker is a region of 5000 or fewer bases, or, in some embodiments, the marker is one base. In some embodiments the marker is in a high CpG density promoter.

The technology is not limited by sample type. For example, in some embodiments the sample is a stool sample, a tissue sample (e.g., breast tissue sample), a blood sample (e.g., plasma, serum, whole blood), an excretion, or a urine sample.

Furthermore, the technology is not limited in the method used to determine methylation state. In some embodiments the assaying comprises using methylation specific polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation specific nuclease, mass-based separation, or target capture. In some embodiments, the assaying comprises use of a methylation specific oligonucleotide. In some embodiments, the technology uses massively parallel sequencing (e.g., next-generation sequencing) to determine methylation state, e.g., sequencing-by-synthesis, real-time (e.g., single-molecule) sequencing, bead emulsion sequencing, nanopore sequencing, etc.

The technology provides reagents for detecting a DMR, e.g., in some embodiments are provided a set of oligonucleotides comprising the sequences provided by SEQ ID NO: 1–422 (see, Tables 10, 19 and 20). In some embodiments are provided an oligonucleotide comprising a sequence complementary to a chromosomal region having a base in a DMR, e.g., an oligonucleotide sensitive to methylation state of a DMR.

The technology provides various panels of markers use for identifying breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968,

MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1 (see, Table 16E, Example II).

The technology provides various panels of markers use for identifying breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B, FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TRH_A, and TRIM67_B (see, Table 22, Example III).

The technology provides various panels of markers use for identifying breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is CD1D, ITPRIPL1, FAM59B, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B (see, Table 27, Example III).

The technology provides various panels of markers use for identifying triple negative breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ABLIM1, AJAP1_B, ASCL2, ATP6V1B1, BANK1, CALN1_A, CALN1_B, CLIC6, DSCR6, FOXP4, GAD2, GCGR, GP5, GRASP, HBM, HNF1B_B, KLF16, MAGI2, MAX.chr11.14926602-14927148, MAX.chr12.4273906-4274012, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr2.97193478-97193562, MAX.chr22.42679578-42679917, MAX.chr4.8859253-8859329, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr6.157557371-157557657, MPZ, NKX2-6, PDX1, PLXNC1_A, PPARG, PRKCB, PTPRN2, RBFOX_A, SCRT2_A, SLC7A4, STAC2_B, STX16_A, STX16_B, TBX1, TRH_A, VSTM2B_A, ZBTB16, ZNF132, and ZSCAN23 (see, Table 3, Example I).

The technology provides various panels of markers use for identifying triple negative breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is CALN1_A, LOC100132891, NACAD, TRIM67, ATP6V1B1, DLX4, GP5, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ST8SIA4, STX16_B ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, KCNK9, SCRT2_B, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, and DSCR6 (see, Table 11, Example I).

The technology provides various panels of markers use for identifying triple negative breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4 (see, Table 16A, Example II).

The technology provides various panels of markers use for identifying $HER2^+$ breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ABLIM1, AFAP1L1, AKR1B1, ALOX5, AMN, ARL5C, BANK1, BCAT1, BEGAIN, BEST4, BHLHE23_B, BHLHE23_C, C17orf64, C1QL2, C7orf52, CALN1_B, CAV2, CD8A, CDH4_A, CDH4_B, CDH4_C, CDH4_D, CDH4_E, CDH4_F, CHST2_B, CLIP4, CR1, DLK1, DNAJC6, DNM3_A, EMX1_A, ESPN, FABP5, FAM150A, FLJ42875, GLP1R, GNG4, GYPC_A, HAND2, HES5, HNF1B_A, HNF1B_B, HOXA1_A, HOXA1_B, HOXA7_A, HOXA7_B, HOXA7_C, HOXD9, IGF2BP3_A, IGF2BP3_B, IGSF9B_A, IL15RA, INSM1, ITPKA_B, ITPRIPL1, KCNE3, KCNK17_B, LIME1, LOC100132891, LOC283999, LY6H, MAST1, MAX.chr1.158083198-158083476, MAX.chr1.228074764-228074977, MAX.chr1.46913931-46913950, MAX.chr10.130085265-130085312, MAX.chr11.68622869-68622968, MAX.chr14.101176106-101176260, MAX.chr15.96889069-96889128, MAX.chr17.8230197-8230314, MAX.chr19.46379903-46380197, MAX.chr2.97193163-97193287, MAX.chr2.97193478-97193562, MAX.chr20.1784209-1784461, MAX.chr21.44782441-44782498, MAX.chr22.23908718-23908782, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.180101084-180101094, MAX.chr5.42952185-42952280, MAX.chr5.42994866-42994936, MAX.chr6.27064703-

27064783, MAX.chr7.152622607-152622638, MAX.chr8.145104132-145104218, MAX.chr9.136474504-136474527, MCF2L2, MSX2P1, NACAD, NID2_B, NID2_C, ODC1, OSR2_B, PAQR6, PCDH8, PIF1, PPARA, PPP2R5C, PRDM13_A, PRHOXNB, PRKCB, RBFOX3_A, RBFOX3_B, RFX8, SNCA, STAC2_A, STAC2_B, STX16_B SYT5, TIMP2, TMEFF2, TNFRSF10D, TRH_B, TRIM67, TRIM71_C, USP44_A, USP44_B, UTF1, UTS2R, VSTM2B_A, VSTM2B_B, ZFP64, and ZNF132 (see, Table 4, Example I).

The technology provides various panels of markers use for identifying $HER2^+$ breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is BHLHE23_C, CALN1_A, CD1D, CHST2_A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, NACAD, TRIM67, ATP6V1B1, C17orf64, CHST2_B, DLX4, DNM3_A, EMX1_A, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, COL23A1, KCNK9, LAYN, PLXNC1_A, RIC3, SCRT2_B, ALOX5, CDH4_E, HNF1B_B, TRH_A, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, AJAP1_B, DSCR6, and MAX.chr11.68622869-68622968 (see, Table 11, Example I).

The technology provides various panels of markers use for identifying $HER2^+$ breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, C10orf125 (see, Table 16B, Example II).

The technology provides various panels of markers use for identifying Luminal A breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having

an annotation that is ARL5C, BHLHE23_C, BMP6, C10orf125, C17orf64, C19orf66, CAMKV, CD1D, CDH4_E, CDH4_F, CHST2_A, CRHBP, DLX6, DNM3_A, DNM3_B, DNM3_C, ESYT3, ETS1_A, ETS1_B, FAM126A, FAM189A1, FAM20A, FAM59B, FBN1, FLRT2, FMN2, FOXP4, GAS7, GYPC_A, GYPC_B, HAND2, HES5, HMGA2, HNF1B_B, IGF2BP3_A, IGF2BP3_B, KCNH8, KCNK17_A, KCNQ2, KLHDC7B, LOC100132891, MAX.chr1.46913931-46913950, MAX.chr11.68622869-68622968, MAX.chr12.4273906-4274012, MAX.chr12.59990591-59990895, MAX.chr17.73073682-73073814, MAX.chr20.1783841-1784054, MAX.chr21.47063802-47063851, MAX.chr4.8860002-8860038, MAX.chr5.172234248-172234494, MAX.chr5.178957564-178957598, MAX.chr6.130686865-130686985, MAX.chr8.687688-687736, MAX.chr8.688863-688924, MAX.chr9.114010-114207, MPZ, NID2_A, NKX2-6, ODC1, OSR2_A, POU4F1, PRDM13_B, PRKCB, RASGRF2, RIPPLY2, SLC30A10, ST8SIA4, SYN2, TRIM71_A, TRIM71_B, TRIM71_C, UBTF, ULBP1, USP44_B, and VSTM2B_A (see, Table 5, Example I).

The technology provides various panels of markers use for identifying Luminal A breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is BHLHE23_C, CD1D, CHST2_A, FAM126A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, SLC30A10, TRIM67, ATP6V1B1, BANK1, C10orf125, C17orf64, CHST2_B, DNM3_A, EMX1_A, GP5, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, ST8SIA4, STX16_B UBTF, LOC100132891, ITPRIPL1, MAX.chr12.4273906-4274012, MAX.chr12.59990671-59990859, BHLHE23_D, COL23A1, KCNK9, OTX1, PLXNC1_A, HNF1B_B, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, DSCR6, MAX.chr11.68622869-68622968 (see, Table 11, Example I).

The technology provides various panels of markers use for identifying Luminal A breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B,

MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125 (see, Table 16C, Example II).

The technology provides various panels of markers use for identifying Luminal B breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ACCN1, AJAP1_A, AJAP1_B, BEST4, CALN1_B, CBLN1_B, CDH4_E, DLX4, FOXP4, IGSF9B_B, ITPRIPL1, KCNA1, KLF16, LMX1B_A, MAST1, MAX.chr11.14926602-14927148, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr22.42679578-42679917, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr8.124173128-124173268, MPZ, PPARA, PRMT1, RBFOX3_B, RYR2_A, SALL3, SCRT2_A, SPHK2, STX16_B SYNJ2, TMEM176A, TSHZ3, and VIPR2 (see, Table 6, Example I).

The technology provides various panels of markers use for identifying Luminal B breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is CALN1_A, LOC100132891, MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, DLX4, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, BHLHE23_D, HNF1B_B, TRH_A, ASCL2, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, AJAP1_B, and DSCR6 (see, Table 11, Example I).

The technology provides various panels of markers use for identifying Luminal B breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5,

MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D (see, Table 16D, Example II).

The technology provides various panels of markers use for identifying BRCA1 breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is C10orf93, C20orf195_A, C20orf195_B, CALN1_B, CBLN1_A, CBLN1_B, CCDC61, CCND2_A, CCND2_B, CCND2_C, EMX1_B, FAM150B, GRASP, HBM, ITPRIPL1, KCNK17_A, KIAA1949, LOC100131176, MAST1, MAX.chr1.8277285-8277316, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr18.5629721-5629791, MAX.chr19.30719261-30719354, MAX.chr22.42679767-42679917, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr6.157556793-157556856, MAX.chr8.124173030-124173395, MN1, MPZ, NR2F6, PDXK_A, PDXK_B, PTPRM, RYR2_B, SERPINB9_A, SERPINB9_B, SLC8A3, STX16_B TEPP, TOX, VIPR2, VSTM2B_A, ZNF486, ZNF626, and ZNF671 (see, Table 7, Example I).

The technology provides various panels of markers use for identifying BRCA1 breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6 (see, Table 11, Example I).

The technology provides various panels of markers use for identifying BRCA2 breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is ANTXR2, B3GNT5, BHLHE23_C, BMP4, CHRNA7, EPHA4, FAM171A1, FAM20A, FMNL2, FSCN1, GSTP1, HBM, IGFBP5, IL17REL, ITGA9, ITPRIPL1, KIRREL2, LRRC34, MAX.chr1.239549742-239549886, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr2.238864674-238864735, MAX.chr5.81148300-

81148332, MAX.chr7.151145632-151145743, MAX.chr8.124173030-124173395, MAX.chr8.143533298-143533558, MERTK, MPZ, NID2_C, NTRK3, OLIG3_A, OLIG3_B, OSR2_C, PROM1, RGS17, SBNO2, STX16_B TBKBP1, TLX1NB, VIPR2, VN1R2, VSNL1, and ZFP64 (see, Table 8, Example I).

The technology provides various panels of markers use for identifying BRCA2 breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968 (see, Table 11, Example I).

The technology provides various panels of markers use for identifying invasive breast cancer, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is CDH4_E, FLJ42875, GAD2, GRASP, ITPRIPL1, KCNA1, MAX.chr12.4273906-4274012, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MPZ, NKX2-6, PRKCB, RBFOX3_B, SALL3, and VSTM2B_A (see, Table 9, Example I).

The technology provides various panels of markers use for distinguishing between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is SCRT2_B, MPZ, MAX.chr8.124173030-124173395, ITPRIPL1, ITPRIPL1, DLX4, CALN1_A, and IGF2BP3_B (see, Table 15, Example I).

The technology provides various panels of markers use for distinguishing between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is SCRT2_B, ITPRIPL1, and MAX.chr8.124173030-12417339 (100% sensitive at 91% specificity) (see, Table 15, Example I).

The technology provides various panels of markers use for distinguishing between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in

situ low grade (DCIS-LG) breast tissue, e.g., in some embodiments the marker comprises a chromosomal region having an annotation that is DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, and ITPRIPL1 (see, Table 17, Example II).

Kit embodiments are provided, e.g., a kit comprising a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent); and a control nucleic acid comprising a sequence from a DMR selected from a group consisting of DMR 1–375 (from Tables 2 and 18) and having a methylation state associated with a subject who does not have breast cancer. In some embodiments, kits comprise a bisulfite reagent and an oligonucleotide as described herein. In some embodiments, kits comprise a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent); and a control nucleic acid comprising a sequence from a DMR selected from a group consisting of DMR 1–375 (from Tables 2 and 18) and having a methylation state associated with a subject who has breast cancer. Some kit embodiments comprise a sample collector for obtaining a sample from a subject (e.g., a stool sample; breast tissue sample; plasma sample, serum sample, whole blood sample); a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent); and an oligonucleotide as described herein.

The technology is related to embodiments of compositions (e.g., reaction mixtures). In some embodiments are provided a composition comprising a nucleic acid comprising a DMR and a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent). Some embodiments provide a composition comprising a nucleic acid comprising a DMR and an oligonucleotide as described herein. Some embodiments provide a composition comprising a nucleic acid comprising a DMR and a methylation-sensitive restriction enzyme. Some embodiments provide a composition comprising a nucleic acid comprising a DMR and a polymerase.

Additional related method embodiments are provided for screening for breast cancer and/or various forms of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2

breast cancer) in a sample obtained from a subject (e.g., breast tissue sample; plasma sample; stool sample), e.g., a method comprising determining a methylation state of a marker in the sample comprising a base in a DMR that is one or more of DMR 1–375 (from Tables 2 and 18); comparing the methylation state of the marker from the subject sample to a methylation state of the marker from a normal control sample from a subject who does not have breast cancer (e.g., breast cancer and/or a form of breast cancer: triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer); and determining a confidence interval and/or a p value of the difference in the methylation state of the subject sample and the normal control sample. In some embodiments, the confidence interval is 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% or 99.99% and the p value is 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, or 0.0001. Some embodiments of methods provide steps of reacting a nucleic acid comprising a DMR with a reagent capable of modifying nucleic acid in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) to produce, for example, nucleic acid modified in a methylation-specific manner; sequencing the nucleic acid modified in a methylation-specific manner to provide a nucleotide sequence of the nucleic acid modified in a methylation-specific manner; comparing the nucleotide sequence of the nucleic acid modified in a methylation-specific manner with a nucleotide sequence of a nucleic acid comprising the DMRfrom a subject who does not have breast cancer and/or a form of breast cancer to identify differences in the two sequences; and identifying the subject as having breast cancer (e.g., breast cancer and/or a form of breast cancer: triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer) when a difference is present.

Systems for screening for breast cancer in a sample obtained from a subject are provided by the technology. Exemplary embodiments of systems include, e.g., a system for screening for breast cancer and/or types of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer) in a sample obtained from a subject (e.g., breast tissue sample; plasma sample; stool sample), the system comprising an analysis component configured to determine the methylation state of a sample, a software component configured to compare the methylation state of the sample with a control sample or a reference sample methylation state recorded in a database, and an alert component configured to alert a user of a breast-cancer-

associated methylation state. An alert is determined in some embodiments by a software component that receives the results from multiple assays (e.g., determining the methylation states of multiple markers, e.g., DMR, e.g., as provided in Tables 2 and 18) and calculating a value or result to report based on the multiple results. Some embodiments provide a database of weighted parameters associated with each DMR provided herein for use in calculating a value or result and/or an alert to report to a user (e.g., such as a physician, nurse, clinician, etc.). In some embodiments all results from multiple assays are reported and in some embodiments one or more results are used to provide a score, value, or result based on a composite of one or more results from multiple assays that is indicative of a cancer risk in a subject.

In some embodiments of systems, a sample comprises a nucleic acid comprising a DMR. In some embodiments the system further comprises a component for isolating a nucleic acid, a component for collecting a sample such as a component for collecting a stool sample. In some embodiments, the system comprises nucleic acid sequences comprising a DMR. In some embodiments the database comprises nucleic acid sequences from subjects who do not have breast cancer and/or specific types of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer). Also provided are nucleic acids, e.g., a set of nucleic acids, each nucleic acid having a sequence comprising a DMR. In some embodiments the set of nucleic acids wherein each nucleic acid has a sequence from a subject who does not have breast cancer and/or specific types of breast cancer. Related system embodiments comprise a set of nucleic acids as described and a database of nucleic acid sequences associated with the set of nucleic acids. Some embodiments further comprise a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent). And, some embodiments further comprise a nucleic acid sequencer.

In certain embodiments, methods for characterizing a sample (e.g., breast tissue sample; plasma sample; whole blood sample; serum sample; stool sample) from a human patient are provided. For example, in some embodiments such embodiments comprise obtaining DNA from a sample of a human patient; assaying a methylation state of a DNA methylation marker comprising a base in a differentially methylated region (DMR) selected from a group consisting of DMR 1–375 from Tables 2 and 18; and comparing the assayed methylation state of the one or more DNA methylation markers with methylation level

references for the one or more DNA methylation markers for human patients not having breast cancer and/or specific types of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer).

Such methods are not limited to a particular type of sample from a human patient. In some embodiments, the sample is a breast tissue sample. In some embodiments, the sample is a plasma sample. In some embodiments, the sample is a stool sample, a tissue sample, a breast tissue sample, a blood sample (e.g., plasma sample, whole blood sample, serum sample), or a urine sample.

In some embodiments, such methods comprise assaying a plurality of DNA methylation markers. In some embodiments, such methods comprise assaying 2 to 11 DNA methylation markers. In some embodiments, such methods comprise assaying 12 to 120 DNA methylation markers. In some embodiments, such methods comprise assaying 2 to 375 DNA methylation markers. In some embodiments, such methods comprise assaying the methylation state of the one or more DNA methylation markers in the sample comprises determining the methylation state of one base. In some embodiments, such methods comprise assaying the methylation state of the one or more DNA methylation markers in the sample comprises determining the extent of methylation at a plurality of bases. In some embodiments, such methods comprise assaying a methylation state of a forward strand or assaying a methylation state of a reverse strand.

In some embodiments, the DNA methylation marker is a region of 100 or fewer bases. In some embodiments, the DNA methylation marker is a region of 500 or fewer bases. In some embodiments, the DNA methylation marker is a region of 1000 or fewer bases. In some embodiments, the DNA methylation marker is a region of 5000 or fewer bases. In some embodiments, the DNA methylation marker is one base. In some embodiments, the DNA methylation marker is in a high CpG density promoter.

In some embodiments, the assaying comprises using methylation specific polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation specific nuclease, mass-based separation, or target capture.

In some embodiments, the assaying comprises use of a methylation specific oligonucleotide. In some embodiments, the methylation specific oligonucleotide is selected from the group consisting of SEQ ID NO: 1–422 (Tables 10, 19 and 20).

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1 (see, Table 16E, Example II) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B, FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TRH_A, and TRIM67_B (see, Table 22, Example III) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of CD1D, ITPRIPL1, FAM59B, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B (see, Table 27, Example III) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ABLIM1, AJAP1_B, ASCL2, ATP6V1B1, BANK1, CALN1_A, CALN1_B, CLIC6, DSCR6, FOXP4, GAD2, GCGR, GP5, GRASP, HBM, HNF1B_B, KLF16, MAGI2, MAX.chr11.14926602-14927148, MAX.chr12.4273906-4274012, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr2.97193478-97193562, MAX.chr22.42679578-42679917, MAX.chr4.8859253-8859329, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr6.157557371-157557657, MPZ, NKX2-6, PDX1, PLXNC1_A, PPARG, PRKCB, PTPRN2, RBFOX_A, SCRT2_A, SLC7A4, STAC2_B, STX16_A,

STX16_B, TBX1, TRH_A, VSTM2B_A, ZBTB16, ZNF132, and ZSCAN23 (see, Table 3, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of CALN1_A, LOC100132891, NACAD, TRIM67, ATP6V1B1, DLX4, GP5, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ST8SIA4, STX16_B ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, KCNK9, SCRT2_B, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, and DSCR6 (see, Table 11, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4 (see, Table 16A, Example II) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ABLIM1, AFAP1L1, AKR1B1, ALOX5, AMN, ARL5C, BANK1, BCAT1, BEGAIN, BEST4, BHLHE23_B, BHLHE23_C, C17orf64, C1QL2, C7orf52, CALN1_B, CAV2, CD8A, CDH4_A, CDH4_B, CDH4_C, CDH4_D, CDH4_E, CDH4_F, CHST2_B, CLIP4, CR1, DLK1, DNAJC6, DNM3_A, EMX1_A, ESPN, FABP5, FAM150A, FLJ42875, GLP1R, GNG4, GYPC_A, HAND2, HES5, HNF1B_A, HNF1B_B, HOXA1_A, HOXA1_B, HOXA7_A, HOXA7_B, HOXA7_C, HOXD9, IGF2BP3_A, IGF2BP3_B, IGSF9B_A, IL15RA, INSM1, ITPKA_B, ITPRIPL1, KCNE3, KCNK17_B, LIME1, LOC100132891, LOC283999, LY6H, MAST1, MAX.chr1.158083198-158083476, MAX.chr1.228074764-228074977, MAX.chr1.46913931-46913950, MAX.chr10.130085265-130085312, MAX.chr11.68622869-68622968, MAX.chr14.101176106-101176260, MAX.chr15.96889069-96889128, MAX.chr17.8230197-8230314, MAX.chr19.46379903-46380197, MAX.chr2.97193163-97193287, MAX.chr2.97193478-97193562, MAX.chr20.1784209-1784461, MAX.chr21.44782441-44782498, MAX.chr22.23908718-23908782, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.180101084-180101094, MAX.chr5.42952185-42952280, MAX.chr5.42994866-42994936, MAX.chr6.27064703-

27064783, MAX.chr7.152622607-152622638, MAX.chr8.145104132-145104218, MAX.chr9.136474504-136474527, MCF2L2, MSX2P1, NACAD, NID2_B, NID2_C, ODC1, OSR2_B, PAQR6, PCDH8, PIF1, PPARA, PPP2R5C, PRDM13_A, PRHOXNB, PRKCB, RBFOX3_A, RBFOX3_B, RFX8, SNCA, STAC2_A, STAC2_B, STX16_B SYT5, TIMP2, TMEFF2, TNFRSF10D, TRH_B, TRIM67, TRIM71_C, USP44_A, USP44_B, UTF1, UTS2R, VSTM2B_A, VSTM2B_B, ZFP64, and ZNF132 (see, Table 4, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of BHLHE23_C, CALN1_A, CD1D, CHST2_A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, NACAD, TRIM67, ATP6V1B1, C17orf64, CHST2_B, DLX4, DNM3_A, EMX1_A, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, COL23A1, KCNK9, LAYN, PLXNC1_A, RIC3, SCRT2_B, ALOX5, CDH4_E, HNF1B_B, TRH_A, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, AJAP1_B, DSCR6, and MAX.chr11.68622869-68622968 (see, Table 11, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, C10orf125 (see, Table 16B, Example II) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ARL5C, BHLHE23_C, BMP6, C10orf125, C17orf64, C19orf66,

CAMKV, CD1D, CDH4_E, CDH4_F, CHST2_A, CRHBP, DLX6, DNM3_A, DNM3_B, DNM3_C, ESYT3, ETS1_A, ETS1_B, FAM126A, FAM189A1, FAM20A, FAM59B, FBN1, FLRT2, FMN2, FOXP4, GAS7, GYPC_A, GYPC_B, HAND2, HES5, HMGA2, HNF1B_B, IGF2BP3_A, IGF2BP3_B, KCNH8, KCNK17_A, KCNQ2, KLHDC7B, LOC100132891, MAX.chr1.46913931-46913950, MAX.chr11.68622869-68622968, MAX.chr12.4273906-4274012, MAX.chr12.59990591-59990895, MAX.chr17.73073682-73073814, MAX.chr20.1783841-1784054, MAX.chr21.47063802-47063851, MAX.chr4.8860002-8860038, MAX.chr5.172234248-172234494, MAX.chr5.178957564-178957598, MAX.chr6.130686865-130686985, MAX.chr8.687688-687736, MAX.chr8.688863-688924, MAX.chr9.114010-114207, MPZ, NID2_A, NKX2-6, ODC1, OSR2_A, POU4F1, PRDM13_B, PRKCB, RASGRF2, RIPPLY2, SLC30A10, ST8SIA4, SYN2, TRIM71_A, TRIM71_B, TRIM71_C, UBTF, ULBP1, USP44_B, and VSTM2B_A (see, Table 5, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of BHLHE23_C, CD1D, CHST2_A, FAM126A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, SLC30A10, TRIM67, ATP6V1B1, BANK1, C10orf125, C17orf64, CHST2_B, DNM3_A, EMX1_A, GP5, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, ST8SIA4, STX16_B UBTF, LOC100132891, ITPRIPL1, MAX.chr12.4273906-4274012, MAX.chr12.59990671-59990859, BHLHE23_D, COL23A1, KCNK9, OTX1, PLXNC1_A, HNF1B_B, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, DSCR6, MAX.chr11.68622869-68622968 (see, Table 11, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891,

BHLHE23_D, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125 (see, Table 16C, Example II) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ACCN1, AJAP1_A, AJAP1_B, BEST4, CALN1_B, CBLN1_B, CDH4_E, DLX4, FOXP4, IGSF9B_B, ITPRIPL1, KCNA1, KLF16, LMX1B_A, MAST1, MAX.chr11.14926602-14927148, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr22.42679578-42679917, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr8.124173128-124173268, MPZ, PPARA, PRMT1, RBFOX3_B, RYR2_A, SALL3, SCRT2_A, SPHK2, STX16_B SYNJ2, TMEM176A, TSHZ3, and VIPR2 (see, Table 6, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of CALN1_A, LOC100132891, MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, DLX4, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, BHLHE23_D, HNF1B_B, TRH_A, ASCL2, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, AJAP1_B, and DSCR6 (see, Table 11, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D (see, Table 16D, Example II) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting C10orf93, C20orf195_A, C20orf195_B, CALN1_B, CBLN1_A, CBLN1_B, CCDC61, CCND2_A, CCND2_B, CCND2_C, EMX1_B, FAM150B, GRASP, HBM, ITPRIPL1, KCNK17_A, KIAA1949, LOC100131176, MAST1, MAX.chr1.8277285-8277316, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr18.5629721-5629791, MAX.chr19.30719261-30719354, MAX.chr22.42679767-42679917, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr6.157556793-157556856, MAX.chr8.124173030-124173395, MN1, MPZ, NR2F6, PDXK_A, PDXK_B, PTPRM, RYR2_B, SERPINB9_A, SERPINB9_B, SLC8A3, STX16_B TEPP, TOX, VIPR2, VSTM2B_A, ZNF486, ZNF626, and ZNF671 (see, Table 7, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6 (see, Table 11, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of ANTXR2, B3GNT5, BHLHE23_C, BMP4, CHRNA7, EPHA4, FAM171A1, FAM20A, FMNL2, FSCN1, GSTP1, HBM, IGFBP5, IL17REL, ITGA9, ITPRIPL1, KIRREL2, LRRC34, MAX.chr1.239549742-239549886, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr2.238864674-238864735, MAX.chr5.81148300-81148332, MAX.chr7.151145632-151145743, MAX.chr8.124173030-124173395, MAX.chr8.143533298-143533558, MERTK, MPZ, NID2_C, NTRK3, OLIG3_A, OLIG3_B, OSR2_C, PROM1, RGS17, SBNO2, STX16_B TBKBP1, TLX1NB, VIPR2, VN1R2, VSNL1, and ZFP64 (see, Table 8, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968 (see, Table 11, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of CDH4_E, FLJ42875, GAD2, GRASP, ITPRIPL1, KCNA1, MAX.chr12.4273906-4274012, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MPZ, NKX2-6, PRKCB, RBFOX3_B, SALL3, and VSTM2B_A (see, Table 9, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of SCRT2_B, MPZ, MAX.chr8.124173030-124173395, ITPRIPL1, ITPRIPL1, DLX4, CALN1_A, and IGF2BP3_B (see, Table 15, Example I) comprises the DNA methylation marker.

In some embodiments, a chromosomal region having an annotation selected from the group consisting of DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, and ITPRIPL1 (see, Table 17, Example II) comprises the DNA methylation marker.

In some embodiments, such methods comprise determining the methylation state of two DNA methylation markers. In some embodiments, such methods comprise determining the methylation state of a pair of DNA methylation markers provided in a row of Tables 2 and/or 18.

In certain embodiments, the technology provides methods for characterizing a sample (e.g., breast tissue sample; plasma sample; whole blood sample; serum sample; stool sample) obtained from a human patient. In some embodiments, such methods comprise determining a methylation state of a DNA methylation marker in the sample comprising a base in a DMR selected from a group consisting of DMR 1–375 from Tables 2 and 18; comparing the methylation state of the DNA methylation marker from the patient sample to a methylation state of the DNA methylation marker from a normal control sample from a human subject

who does not have a breast cancer and/or a specific form of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer); and determining a confidence interval and/or a *p* value of the difference in the methylation state of the human patient and the normal control sample. In some embodiments, the confidence interval is 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% or 99.99% and the *p* value is 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, or 0.0001.

In certain embodiments, the technology provides methods for characterizing a sample obtained from a human subject (e.g., breast tissue sample; plasma sample; whole blood sample; serum sample; stool sample), the method comprising reacting a nucleic acid comprising a DMR with a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) to produce nucleic acid modified in a methylation-specific manner; sequencing the nucleic acid modified in a methylation-specific manner to provide a nucleotide sequence of the nucleic acid modified in a methylation-specific manner; comparing the nucleotide sequence of the nucleic acid modified in a methylation-specific manner with a nucleotide sequence of a nucleic acid comprising the DMR from a subject who does not have breast cancer to identify differences in the two sequences.

In certain embodiments, the technology provides systems for characterizing a sample obtained from a human subject (e.g., breast tissue sample; plasma sample; stool sample), the system comprising an analysis component configured to determine the methylation state of a sample, a software component configured to compare the methylation state of the sample with a control sample or a reference sample methylation state recorded in a database, and an alert component configured to determine a single value based on a combination of methylation states and alert a user of a breast cancer-associated methylation state. In some embodiments, the sample comprises a nucleic acid comprising a DMR.

In some embodiments, such systems further comprise a component for isolating a nucleic acid. In some embodiments, such systems further comprise a component for collecting a sample.

In some embodiments, the sample is a stool sample, a tissue sample, a breast tissue sample, a blood sample (e.g., plasma sample, whole blood sample, serum sample), or a urine sample.

In some embodiments, the database comprises nucleic acid sequences comprising a DMR. In some embodiments, the database comprises nucleic acid sequences from subjects who do not have a breast cancer.

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

**DEFINITIONS**

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

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

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

The transitional phrase "consisting essentially of" as used in claims in the present application limits the scope of a claim to the specified materials or steps "and those that do not materially affect the basic and novel characteristic(s)" of the claimed invention, as discussed in *In re Herz, 537 F.2d 549, 551-52, 190 USPQ 461, 463 (CCPA 1976)*. For example, a composition "consisting essentially of" recited elements may contain an unrecited contaminant at a level such that, though present, the contaminant does not alter the function of the recited composition as compared to a pure composition, *i.e.*, a composition "consisting of" the recited components.

As used herein, a "nucleic acid" or "nucleic acid molecule" generally refers to any ribonucleic acid or deoxyribonucleic acid, which may be unmodified or modified DNA or RNA. "Nucleic acids" include, without limitation, single- and double-stranded nucleic acids.

As used herein, the term "nucleic acid" also includes DNA as described above that contains one or more modified bases. Thus, DNA with a backbone modified for stability or for other reasons is a "nucleic acid". The term "nucleic acid" as it is used herein embraces such chemically, enzymatically, or metabolically modified forms of nucleic acids, as well as the chemical forms of DNA characteristic of viruses and cells, including for example, simple and complex cells.

The terms "oligonucleotide" or "polynucleotide" or "nucleotide" or "nucleic acid" refer to a molecule having two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof. Typical deoxyribonucleotides for DNA are thymine, adenine, cytosine, and guanine. Typical ribonucleotides for RNA are uracil, adenine, cytosine, and guanine.

As used herein, the terms "locus" or "region" of a nucleic acid refer to a subregion of a nucleic acid, e.g., a gene on a chromosome, a single nucleotide, a CpG island, etc.

The terms "complementary" and "complementarity" refer to nucleotides (e.g., 1 nucleotide) or polynucleotides (e.g., a sequence of nucleotides) related by the base-pairing rules. For example, the sequence 5'-A-G-T-3' is complementary to the sequence 3'-T-C-A-5'. Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands effects the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions and in detection methods that depend upon binding between nucleic acids.

The term "gene" refers to a nucleic acid (e.g., DNA or RNA) sequence that comprises coding sequences necessary for the production of an RNA, or of a polypeptide or its precursor. A functional polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence as long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction, etc.) of the polypeptide are retained. The term "portion" when used in reference to a gene refers to fragments of that gene. The fragments may range in size from a few nucleotides to the entire gene sequence

minus one nucleotide. Thus, "a nucleotide comprising at least a portion of a gene" may comprise fragments of the gene or the entire gene.

The term "gene" also encompasses the coding regions of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends, e.g., for a distance of about 1 kb on either end, such that the gene corresponds to the length of the full-length mRNA (e.g., comprising coding, regulatory, structural and other sequences). The sequences that are located 5' of the coding region and that are present on the mRNA are referred to as 5' non-translated or untranslated sequences. The sequences that are located 3' or downstream of the coding region and that are present on the mRNA are referred to as 3' non-translated or 3' untranslated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. In some organisms (e.g., eukaryotes), a genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' ends of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, posttranscriptional cleavage, and polyadenylation.

The term "wild-type" when made in reference to a gene refers to a gene that has the characteristics of a gene isolated from a naturally occurring source. The term "wild-type" when made in reference to a gene product refers to a gene product that has the characteristics of a gene product isolated from a naturally occurring source. The term "naturally-occurring" as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by the hand of a person in the laboratory is naturally-occurring. A wild-type gene is often that gene

or allele that is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" when made in reference to a gene or to a gene product refers, respectively, to a gene or to a gene product that displays modifications in sequence and/or functional properties (e.g., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term "allele" refers to a variation of a gene; the variations include but are not limited to variants and mutants, polymorphic loci, and single nucleotide polymorphic loci, frameshift, and splice mutations. An allele may occur naturally in a population or it might arise during the lifetime of any particular individual of the population.

Thus, the terms "variant" and "mutant" when used in reference to a nucleotide sequence refer to a nucleic acid sequence that differs by one or more nucleotides from another, usually related, nucleotide acid sequence. A "variation" is a difference between two different nucleotide sequences; typically, one sequence is a reference sequence.

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (e.g., replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (e.g., synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

The term "amplifying" or "amplification" in the context of nucleic acids refers to the production of multiple copies of a polynucleotide, or a portion of the polynucleotide, typically starting from a small amount of the polynucleotide (*e.g.*, a single polynucleotide molecule), where the amplification products or amplicons are generally detectable. Amplification of polynucleotides encompasses a variety of chemical and enzymatic processes. The generation of multiple DNA copies from one or a few copies of a target or template DNA molecule during a polymerase chain reaction (PCR) or a ligase chain reaction (LCR; see, *e.g.*, U.S. Patent No. 5,494,810; herein incorporated by reference in its entirety) are forms of amplification. Additional types of amplification include, but are not limited to, allele-specific PCR (see, *e.g.*, U.S. Patent No. 5,639,611; herein incorporated by reference in

its entirety), assembly PCR (see, *e.g.*, U.S. Patent No. 5,965,408; herein incorporated by reference in its entirety), helicase-dependent amplification (see, *e.g.*, U.S. Patent No. 7,662,594; herein incorporated by reference in its entirety), hot-start PCR (see, *e.g.*, U.S. Patent Nos. 5,773,258 and 5,338,671; each herein incorporated by reference in their entireties), intersequence-specific PCR, inverse PCR (see, *e.g.*, Triglia, *et al.*(1988) Nucleic Acids Res., 16:8186; herein incorporated by reference in its entirety), ligation-mediated PCR (see, *e.g.*, Guilfoyle, R. *et al.*, Nucleic Acids Research, 25:1854-1858 (1997); U.S. Patent No. 5,508,169; each of which are herein incorporated by reference in their entireties), methylation-specific PCR (see, *e.g.*, Herman, *et al.*, (1996) PNAS 93(13) 9821-9826; herein incorporated by reference in its entirety), miniprimer PCR, multiplex ligation-dependent probe amplification (see, *e.g.*, Schouten, *et al.*, (2002) Nucleic Acids Research 30(12): e57; herein incorporated by reference in its entirety), multiplex PCR (see, *e.g.*, Chamberlain, *et al.*, (1988) Nucleic Acids Research 16(23) 11141-11156; Ballabio, *et al.*, (1990) Human Genetics 84(6) 571-573; Hayden, *et al.*, (2008) BMC Genetics 9:80; each of which are herein incorporated by reference in their entireties), nested PCR, overlap-extension PCR (see, *e.g.*, Higuchi, *et al.*, (1988) Nucleic Acids Research 16(15) 7351-7367; herein incorporated by reference in its entirety), real time PCR (see, *e.g.*, Higuchi, *et al.*, (1992) Biotechnology 10:413-417; Higuchi, *et al.*, (1993) Biotechnology 11:1026-1030; each of which are herein incorporated by reference in their entireties), reverse transcription PCR (see, *e.g.*, Bustin, S.A. (2000) J. Molecular Endocrinology 25:169-193; herein incorporated by reference in its entirety), solid phase PCR, thermal asymmetric interlaced PCR, and Touchdown PCR (see, *e.g.*, Don, *et al.*, Nucleic Acids Research (1991) 19(14) 4008; Roux, K. (1994) Biotechniques 16(5) 812-814; Hecker, *et al.*, (1996) Biotechniques 20(3) 478-485; each of which are herein incorporated by reference in their entireties). Polynucleotide amplification also can be accomplished using digital PCR (see, *e.g.*, Kalinina, *et al.*, Nucleic Acids Research. 25; 1999-2004, (1997); Vogelstein and Kinzler, Proc Natl Acad Sci USA. 96; 9236-41, (1999); International Patent Publication No. WO05023091A2; US Patent Application Publication No. 20070202525; each of which are incorporated herein by reference in their entireties).

The term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic or other DNA or RNA, without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture

containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" ("PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified" and are "PCR products" or "amplicons." Those of skill in the art will understand the term "PCR" encompasses many variants of the originally described method using, *e.g.*, real time PCR, nested PCR, reverse transcription PCR (RT-PCR), single primer and arbitrarily primed PCR, *etc.*

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q-beta replicase, MDV-1 RNA is the specific template for the replicase (Kacian et al., Proc. Natl. Acad. Sci. USA, 69:3038 [1972]). Other nucleic acid will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin et al, Nature, 228:227 [1970]). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace (1989) Genomics 4:560). Finally, thermostable template-dependant DNA polymerases (e.g., Taq and Pfu DNA polymerases), by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the

target sequences and not hybridization with non-target sequences (H. A. Erlich (ed.), PCR Technology, Stockton Press [1989]).

As used herein, the term "nucleic acid detection assay" refers to any method of determining the nucleotide composition of a nucleic acid of interest. Nucleic acid detection assay include but are not limited to, DNA sequencing methods, probe hybridization methods, structure specific cleavage assays (*e.g.*, the INVADER assay, (Hologic, Inc.) and are described, *e.g.*, in U.S. Patent Nos. 5,846,717, 5,985,557, 5,994,069, 6,001,567, 6,090,543, and 6,872,816; Lyamichev et al., Nat. Biotech., 17:292 (1999), Hall et al., PNAS, USA, 97:8272 (2000), and US Pat. No. 9,096,893, each of which is herein incorporated by reference in its entirety for all purposes); enzyme mismatch cleavage methods (*e.g.*, Variagenics, U.S. Pat. Nos. 6,110,684, 5,958,692, 5,851,770, herein incorporated by reference in their entireties); polymerase chain reaction (PCR), described above; branched hybridization methods (*e.g.*, Chiron, U.S. Pat. Nos. 5,849,481, 5,710,264, 5,124,246, and 5,624,802, herein incorporated by reference in their entireties); rolling circle replication (*e.g.*, U.S. Pat. Nos. 6,210,884, 6,183,960 and 6,235,502, herein incorporated by reference in their entireties); NASBA (*e.g.*, U.S. Pat. No. 5,409,818, herein incorporated by reference in its entirety); molecular beacon technology (*e.g.*, U.S. Pat. No. 6,150,097, herein incorporated by reference in its entirety); E-sensor technology (Motorola, U.S. Pat. Nos. 6,248,229, 6,221,583, 6,013,170, and 6,063,573, herein incorporated by reference in their entireties); cycling probe technology (*e.g.*, U.S. Pat. Nos. 5,403,711, 5,011,769, and 5,660,988, herein incorporated by reference in their entireties); Dade Behring signal amplification methods (*e.g.*, U.S. Pat. Nos. 6,121,001, 6,110,677, 5,914,230, 5,882,867, and 5,792,614, herein incorporated by reference in their entireties); ligase chain reaction (*e.g.*, Baranay Proc. Natl. Acad. Sci USA 88, 189-93 (1991)); and sandwich hybridization methods (*e.g.*, U.S. Pat. No. 5,288,609, herein incorporated by reference in its entirety).

The term "amplifiable nucleic acid" refers to a nucleic acid that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

The term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover or it may be due to the presence of nucleic acid contaminants sought to be purified away from

the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

The term "primer" refers to an oligonucleotide, whether occurring naturally as, *e.g.*, a nucleic acid fragment from a restriction digest, or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid template strand is induced, (*e.g.*, in the presence of nucleotides and an inducing agent such as a DNA polymerase, and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer, and the use of the method.

The term "probe" refers to an oligonucleotide (*e.g.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly, or by PCR amplification, that is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification, and isolation of particular gene sequences (*e.g.*, a "capture probe"). It is contemplated that any probe used in the present invention may, in some embodiments, be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The term "target," as used herein refers to a nucleic acid sought to be sorted out from other nucleic acids, *e.g.*, by probe binding, amplification, isolation, capture, *etc.* For example, when used in reference to the polymerase chain reaction, "target" refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction, while when used in an assay in which target DNA is not amplified, *e.g.*, in some embodiments of an invasive cleavage assay, a target comprises the site at which a probe and invasive oligonucleotides (*e.g.*, INVADER oligonucleotide) bind to form an invasive cleavage structure, such that the presence of the target nucleic acid can be detected. A "segment" is defined as a region of nucleic acid within the target sequence.

As used herein, "methylation" refers to cytosine methylation at positions C5 or N4 of cytosine, the N6 position of adenine, or other types of nucleic acid methylation. In vitro amplified DNA is usually unmethylated because typical in vitro DNA amplification methods do not retain the methylation pattern of the amplification template. However, "unmethylated DNA" or "methylated DNA" can also refer to amplified DNA whose original template was unmethylated or methylated, respectively.

Accordingly, as used herein a "methylated nucleotide" or a "methylated nucleotide base" refers to the presence of a methyl moiety on a nucleotide base, where the methyl moiety is not present in a recognized typical nucleotide base. For example, cytosine does not contain a methyl moiety on its pyrimidine ring, but 5-methylcytosine contains a methyl moiety at position 5 of its pyrimidine ring. Therefore, cytosine is not a methylated nucleotide and 5-methylcytosine is a methylated nucleotide. In another example, thymine contains a methyl moiety at position 5 of its pyrimidine ring; however, for purposes herein, thymine is not considered a methylated nucleotide when present in DNA since thymine is a typical nucleotide base of DNA.

As used herein, a "methylated nucleic acid molecule" refers to a nucleic acid molecule that contains one or more methylated nucleotides.

As used herein, a "methylation state", "methylation profile", and "methylation status" of a nucleic acid molecule refers to the presence of absence of one or more methylated nucleotide bases in the nucleic acid molecule. For example, a nucleic acid molecule containing a methylated cytosine is considered methylated (e.g., the methylation state of the nucleic acid molecule is methylated). A nucleic acid molecule that does not contain any methylated nucleotides is considered unmethylated.

The methylation state of a particular nucleic acid sequence (e.g., a gene marker or DNA region as described herein) can indicate the methylation state of every base in the sequence or can indicate the methylation state of a subset of the bases (e.g., of one or more cytosines) within the sequence, or can indicate information regarding regional methylation density within the sequence with or without providing precise information of the locations within the sequence the methylation occurs.

The methylation state of a nucleotide locus in a nucleic acid molecule refers to the presence or absence of a methylated nucleotide at a particular locus in the nucleic acid molecule. For example, the methylation state of a cytosine at the 7th nucleotide in a nucleic acid molecule is methylated when the nucleotide present at the 7th nucleotide in the nucleic

acid molecule is 5-methylcytosine. Similarly, the methylation state of a cytosine at the 7th nucleotide in a nucleic acid molecule is unmethylated when the nucleotide present at the 7th nucleotide in the nucleic acid molecule is cytosine (and not 5-methylcytosine).

The methylation status can optionally be represented or indicated by a "methylation value" (e.g., representing a methylation frequency, fraction, ratio, percent, etc.) A methylation value can be generated, for example, by quantifying the amount of intact nucleic acid present following restriction digestion with a methylation dependent restriction enzyme or by comparing amplification profiles after bisulfite reaction or by comparing sequences of bisulfite-treated and untreated nucleic acids. Accordingly, a value, e.g., a methylation value, represents the methylation status and can thus be used as a quantitative indicator of methylation status across multiple copies of a locus. This is of particular use when it is desirable to compare the methylation status of a sequence in a sample to a threshold or reference value.

As used herein, "methylation frequency" or "methylation percent (%)" refer to the number of instances in which a molecule or locus is methylated relative to the number of instances the molecule or locus is unmethylated.

As such, the methylation state describes the state of methylation of a nucleic acid (e.g., a genomic sequence). In addition, the methylation state refers to the characteristics of a nucleic acid segment at a particular genomic locus relevant to methylation. Such characteristics include, but are not limited to, whether any of the cytosine (C) residues within this DNA sequence are methylated, the location of methylated C residue(s), the frequency or percentage of methylated C throughout any particular region of a nucleic acid, and allelic differences in methylation due to, e.g., difference in the origin of the alleles. The terms "methylation state", "methylation profile", and "methylation status" also refer to the relative concentration, absolute concentration, or pattern of methylated C or unmethylated C throughout any particular region of a nucleic acid in a biological sample. For example, if the cytosine (C) residue(s) within a nucleic acid sequence are methylated it may be referred to as "hypermethylated" or having "increased methylation", whereas if the cytosine (C) residue(s) within a DNA sequence are not methylated it may be referred to as "hypomethylated" or having "decreased methylation". Likewise, if the cytosine (C) residue(s) within a nucleic acid sequence are methylated as compared to another nucleic acid sequence (e.g., from a different region or from a different individual, etc.) that sequence is considered hypermethylated or having increased methylation compared to the other nucleic acid sequence. Alternatively, if

the cytosine (C) residue(s) within a DNA sequence are not methylated as compared to another nucleic acid sequence (e.g., from a different region or from a different individual, etc.) that sequence is considered hypomethylated or having decreased methylation compared to the other nucleic acid sequence. Additionally, the term "methylation pattern" as used herein refers to the collective sites of methylated and unmethylated nucleotides over a region of a nucleic acid. Two nucleic acids may have the same or similar methylation frequency or methylation percent but have different methylation patterns when the number of methylated and unmethylated nucleotides are the same or similar throughout the region but the locations of methylated and unmethylated nucleotides are different. Sequences are said to be "differentially methylated" or as having a "difference in methylation" or having a "different methylation state" when they differ in the extent (e.g., one has increased or decreased methylation relative to the other), frequency, or pattern of methylation. The term "differential methylation" refers to a difference in the level or pattern of nucleic acid methylation in a cancer positive sample as compared with the level or pattern of nucleic acid methylation in a cancer negative sample. It may also refer to the difference in levels or patterns between patients that have recurrence of cancer after surgery versus patients who not have recurrence. Differential methylation and specific levels or patterns of DNA methylation are prognostic and predictive biomarkers, e.g., once the correct cut-off or predictive characteristics have been defined.

Methylation state frequency can be used to describe a population of individuals or a sample from a single individual. For example, a nucleotide locus having a methylation state frequency of 50% is methylated in 50% of instances and unmethylated in 50% of instances. Such a frequency can be used, for example, to describe the degree to which a nucleotide locus or nucleic acid region is methylated in a population of individuals or a collection of nucleic acids. Thus, when methylation in a first population or pool of nucleic acid molecules is different from methylation in a second population or pool of nucleic acid molecules, the methylation state frequency of the first population or pool will be different from the methylation state frequency of the second population or pool. Such a frequency also can be used, for example, to describe the degree to which a nucleotide locus or nucleic acid region is methylated in a single individual. For example, such a frequency can be used to describe the degree to which a group of cells from a tissue sample are methylated or unmethylated at a nucleotide locus or nucleic acid region.

As used herein a "nucleotide locus" refers to the location of a nucleotide in a nucleic acid molecule. A nucleotide locus of a methylated nucleotide refers to the location of a methylated nucleotide in a nucleic acid molecule.

Typically, methylation of human DNA occurs on a dinucleotide sequence including an adjacent guanine and cytosine where the cytosine is located 5' of the guanine (also termed CpG dinucleotide sequences). Most cytosines within the CpG dinucleotides are methylated in the human genome, however some remain unmethylated in specific CpG dinucleotide rich genomic regions, known as CpG islands (see, e.g, Antequera et al. (1990) *Cell* **62**: 503–514).

As used herein, a "CpG island" refers to a G:C-rich region of genomic DNA containing an increased number of CpG dinucleotides relative to total genomic DNA. A CpG island can be at least 100, 200, or more base pairs in length, where the G:C content of the region is at least 50% and the ratio of observed CpG frequency over expected frequency is 0.6; in some instances, a CpG island can be at least 500 base pairs in length, where the G:C content of the region is at least 55%) and the ratio of observed CpG frequency over expected frequency is 0.65. The observed CpG frequency over expected frequency can be calculated according to the method provided in Gardiner-Garden et al (1987) *J. Mol. Biol.* **196**: 261–281. For example, the observed CpG frequency over expected frequency can be calculated according to the formula R = (A × B) / (C × D), where R is the ratio of observed CpG frequency over expected frequency, A is the number of CpG dinucleotides in an analyzed sequence, B is the total number of nucleotides in the analyzed sequence, C is the total number of C nucleotides in the analyzed sequence, and D is the total number of G nucleotides in the analyzed sequence. Methylation state is typically determined in CpG islands, e.g., at promoter regions. It will be appreciated though that other sequences in the human genome are prone to DNA methylation such as CpA and CpT (see Ramsahoye (2000) Proc. Natl. Acad. Sci. USA 97: 5237–5242; Salmon and Kaye (1970) Biochim. Biophys. Acta. 204: 340–351; Grafstrom (1985) Nucleic Acids Res. 13: 2827–2842; Nyce (1986) Nucleic Acids Res. 14: 4353–4367; Woodcock (1987) Biochem. Biophys. Res. Commun. 145: 888-894).

As used herein, a "methylation-specific reagent" refers to a reagent that modifies a nucleotide of the nucleic acid molecule as a function of the methylation state of the nucleic acid molecule, or a methylation-specific reagent, refers to a compound or composition or other agent that can change the nucleotide sequence of a nucleic acid molecule in a manner that reflects the methylation state of the nucleic acid molecule. Methods of treating a nucleic acid molecule with such a reagent can include contacting the nucleic acid molecule with the

reagent, coupled with additional steps, if desired, to accomplish the desired change of nucleotide sequence. Such methods can be applied in a manner in which unmethylated nucleotides (*e.g.*, each unmethylated cytosine) is modified to a different nucleotide. For example, in some embodiments, such a reagent can deaminate unmethylated cytosine nucleotides to produce deoxy uracil residues. Examples of such reagents include, but are not limited to, a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent.

A change in the nucleic acid nucleotide sequence by a methylation –specific reagent can also result in a nucleic acid molecule in which each methylated nucleotide is modified to a different nucleotide.

The term "methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of a nucleic acid.

The term "MS AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al. (1997) *Cancer Research* **57**: 594–599.

The term "MethyLight™" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al. (1999) *Cancer Res.* **59**: 2302–2306.

The term "HeavyMethyl™" refers to an assay wherein methylation specific blocking probes (also referred to herein as blockers) covering CpG positions between, or covered by, the amplification primers enable methylation-specific selective amplification of a nucleic acid sample.

The term "HeavyMethyl™ MethyLight™" assay refers to a HeavyMethyl™ MethyLight™ assay, which is a variation of the MethyLight™ assay, wherein the MethyLight™ assay is combined with methylation specific blocking probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones (1997) *Nucleic Acids Res.* 25: 2529–2531.

The term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. (1996) *Proc. Natl. Acad. Sci. USA* **93**: 9821–9826, and by U.S. Pat. No. 5,786,146.

The term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird (1997) *Nucleic Acids Res.* **25**: 2532–2534.

The term "MCA" (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al. (1999) *Cancer Res.* **59**: 2307–12, and in WO 00/26401A1.

As used herein, a "selected nucleotide" refers to one nucleotide of the four typically occurring nucleotides in a nucleic acid molecule (C, G, T, and A for DNA and C, G, U, and A for RNA), and can include methylated derivatives of the typically occurring nucleotides (e.g., when C is the selected nucleotide, both methylated and unmethylated C are included within the meaning of a selected nucleotide), whereas a methylated selected nucleotide refers specifically to a methylated typically occurring nucleotide and an unmethylated selected nucleotides refers specifically to an unmethylated typically occurring nucleotide.

The term "methylation-specific restriction enzyme" refers to a restriction enzyme that selectively digests a nucleic acid dependent on the methylation state of its recognition site. In the case of a restriction enzyme that specifically cuts if the recognition site is not methylated or is hemi-methylated (a methylation-sensitive enzyme), the cut will not take place (or will take place with a significantly reduced efficiency) if the recognition site is methylated on one or both strands. In the case of a restriction enzyme that specifically cuts only if the recognition site is methylated (a methylation-dependent enzyme), the cut will not take place (or will take place with a significantly reduced efficiency) if the recognition site is not methylated. Preferred are methylation-specific restriction enzymes, the recognition sequence of which contains a CG dinucleotide (for instance a recognition sequence such as CGCG or CCCGGG). Further preferred for some embodiments are restriction enzymes that do not cut if the cytosine in this dinucleotide is methylated at the carbon atom C5.

As used herein, a "different nucleotide" refers to a nucleotide that is chemically different from a selected nucleotide, typically such that the different nucleotide has Watson-Crick base-pairing properties that differ from the selected nucleotide, whereby the typically occurring nucleotide that is complementary to the selected nucleotide is not the same as the typically occurring nucleotide that is complementary to the different nucleotide. For example, when C is the selected nucleotide, U or T can be the different nucleotide, which is exemplified by the complementarity of C to G and the complementarity of U or T to A. As used herein, a nucleotide that is complementary to the selected nucleotide or that is complementary to the different nucleotide refers to a nucleotide that base-pairs, under high

stringency conditions, with the selected nucleotide or different nucleotide with higher affinity than the complementary nucleotide's base-paring with three of the four typically occurring nucleotides. An example of complementarity is Watson-Crick base pairing in DNA (e.g., A-T and C-G) and RNA (e.g., A-U and C-G). Thus, for example, G base-pairs, under high stringency conditions, with higher affinity to C than G base-pairs to G, A, or T and, therefore, when C is the selected nucleotide, G is a nucleotide complementary to the selected nucleotide.

As used herein, the "sensitivity" of a given marker (or set of markers used together) refers to the percentage of samples that report a DNA methylation value above a threshold value that distinguishes between neoplastic and non-neoplastic samples. In some embodiments, a positive is defined as a histology-confirmed neoplasia that reports a DNA methylation value above a threshold value (*e.g.*, the range associated with disease), and a false negative is defined as a histology-confirmed neoplasia that reports a DNA methylation value below the threshold value (*e.g.*, the range associated with no disease). The value of sensitivity, therefore, reflects the probability that a DNA methylation measurement for a given marker obtained from a known diseased sample will be in the range of disease-associated measurements. As defined here, the clinical relevance of the calculated sensitivity value represents an estimation of the probability that a given marker would detect the presence of a clinical condition when applied to a subject with that condition.

As used herein, the "specificity" of a given marker (or set of markers used together) refers to the percentage of non-neoplastic samples that report a DNA methylation value below a threshold value that distinguishes between neoplastic and non-neoplastic samples. In some embodiments, a negative is defined as a histology-confirmed non-neoplastic sample that reports a DNA methylation value below the threshold value (*e.g.*, the range associated with no disease) and a false positive is defined as a histology-confirmed non-neoplastic sample that reports a DNA methylation value above the threshold value (*e.g.*, the range associated with disease). The value of specificity, therefore, reflects the probability that a DNA methylation measurement for a given marker obtained from a known non-neoplastic sample will be in the range of non-disease associated measurements. As defined here, the clinical relevance of the calculated specificity value represents an estimation of the probability that a given marker would detect the absence of a clinical condition when applied to a patient without that condition.

The term “AUC” as used herein is an abbreviation for the “area under a curve”. In particular it refers to the area under a Receiver Operating Characteristic (ROC) curve. The ROC curve is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test. It shows the trade-off between sensitivity and specificity depending on the selected cut point (any increase in sensitivity will be accompanied by a decrease in specificity). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test (the larger the area the better; the optimum is 1; a random test would have a ROC curve lying on the diagonal with an area of 0.5; for reference: J. P. Egan. (1975) *Signal Detection Theory and ROC Analysis*, Academic Press, New York).

The term "neoplasm" as used herein refers to any new and abnormal growth of tissue. Thus, a neoplasm can be a premalignant neoplasm or a malignant neoplasm.

The term "neoplasm-specific marker," as used herein, refers to any biological material or element that can be used to indicate the presence of a neoplasm. Examples of biological materials include, without limitation, nucleic acids, polypeptides, carbohydrates, fatty acids, cellular components (*e.g.*, cell membranes and mitochondria), and whole cells. In some instances, markers are particular nucleic acid regions (*e.g.*, genes, intragenic regions, specific loci, etc.). Regions of nucleic acid that are markers may be referred to, *e.g.*, as "marker genes," "marker regions," "marker sequences," "marker loci," etc.

As used herein, the term “adenoma” refers to a benign tumor of glandular origin. Although these growths are benign, over time they may progress to become malignant.

The term “pre-cancerous” or “pre-neoplastic” and equivalents thereof refer to any cellular proliferative disorder that is undergoing malignant transformation.

A “site” of a neoplasm, adenoma, cancer, etc. is the tissue, organ, cell type, anatomical area, body part, etc. in a subject’s body where the neoplasm, adenoma, cancer, etc. is located.

As used herein, a “diagnostic” test application includes the detection or identification of a disease state or condition of a subject, determining the likelihood that a subject will contract a given disease or condition, determining the likelihood that a subject with a disease or condition will respond to therapy, determining the prognosis of a subject with a disease or condition (or its likely progression or regression), and determining the effect of a treatment on a subject with a disease or condition. For example, a diagnostic can be used for detecting the presence or likelihood of a subject contracting a neoplasm or the likelihood that such a

subject will respond favorably to a compound (e.g., a pharmaceutical, e.g., a drug) or other treatment.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids, such as DNA and RNA, are found in the state they exist in nature. Examples of non-isolated nucleic acids include: a given DNA sequence (e.g., a gene) found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, isolated nucleic acid encoding a particular protein includes, by way of example, such nucleic acid in cells ordinarily expressing the protein, where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or oligonucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide may be single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide may be double-stranded). An isolated nucleic acid may, after isolation from its natural or typical environment, by be combined with other nucleic acids or molecules. For example, an isolated nucleic acid may be present in a host cell in which into which it has been placed, e.g., for heterologous expression.

The term "purified" refers to molecules, either nucleic acid or amino acid sequences that are removed from their natural environment, isolated, or separated. An "isolated nucleic acid sequence" may therefore be a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated. As used herein, the terms "purified" or "to purify" also refer to the removal of contaminants from a sample. The removal of contaminating proteins results in an increase in the percent of polypeptide or nucleic acid of interest in the sample. In another example, recombinant polypeptides are expressed in plant, bacterial, yeast, or mammalian host cells and the polypeptides are purified

by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

The term "composition comprising" a given polynucleotide sequence or polypeptide refers broadly to any composition containing the given polynucleotide sequence or polypeptide. The composition may comprise an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The term "sample" is used in its broadest sense. In one sense it can refer to an animal cell or tissue. In another sense, it refers to a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. Environmental samples include environmental material such as surface matter, soil, water, and industrial samples. These examples are not to be construed as limiting the sample types applicable to the present invention.

As used herein, a "remote sample" as used in some contexts relates to a sample indirectly collected from a site that is not the cell, tissue, or organ source of the sample. For instance, when sample material originating from the pancreas is assessed in a stool sample (e.g., not from a sample taken directly from a breast), the sample is a remote sample.

As used herein, the terms "patient" or "subject" refer to organisms to be subject to various tests provided by the technology. The term "subject" includes animals, preferably mammals, including humans. In a preferred embodiment, the subject is a primate. In an even more preferred embodiment, the subject is a human. Further with respect to diagnostic methods, a preferred subject is a vertebrate subject. A preferred vertebrate is warm-blooded; a preferred warm-blooded vertebrate is a mammal. A preferred mammal is most preferably a human. As used herein, the term "subject' includes both human and animal subjects. Thus, veterinary therapeutic uses are provided herein. As such, the present technology provides for the diagnosis of mammals such as humans, as well as those mammals of importance due to being endangered, such as Siberian tigers; of economic importance, such as animals raised on farms for consumption by humans; and/or animals of social importance to humans, such as animals kept as pets or in zoos. Examples of such animals include but are not limited to: carnivores such as cats and dogs; swine, including pigs, hogs, and wild boars; ruminants and/or ungulates such as cattle, oxen, sheep, giraffes, deer, goats, bison, and camels; pinnipeds; and horses. Thus, also provided is the diagnosis and treatment of livestock,

including, but not limited to, domesticated swine, ruminants, ungulates, horses (including race horses), and the like. The presently-disclosed subject matter further includes a system for diagnosing a lung cancer in a subject. The system can be provided, for example, as a commercial kit that can be used to screen for a risk of lung cancer or diagnose a lung cancer in a subject from whom a biological sample has been collected. An exemplary system provided in accordance with the present technology includes assessing the methylation state of a marker described herein.

As used herein, the term "kit" refers to any delivery system for delivering materials. In the context of reaction assays, such delivery systems include systems that allow for the storage, transport, or delivery of reaction reagents (e.g., oligonucleotides, enzymes, etc. in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions for performing the assay etc.) from one location to another. For example, kits include one or more enclosures (e.g., boxes) containing the relevant reaction reagents and/or supporting materials. As used herein, the term "fragmented kit" refers to delivery systems comprising two or more separate containers that each contain a subportion of the total kit components. The containers may be delivered to the intended recipient together or separately. For example, a first container may contain an enzyme for use in an assay, while a second container contains oligonucleotides. The term "fragmented kit" is intended to encompass kits containing Analyte specific reagents (ASR's) regulated under section 520(e) of the Federal Food, Drug, and Cosmetic Act, but are not limited thereto. Indeed, any delivery system comprising two or more separate containers that each contains a subportion of the total kit components are included in the term "fragmented kit." In contrast, a "combined kit" refers to a delivery system containing all of the components of a reaction assay in a single container (e.g., in a single box housing each of the desired components). The term "kit" includes both fragmented and combined kits.

As used herein, the term "breast cancer" refers generally to the uncontrolled growth of breast tissue and, more specifically, to a condition characterized by anomalous rapid proliferation of abnormal cells in one or both breasts of a subject. The abnormal cells often are referred to as malignant or "neoplastic cells," which are transformed cells that can form a solid tumor. The term "tumor" refers to an abnormal mass or population of cells (i.e., two or more cells) that result from excessive or abnormal cell division, whether malignant or benign, and pre-cancerous and cancerous cells. Malignant tumors are distinguished from benign

growths or tumors in that, in addition to uncontrolled cellular proliferation, they can invade surrounding tissues and can metastasize.

As used herein, the term "HER2$^+$ breast cancer" refers to breast cancers wherein at least a portion of the cancer cells express elevated levels of HER2 protein (HER2 (from human epidermal growth factor receptor 2) or HER2/neu) which promotes rapid growth of cells.

As used herein, the term "Luminal A breast cancer" refers to breast cancers wherein at least a portion of the cancer cells are estrogen receptor (ER) positive and progesterone receptor (PR) positive, but negative for HER2.

As used herein, the term "Luminal B breast cancer" refers to breast cancers wherein at least a portion of the cancer cells are ER positive, HER2 positive, and negative for PR.

As used herein, the term "triple negative breast cancer" refers to breast cancers wherein at least a portion of the cancer cells are negative for ER, HER2, and PR.

As used herein, the term "HER2$^+$ breast cancer" refers to breast cancers wherein at least a portion of the cancer cells are negative for ER and PR, but positive for HER2.

As used herein, the term "BRCA1 breast cancer" refers to breast cancers wherein at least a portion of the cancer cells are characterized with a mutation in the BRCA1 gene and/or reduced wild type BRCA1 expression.

As used herein, the term "BRCA2 breast cancer" refers to breast cancers wherein at least a portion of the cancer cells are characterized with a mutation in the BRCA2 gene and/or reduced wild type BRCA2 expression.

As used herein the term "ductal carcinoma in situ" (DCIS) refers to a non-invasive cancer where abnormal cells are found in the lining of the breast milk duct. "Low grade" DCIS refers to a DCIS that is nuclear grade 1 or has a low mitotic rate. "High grade" DCIS refers to a DCIS that nuclear grade 3 or has a high mitotic rate. "Invasive" DCIS refers to a ductal carcinoma that has spread to non-ductal tissue.

As used herein, the term "information" refers to any collection of facts or data. In reference to information stored or processed using a computer system(s), including but not limited to internets, the term refers to any data stored in any format (*e.g.*, analog, digital, optical, *etc.*). As used herein, the term "information related to a subject" refers to facts or data pertaining to a subject (*e.g.*, a human, plant, or animal). The term "genomic information" refers to information pertaining to a genome including, but not limited to, nucleic acid sequences, genes, percentage methylation, allele frequencies, RNA expression levels, protein

expression, phenotypes correlating to genotypes, *etc.* "Allele frequency information" refers to facts or data pertaining to allele frequencies, including, but not limited to, allele identities, statistical correlations between the presence of an allele and a characteristic of a subject (*e.g.*, a human subject), the presence or absence of an allele in an individual or population, the percentage likelihood of an allele being present in an individual having one or more particular characteristics, *etc.*

**DETAILED DESCRIPTION**

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

Provided herein is technology for breast cancer screening and particularly, but not exclusively, to methods, compositions, and related uses for detecting the presence of breast cancer and/or specific forms of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer). As the technology is described herein, the section headings used are for organizational purposes only and are not to be construed as limiting the subject matter in any way.

Indeed, as described in Examples I, II and III, experiments conducted during the course for identifying embodiments for the present invention identified a novel set of 375 differentially methylated regions (DMRs) for discriminating cancer of the breast derived DNA from non-neoplastic control DNA. From these 375 novel DNA methylation markers, further experiments identified markers capable of distinguishing different types of breast cancer from normal breast tissue. For example, separate sets of DMRs were identified capable of distinguishing 1) triple negative breast cancer tissue from normal breast tissue, 2) $HER2^+$ breast cancer tissue from normal breast tissue, 3) Luminal A breast cancer tissue from normal breast tissue, 4) Luminal B breast cancer tissue from normal breast tissue, 5) BRCA1 breast cancer tissue from normal breast tissue, 6) BRCA2 breast cancer tissue from normal

breast tissue, and 7) invasive breast cancer tissue from normal breast tissue. In addition, DMRs were identified capable of distinguishing between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue. In addition, DMRs were identified capable of plasma from subjects having breast cancer from plasma from subjects not having breast cancer.

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

In particular aspects, the present technology provides compositions and methods for identifying, determining, and/or classifying a cancer such as breast cancer. The methods comprise determining the methylation status of at least one methylation marker in a biological sample isolated from a subject (e.g., stool sample, breast tissue sample, plasma sample), wherein a change in the methylation state of the marker is indicative of the presence, class, or site of a breast cancer. Particular embodiments relate to markers comprising a differentially methylated region (DMR, e.g., DMR 1–375, see Tables 2 and 18) that are used for diagnosis (e.g., screening) of breast cancer and various types of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer).

In addition to embodiments wherein the methylation analysis of at least one marker, a region of a marker, or a base of a marker comprising a DMR (e.g., DMR, e.g., DMR 1–375) provided herein and listed in Tables 2 and 18 is analyzed, the technology also provides panels of markers comprising at least one marker, region of a marker, or base of a marker comprising a DMR with utility for the detection of cancers, in particular breast cancer.

Some embodiments of the technology are based upon the analysis of the CpG methylation status of at least one marker, region of a marker, or base of a marker comprising a DMR.

In some embodiments, the present technology provides for the use of a reagent that modifies DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent) in combination with one or more methylation assays to determine the methylation status of CpG dinucleotide sequences within at least one marker comprising a DMR (e.g., DMR 1–375, see Tables 2 and 18). Genomic CpG dinucleotides can be methylated or unmethylated (alternatively known as up- and down-methylated respectively). However the methods of the present invention are

suitable for the analysis of biological samples of a heterogeneous nature, e.g., a low concentration of tumor cells, or biological materials therefrom, within a background of a remote sample (e.g., blood, organ effluent, or stool). Accordingly, when analyzing the methylation status of a CpG position within such a sample one may use a quantitative assay for determining the level (e.g., percent, fraction, ratio, proportion, or degree) of methylation at a particular CpG position.

According to the present technology, determination of the methylation status of CpG dinucleotide sequences in markers comprising a DMR has utility both in the diagnosis and characterization of cancers such as breast cancer.

**Combinations of markers**

In some embodiments, the technology relates to assessing the methylation state of combinations of markers comprising a DMR from Tables 2 and 18 (e.g., DMR Nos. 1-375). In some embodiments, assessing the methylation state of more than one marker increases the specificity and/or sensitivity of a screen or diagnostic for identifying a neoplasm in a subject (e.g., breast cancer).

Various cancers are predicted by various combinations of markers, e.g., as identified by statistical techniques related to specificity and sensitivity of prediction. The technology provides methods for identifying predictive combinations and validated predictive combinations for some cancers.

**Methods for assaying methylation state**

In certain embodiments, methods for analyzing a nucleic acid for the presence of 5-methylcytosine involves treatment of DNA with a reagent that modifies DNA in a methylation-specific manner. Examples of such reagents include, but are not limited to, a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent.

A frequently used method for analyzing a nucleic acid for the presence of 5-methylcytosine is based upon the bisulfite method described by Frommer, et al. for the detection of 5-methylcytosines in DNA (Frommer et al. (1992) *Proc. Natl. Acad. Sci. USA* 89: 1827–31 explicitly incorporated herein by reference in its entirety for all purposes) or variations thereof. The bisulfite method of mapping 5-methylcytosines is based on the observation that cytosine, but not 5-methylcytosine, reacts with hydrogen sulfite ion (also

known as bisulfite). The reaction is usually performed according to the following steps: first, cytosine reacts with hydrogen sulfite to form a sulfonated cytosine. Next, spontaneous deamination of the sulfonated reaction intermediate results in a sulfonated uracil. Finally, the sulfonated uracil is desulfonated under alkaline conditions to form uracil. Detection is possible because uracil base pairs with adenine (thus behaving like thymine), whereas 5-methylcytosine base pairs with guanine (thus behaving like cytosine). This makes the discrimination of methylated cytosines from non-methylated cytosines possible by, *e.g.*, bisulfite genomic sequencing (Grigg G, & Clark S, Bioessays (1994) 16: 431–36; Grigg G, DNA Seq. (1996) 6: 189–98),methylation-specific PCR (MSP) as is disclosed, *e.g.*, in U.S. Patent No. 5,786,146, or using an assay comprising sequence-specific probe cleavage, *e.g.*, a QuARTS flap endonuclease assay (see, *e.g.*, Zou et al. (2010) "Sensitive quantification of methylated markers with a novel methylation specific technology" *Clin Chem* **56**: A199; and in U.S. Pat. Nos. 8,361,720; 8,715,937; 8,916,344; and 9,212,392.

Some conventional technologies are related to methods comprising enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing precipitation and purification steps with a fast dialysis (Olek A, et al. (1996) "A modified and improved method for bisulfite based cytosine methylation analysis" *Nucleic Acids Res.* 24: 5064-6). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of conventional methods for detecting 5-methylcytosine is provided by Rein, T., et al. (1998) *Nucleic Acids Res.* 26: 2255.

The bisulfite technique typically involves amplifying short, specific fragments of a known nucleic acid subsequent to a bisulfite treatment, then either assaying the product by sequencing (Olek & Walter (1997) *Nat. Genet.* 17: 275–6) or a primer extension reaction (Gonzalgo & Jones (1997) *Nucleic Acids Res.* 25: 2529–31; WO 95/00669; U.S. Pat. No. 6,251,594) to analyze individual cytosine positions. Some methods use enzymatic digestion (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532–4). Detection by hybridization has also been described in the art (Olek et al., WO 99/28498). Additionally, use of the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg & Clark (1994) *Bioessays* 16: 431–6,; Zeschnigk et al. (1997) *Hum Mol Genet.* 6: 387–95; Feil et al. (1994) *Nucleic Acids Res.* 22: 695; Martin et al. (1995) *Gene* 157: 261–4; WO 9746705; WO 9515373).

Various methylation assay procedures can be used in conjunction with bisulfite treatment according to the present technology. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (*e.g.*, CpG islands) within a nucleic acid sequence. Such assays involve, among other techniques, sequencing of bisulfite-treated nucleic acid, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-specific restriction enzymes, *e.g.*, methylation-sensitive or methylation-dependent enzymes.

For example, genomic sequencing has been simplified for analysis of methylation patterns and 5-methylcytosine distributions by using bisulfite treatment (Frommer et al. (1992) *Proc. Natl. Acad. Sci. USA* 89: 1827–1831). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA finds use in assessing methylation state, e.g., as described by Sadri & Hornsby (1997) *Nucl. Acids Res.* 24: 5058–5059 or as embodied in the method known as COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532–2534).

COBRA™ analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific loci in small amounts of genomic DNA (Xiong & Laird, Nucleic Acids Res. 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (Proc. Natl. Acad. Sci. USA 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the CpG islands of interest, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples.

Typical reagents (e.g., as might be found in a typical COBRA™-based kit) for COBRA™ analysis may include, but are not limited to: PCR primers for specific loci (e.g., specific genes, markers, DMR, regions of genes, regions of markers, bisulfite treated DNA sequence , CpG island, etc.); restriction enzyme and appropriate buffer; gene-hybridization oligonucleotide; control hybridization oligonucleotide; kinase labeling kit for oligonucleotide

probe; and labeled nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components. Assays such as "MethyLight™" (a fluorescence-based real-time PCR technique) (Eads et al., Cancer Res. 59:2302-2306, 1999), Ms-SNuPE™ (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; U.S. Pat. No. 5,786,146), and methylated CpG island amplification ("MCA"; Toyota et al., Cancer Res. 59:2307-12, 1999) are used alone or in combination with one or more of these methods.

The "HeavyMethyl™" assay, technique is a quantitative method for assessing methylation differences based on methylation-specific amplification of bisulfite-treated DNA. Methylation-specific blocking probes ("blockers") covering CpG positions between, or covered by, the amplification primers enable methylation-specific selective amplification of a nucleic acid sample.

The term "HeavyMethyl™ MethyLight™" assay refers to a HeavyMethyl™ MethyLight™ assay, which is a variation of the MethyLight™ assay, wherein the MethyLight™ assay is combined with methylation specific blocking probes covering CpG positions between the amplification primers. The HeavyMethyl™ assay may also be used in combination with methylation specific amplification primers.

Typical reagents (*e.g.*, as might be found in a typical MethyLight™-based kit) for HeavyMethyl™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, or bisulfite treated DNA sequence or CpG island, *etc.*); blocking oligonucleotides; optimized PCR buffers and deoxynucleotides; and Taq polymerase. MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; U.S. Pat. No. 5,786,146). Briefly, DNA is modified by sodium bisulfite, which converts unmethylated, but not methylated cytosines, to uracil, and the products are subsequently amplified with primers specific for methylated versus unmethylated DNA. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (*e.g.*,

as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); optimized PCR buffers and deoxynucleotides, and specific probes.

The MethyLight™ assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (*e.g.*, TaqMan®) that requires no further manipulations after the PCR step (Eads et al., Cancer Res. 59:2302-2306, 1999). Briefly, the MethyLight™ process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed in a "biased" reaction, *e.g.*, with PCR primers that overlap known CpG dinucleotides. Sequence discrimination occurs both at the level of the amplification process and at the level of the fluorescence detection process.

The MethyLight™ assay is used as a quantitative test for methylation patterns in a nucleic acid, *e.g.*, a genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In a quantitative version, the PCR reaction provides for a methylation specific amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe, overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing the biased PCR pool with either control oligonucleotides that do not cover known methylation sites (*e.g.*, a fluorescence-based version of the HeavyMethyl™ and MSP techniques) or with oligonucleotides covering potential methylation sites.

The MethyLight™ process is used with any suitable probe (*e.g.* a "TaqMan®" probe, a Lightcycler® probe, *etc.*) For example, in some applications double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes, *e.g.*, with MSP primers and/or HeavyMethyl blocker oligonucleotides and a TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules and is designed to be specific for a relatively high GC content region so that it melts at about a 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand

during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5′ to 3′ endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Typical reagents (*e.g.*, as might be found in a typical MethyLight™-based kit) for MethyLight™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); TaqMan® or Lightcycler® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

The QM™ (quantitative methylation) assay is an alternative quantitative test for methylation patterns in genomic DNA samples, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe, overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing the biased PCR pool with either control oligonucleotides that do not cover known methylation sites (a fluorescence-based version of the HeavyMethyl™ and MSP techniques) or with oligonucleotides covering potential methylation sites.

The QM™ process can be used with any suitable probe, *e.g.*, "TaqMan®" probes, Lightcycler® probes, in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to unbiased primers and the TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about a 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5′ to 3′ endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system. Typical reagents (*e.g.*, as might be found in a typical QM™-based kit) for QM™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite

treated DNA sequence, CpG island, *etc.*); TaqMan® or Lightcycler® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

The Ms-SNuPE™ technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site of interest. Small amounts of DNA can be analyzed (*e.g.*, microdissected pathology sections) and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

Typical reagents (*e.g.*, as might be found in a typical Ms-SNuPE™-based kit) for Ms-SNuPE™ analysis may include, but are not limited to: PCR primers for specific loci (*e.g.*, specific genes, markers, regions of genes, regions of markers, bisulfite treated DNA sequence, CpG island, *etc.*); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE™ primers for specific loci; reaction buffer (for the Ms-SNuPE reaction); and labeled nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Reduced Representation Bisulfite Sequencing (RRBS) begins with bisulfite treatment of nucleic acid to convert all unmethylated cytosines to uracil, followed by restriction enzyme digestion (*e.g.*, by an enzyme that recognizes a site including a CG sequence such as MspI) and complete sequencing of fragments after coupling to an adapter ligand. The choice of restriction enzyme enriches the fragments for CpG dense regions, reducing the number of redundant sequences that may map to multiple gene positions during analysis. As such, RRBS reduces the complexity of the nucleic acid sample by selecting a subset (*e.g.*, by size selection using preparative gel electrophoresis) of restriction fragments for sequencing. As opposed to whole-genome bisulfite sequencing, every fragment produced by the restriction enzyme digestion contains DNA methylation information for at least one CpG dinucleotide. As such, RRBS enriches the sample for promoters, CpG islands, and other genomic features

with a high frequency of restriction enzyme cut sites in these regions and thus provides an assay to assess the methylation state of one or more genomic loci.

A typical protocol for RRBS comprises the steps of digesting a nucleic acid sample with a restriction enzyme such as MspI, filling in overhangs and A-tailing, ligating adaptors, bisulfite conversion, and PCR. See, *e.g.*, et al. (2005) "Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution" *Nat Methods* **7**: 133–6; Meissner et al. (2005) "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis" *Nucleic Acids Res.* **33**: 5868–77.

In some embodiments, a quantitative allele-specific real-time target and signal amplification (QuARTS) assay is used to evaluate methylation state. Three reactions sequentially occur in each QuARTS assay, including amplification (reaction 1) and target probe cleavage (reaction 2) in the primary reaction; and FRET cleavage and fluorescent signal generation (reaction 3) in the secondary reaction. When target nucleic acid is amplified with specific primers, a specific detection probe with a flap sequence loosely binds to the amplicon. The presence of the specific invasive oligonucleotide at the target binding site causes a 5′ nuclease, *e.g.*, a FEN-1 endonuclease, to release the flap sequence by cutting between the detection probe and the flap sequence. The flap sequence is complementary to a non-hairpin portion of a corresponding FRET cassette. Accordingly, the flap sequence functions as an invasive oligonucleotide on the FRET cassette and effects a cleavage between the FRET cassette fluorophore and a quencher, which produces a fluorescent signal. The cleavage reaction can cut multiple probes per target and thus release multiple fluorophore per flap, providing exponential signal amplification. QuARTS can detect multiple targets in a single reaction well by using FRET cassettes with different dyes. See, *e.g.*, in Zou et al. (2010) "Sensitive quantification of methylated markers with a novel methylation specific technology" *Clin Chem* **56**: A199), and U.S. Pat. Nos. 8,361,720; 8,715,937; 8,916,344; and 9,212,392, each of which is incorporated herein by reference for all purposes.

The term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite, or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences. Methods of said treatment are known in the art (*e.g.*, PCT/EP2004/011715 and WO 2013/116375, each of which is incorporated by reference in its entirety). In some embodiments, bisulfite treatment is conducted in the presence of denaturing solvents such as but not limited to n-alkyleneglycol or diethylene glycol dimethyl ether (DME), or in the presence of dioxane or dioxane

derivatives. In some embodiments the denaturing solvents are used in concentrations between 1% and 35% (v/v). In some embodiments, the bisulfite reaction is carried out in the presence of scavengers such as but not limited to chromane derivatives, *e.g.*, 6-hydroxy-2,5,7,8,-tetramethylchromane 2-carboxylic acid or trihydroxybenzone acid and derivates thereof, *e.g.*, Gallic acid (see: PCT/EP2004/011715, which is incorporated by reference in its entirety). In certain preferred embodiments, the bisulfite reaction comprises treatment with ammonium hydrogen sulfite, *e.g.*, as described in WO 2013/116375.

In some embodiments, fragments of the treated DNA are amplified using sets of primer oligonucleotides according to the present invention (e.g., see Tables 10, 19 and 20) and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). Amplicons are typically 100 to 2000 base pairs in length.

In another embodiment of the method, the methylation status of CpG positions within or near a marker comprising a DMR (e.g., DMR 1–375, Tables 2 and 18) may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in U.S. Pat. No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primer pairs contain at least one primer that hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the position of the C position in the CpG.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. In some embodiments, the labels are fluorescent labels, radionuclides, or detachable molecule fragments having a typical mass that can be detected in a mass spectrometer. Where said labels are mass labels, some embodiments provide that the labeled amplicons have a single positive or negative net charge, allowing for better delectability in the mass spectrometer. The detection may be carried out and visualized by means of, e.g., matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Methods for isolating DNA suitable for these assay technologies are known in the art. In particular, some embodiments comprise isolation of nucleic acids as described in U.S. Pat.

Appl. Ser. No. 13/470,251 ("Isolation of Nucleic Acids"), incorporated herein by reference in its entirety.

In some embodiments, the markers described herein find use in QUARTS assays performed on stool samples. In some embodiments, methods for producing DNA samples and, in particular, to methods for producing DNA samples that comprise highly purified, low-abundance nucleic acids in a small volume (*e.g.*, less than 100, less than 60 microliters) and that are substantially and/or effectively free of substances that inhibit assays used to test the DNA samples (*e.g.*, PCR, INVADER, QuARTS assays, *etc.*) are provided. Such DNA samples find use in diagnostic assays that qualitatively detect the presence of, or quantitatively measure the activity, expression, or amount of, a gene, a gene variant (*e.g.*, an allele), or a gene modification (*e.g.*, methylation) present in a sample taken from a patient. For example, some cancers are correlated with the presence of particular mutant alleles or particular methylation states, and thus detecting and/or quantifying such mutant alleles or methylation states has predictive value in the diagnosis and treatment of cancer. Many valuable genetic markers are present in extremely low amounts in samples and many of the events that produce such markers are rare. Consequently, even sensitive detection methods such as PCR require a large amount of DNA to provide enough of a low-abundance target to meet or supersede the detection threshold of the assay. Moreover, the presence of even low amounts of inhibitory substances compromise the accuracy and precision of these assays directed to detecting such low amounts of a target. Accordingly, provided herein are methods providing the requisite management of volume and concentration to produce such DNA samples.

In some embodiments, the sample comprises blood, serum, plasma, or saliva. In some embodiments, the subject is human. Such samples can be obtained by any number of means known in the art, such as will be apparent to the skilled person. Cell free or substantially cell free samples can be obtained by subjecting the sample to various techniques known to those of skill in the art which include, but are not limited to, centrifugation and filtration. Although it is generally preferred that no invasive techniques are used to obtain the sample, it still may be preferable to obtain samples such as tissue homogenates, tissue sections, and biopsy specimens. The technology is not limited in the methods used to prepare the samples and provide a nucleic acid for testing. For example, in some embodiments, a DNA is isolated from a stool sample or from blood or from a plasma sample using direct gene capture, *e.g.*, as

detailed in U.S. Pat. Nos. 8,808,990 and 9,169,511, and in WO 2012/155072, or by a related method.

The analysis of markers can be carried out separately or simultaneously with additional markers within one test sample. For example, several markers can be combined into one test for efficient processing of multiple samples and for potentially providing greater diagnostic and/or prognostic accuracy. In addition, one skilled in the art would recognize the value of testing multiple samples (for example, at successive time points) from the same subject. Such testing of serial samples can allow the identification of changes in marker methylation states over time. Changes in methylation state, as well as the absence of change in methylation state, can provide useful information about the disease status that includes, but is not limited to, identifying the approximate time from onset of the event, the presence and amount of salvageable tissue, the appropriateness of drug therapies, the effectiveness of various therapies, and identification of the subject's outcome, including risk of future events. The analysis of biomarkers can be carried out in a variety of physical formats. For example, the use of microtiter plates or automation can be used to facilitate the processing of large numbers of test samples. Alternatively, single sample formats could be developed to facilitate immediate treatment and diagnosis in a timely fashion, for example, in ambulatory transport or emergency room settings.

It is contemplated that embodiments of the technology are provided in the form of a kit. The kits comprise embodiments of the compositions, devices, apparatuses, *etc.* described herein, and instructions for use of the kit. Such instructions describe appropriate methods for preparing an analyte from a sample, *e.g.*, for collecting a sample and preparing a nucleic acid from the sample. Individual components of the kit are packaged in appropriate containers and packaging (*e.g.*, vials, boxes, blister packs, ampules, jars, bottles, tubes, and the like) and the components are packaged together in an appropriate container (*e.g.*, a box or boxes) for convenient storage, shipping, and/or use by the user of the kit. It is understood that liquid components (*e.g.*, a buffer) may be provided in a lyophilized form to be reconstituted by the user. Kits may include a control or reference for assessing, validating, and/or assuring the performance of the kit. For example, a kit for assaying the amount of a nucleic acid present in a sample may include a control comprising a known concentration of the same or another nucleic acid for comparison and, in some embodiments, a detection reagent (*e.g.*, a primer) specific for the control nucleic acid. The kits are appropriate for use in a clinical setting and, in some embodiments, for use in a user's home. The components of a kit, in some

embodiments, provide the functionalities of a system for preparing a nucleic acid solution from a sample. In some embodiments, certain components of the system are provided by the user.

**Methods**

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker comprising a DMR (e.g., DMR 1–375 e.g., as provided in Tables 2 and 18) and
2) detecting breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1, and

2) detecting breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B, FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, KLHDC7B_B, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr17.73073682-73073814, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TBX1_B, TRH_A, and TRIM67_B, and
2) detecting breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of CD1D, ITPRIPL1, FAM59B,

C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B, and

2) detecting breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ABLIM1, AJAP1_B, ASCL2, ATP6V1B1, BANK1, CALN1_A, CALN1_B, CLIC6, DSCR6, FOXP4, GAD2, GCGR, GP5, GRASP, HBM, HNF1B_B, KLF16, MAGI2, MAX.chr11.14926602-14927148, MAX.chr12.4273906-4274012, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr2.97193478-97193562, MAX.chr22.42679578-42679917, MAX.chr4.8859253-8859329, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr6.157557371-157557657, MPZ, NKX2-6, PDX1, PLXNC1_A, PPARG, PRKCB, PTPRN2, RBFOX_A, SCRT2_A, SLC7A4, STAC2_B, STX16_A, STX16_B, TBX1, TRH_A, VSTM2B_A, ZBTB16, ZNF132, and ZSCAN23, and

2) detecting triple negative breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of CALN1_A, LOC100132891,

NACAD, TRIM67, ATP6V1B1, DLX4, GP5, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ST8SIA4, STX16_B ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, KCNK9, SCRT2_B, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, and DSCR6, and

2) detecting triple negative breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4, and
2) detecting triple negative breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ABLIM1, AFAP1L1, AKR1B1, ALOX5, AMN, ARL5C, BANK1, BCAT1, BEGAIN, BEST4, BHLHE23_B,

BHLHE23_C, C17orf64, C1QL2, C7orf52, CALN1_B, CAV2, CD8A, CDH4_A, CDH4_B, CDH4_C, CDH4_D, CDH4_E, CDH4_F, CHST2_B, CLIP4, CR1, DLK1, DNAJC6, DNM3_A, EMX1_A, ESPN, FABP5, FAM150A, FLJ42875, GLP1R, GNG4, GYPC_A, HAND2, HES5, HNF1B_A, HNF1B_B, HOXA1_A, HOXA1_B, HOXA7_A, HOXA7_B, HOXA7_C, HOXD9, IGF2BP3_A, IGF2BP3_B, IGSF9B_A, IL15RA, INSM1, ITPKA_B, ITPRIPL1, KCNE3, KCNK17_B, LIME1, LOC100132891, LOC283999, LY6H, MAST1, MAX.chr1.158083198-158083476, MAX.chr1.228074764-228074977, MAX.chr1.46913931-46913950, MAX.chr10.130085265-130085312, MAX.chr11.68622869-68622968, MAX.chr14.101176106-101176260, MAX.chr15.96889069-96889128, MAX.chr17.8230197-8230314, MAX.chr19.46379903-46380197, MAX.chr2.97193163-97193287, MAX.chr2.97193478-97193562, MAX.chr20.1784209-1784461, MAX.chr21.44782441-44782498, MAX.chr22.23908718-23908782, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.180101084-180101094, MAX.chr5.42952185-42952280, MAX.chr5.42994866-42994936, MAX.chr6.27064703-27064783, MAX.chr7.152622607-152622638, MAX.chr8.145104132-145104218, MAX.chr9.136474504-136474527, MCF2L2, MSX2P1, NACAD, NID2_B, NID2_C, ODC1, OSR2_B, PAQR6, PCDH8, PIF1, PPARA, PPP2R5C, PRDM13_A, PRHOXNB, PRKCB, RBFOX3_A, RBFOX3_B, RFX8, SNCA, STAC2_A, STAC2_B, STX16_B SYT5, TIMP2, TMEFF2, TNFRSF10D, TRH_B, TRIM67, TRIM71_C, USP44_A, USP44_B, UTF1, UTS2R, VSTM2B_A, VSTM2B_B, ZFP64, and ZNF132, and

2) detecting $HER2^+$ breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having

an annotation selected from the group consisting of BHLHE23_C, CALN1_A, CD1D, CHST2_A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, NACAD, TRIM67, ATP6V1B1, C17orf64, CHST2_B, DLX4, DNM3_A, EMX1_A, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, COL23A1, KCNK9, LAYN, PLXNC1_A, RIC3, SCRT2_B, ALOX5, CDH4_E, HNF1B_B, TRH_A, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, AJAP1_B, DSCR6, and MAX.chr11.68622869-68622968, and

2) detecting $HER2^+$ breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, C10orf125, and

2) detecting $HER2^+$ breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ARL5C, BHLHE23_C, BMP6, C10orf125, C17orf64, C19orf66, CAMKV, CD1D, CDH4_E, CDH4_F, CHST2_A, CRHBP, DLX6, DNM3_A, DNM3_B, DNM3_C, ESYT3, ETS1_A, ETS1_B, FAM126A, FAM189A1, FAM20A, FAM59B, FBN1, FLRT2, FMN2, FOXP4, GAS7, GYPC_A, GYPC_B, HAND2, HES5, HMGA2, HNF1B_B, IGF2BP3_A, IGF2BP3_B, KCNH8, KCNK17_A, KCNQ2, KLHDC7B, LOC100132891, MAX.chr1.46913931-46913950, MAX.chr11.68622869-68622968, MAX.chr12.4273906-4274012, MAX.chr12.59990591-59990895, MAX.chr17.73073682-73073814, MAX.chr20.1783841-1784054, MAX.chr21.47063802-47063851, MAX.chr4.8860002-8860038, MAX.chr5.172234248-172234494, MAX.chr5.178957564-178957598, MAX.chr6.130686865-130686985, MAX.chr8.687688-687736, MAX.chr8.688863-688924, MAX.chr9.114010-114207, MPZ, NID2_A, NKX2-6, ODC1, OSR2_A, POU4F1, PRDM13_B, PRKCB, RASGRF2, RIPPLY2, SLC30A10, ST8SIA4, SYN2, TRIM71_A, TRIM71_B, TRIM71_C, UBTF, ULBP1, USP44_B, and VSTM2B_A, and
2) detecting Luminal A breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of BHLHE23_C, CD1D, CHST2_A, FAM126A, FMN2, HOXA1_A, HOXA7_A, KCNH8, LOC100132891, MAX.chr15.96889013-96889128, SLC30A10, TRIM67, ATP6V1B1, BANK1, C10orf125, C17orf64, CHST2_B, DNM3_A, EMX1_A, GP5, IGF2BP3_A, IGF2BP3_B, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, ODC1, PLXNC1_A, PRKCB, ST8SIA4, STX16_B UBTF, LOC100132891, ITPRIPL1, MAX.chr12.4273906-4274012, MAX.chr12.59990671-59990859, BHLHE23_D, COL23A1, KCNK9, OTX1, PLXNC1_A, HNF1B_B, MAST1, ASCL2, MAX.chr20.1784209-1784461, RBFOX_A, MAX.chr12.4273906-4274012, GAS7, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, GYPC_B, DLX6, FBN1, OSR2_A, BEST4, DSCR6, MAX.chr11.68622869-68622968, and
2) detecting Luminal A breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, ALOX5,

MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125, and

2) detecting Luminal A breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ACCN1, AJAP1_A, AJAP1_B, BEST4, CALN1_B, CBLN1_B, CDH4_E, DLX4, FOXP4, IGSF9B_B, ITPRIPL1, KCNA1, KLF16, LMX1B_A, MAST1, MAX.chr11.14926602-14927148, MAX.chr17.73073682-73073814, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr22.42679578-42679917, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr8.124173128-124173268, MPZ, PPARA, PRMT1, RBFOX3_B, RYR2_A, SALL3, SCRT2_A, SPHK2, STX16_B SYNJ2, TMEM176A, TSHZ3, and VIPR2, and

2) detecting Luminal B breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having

an annotation selected from the group consisting of CALN1_A, LOC100132891, MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, DLX4, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, ITPRIPL1, KLF16, MAX.chr12.4273906-4274012, MAX.chr19.46379903-46380197, BHLHE23_D, HNF1B_B, TRH_A, ASCL2, MAX.chr20.1784209-1784461, MAX.chr12.4273906-4274012, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, AJAP1_B, and DSCR6, and

2) detecting Luminal B breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D, and

2) detecting Luminal B breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of C10orf93, C20orf195_A, C20orf195_B, CALN1_B, CBLN1_A, CBLN1_B, CCDC61, CCND2_A, CCND2_B, CCND2_C, EMX1_B, FAM150B, GRASP, HBM, ITPRIPL1, KCNK17_A, KIAA1949, LOC100131176, MAST1, MAX.chr1.8277285-8277316, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr18.5629721-5629791, MAX.chr19.30719261-30719354, MAX.chr22.42679767-42679917, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MAX.chr6.157556793-157556856, MAX.chr8.124173030-124173395, MN1, MPZ, NR2F6, PDXK_A, PDXK_B, PTPRM, RYR2_B, SERPINB9_A, SERPINB9_B, SLC8A3, STX16_B TEPP, TOX, VIPR2, VSTM2B_A, ZNF486, ZNF626, and ZNF671, and
2) detecting BRCA1 breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-

46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6, and

2) detecting BRCA1 breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of ANTXR2, B3GNT5, BHLHE23_C, BMP4, CHRNA7, EPHA4, FAM171A1, FAM20A, FMNL2, FSCN1, GSTP1, HBM, IGFBP5, IL17REL, ITGA9, ITPRIPL1, KIRREL2, LRRC34, MAX.chr1.239549742-239549886, MAX.chr1.8277479-8277527, MAX.chr11.14926602-14926729, MAX.chr11.14926860-14927148, MAX.chr15.96889013-96889128, MAX.chr2.238864674-238864735, MAX.chr5.81148300-81148332, MAX.chr7.151145632-151145743, MAX.chr8.124173030-124173395, MAX.chr8.143533298-143533558, MERTK, MPZ, NID2_C, NTRK3, OLIG3_A, OLIG3_B, OSR2_C, PROM1, RGS17, SBNO2, STX16_B TBKBP1, TLX1NB, VIPR2, VN1R2, VSNL1, and ZFP64, and

2) detecting BRCA2 breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-

methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, and

2) detecting BRCA2 breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of CDH4_E, FLJ42875, GAD2, GRASP, ITPRIPL1, KCNA1, MAX.chr12.4273906-4274012, MAX.chr18.76734362-76734370, MAX.chr18.76734423-76734476, MAX.chr19.30719261-30719354, MAX.chr4.8859602-8859669, MAX.chr4.8860002-8860038, MAX.chr5.145725410-145725459, MAX.chr5.178957564-178957598, MAX.chr5.77268672-77268725, MPZ, NKX2-6, PRKCB, RBFOX3_B, SALL3, and VSTM2B_A, and

2) detecting invasive breast cancer (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a

chromosomal region having an annotation selected from the group consisting of SCRT2_B, MPZ, MAX.chr8.124173030-124173395, ITPRIPL1, ITPRIPL1, DLX4, CALN1_A, and IGF2BP3_B, and

2) distinguishing between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of SCRT2_B, ITPRIPL1, and MAX.chr8.124173030-12417339, and

2) distinguishing between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue (e.g., afforded with a sensitivity of greater than or equal to 100% and a specificity of greater than or equal to 91%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) contacting a nucleic acid (e.g., genomic DNA, e.g., isolated from body fluids such as blood or plasma or breast tissue) obtained from the subject with at least one reagent or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker selected from a chromosomal region having an annotation selected from the group consisting of DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, and ITPRIPL1, and

2) distinguishing between ductal carcinoma in situ high grade (DCIS-HG) breast cancer tissue from ductal carcinoma in situ low grade (DCIS-LG) breast tissue (e.g., afforded with a sensitivity of greater than or equal to 80% and a specificity of greater than or equal to 80%).

In some embodiments of the technology, methods are provided that comprise the following steps:

1) measuring a methylation level for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with a reagent that modifies DNA in a methylation-specific manner (e.g., wherein the reagent is a bisulfite reagent, a methylation-sensitive restriction enzyme, or a methylation-dependent restriction enzyme), wherein the one or more genes is selected from one of the following groups:

(i) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1;

(ii) ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B, FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TRH_A, and TRIM67_B; and

(iii) CD1D, ITPRIPL1, FAM59B, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B;

2) amplifying the treated genomic DNA using a set of primers for the selected one or more genes; and

3) determining the methylation level of the one or more genes by polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and target capture.

In some embodiments of the technology, methods are provided that comprise the following steps:

1) measuring an amount of at least one methylated marker gene in DNA from the sample, wherein the one or more genes is selected from one of the following groups:

(i) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1;

(ii) ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B, FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TRH_A, and TRIM67_B; and

(iii) CD1D, ITPRIPL1, FAM59B, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B;

2) measuring the amount of at least one reference marker in the DNA; and

3) calculating a value for the amount of the at least one methylated marker gene measured in the DNA as a percentage of the amount of the reference marker gene measured in the DNA, wherein the value indicates the amount of the at least one methylated marker DNA measured in the sample.

In some embodiments of the technology, methods are provided that comprise the following steps:

1) measuring a methylation level of a CpG site for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with bisulfite a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent);

2) amplifying the modified genomic DNA using a set of primers for the selected one or more genes; and

3) determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR;

wherein the one or more genes is selected from one of the following groups:

(i) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, and ABLIM1;

(ii) ABLIM1_B, AJAP1_C, ALOX5_B, ASCL2_B, BANK1_B, BHLHE23_E, C10orf125_B, C17orf64_B, CALN1_1520, CALN_1B, CD1D_1058, CDH4_7890, CHST2_8128, CHST2_8384, CHST2_9316, CHST2_9470, CLIC6_B, CXCL12_B, DLX4_B, DNM3_D, EMX1_A, ESPN_B, FAM59B_7764, FOXP4_B, GP5, HOXA1_C, IGF2BP3_C, IPTRIPL1_1138, IPTRIPL1_1200, KCNK9_B, KCNK17_C, LAYN_B, LIME1_B, LMX1B_D, LOC100132891_B, MAST1_B, MAX.chr12.427.br, MAX.chr20.4422, MPZ_5742, MPZ_5554, MSX2P1_B, ODC1_B, OSR2_A, OTX1_B, PLXNC1_B, PRKCB_7570, SCRT2_C, SLC30A10, SPHK2_B, ST8SIA4_B, STX16_C, TRH_A, and TRIM67_B; and

(iii) CD1D, ITPRIPL1, FAM59B, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and C17orf64_B.

In some embodiments of the technology, methods are provided that comprise the following steps:

1) measuring a methylation level for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with a reagent that modifies DNA in a methylation-specific manner (e.g., wherein the reagent is a bisulfite reagent, a methylation-sensitive restriction enzyme, or a methylation-dependent restriction enzyme), wherein the one or more genes is selected from one of the following groups:

(i) BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6;

(ii) MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891,

ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, and MAX.chr11.68622869-68622968;

(iii) ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4;

(iv) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, and C10orf125;

(v) ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125;

(vi) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D; and

(vii) DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, ITPRIPL1;

2) amplifying the treated genomic DNA using a set of primers for the selected one or more genes; and

3) determining the methylation level of the one or more genes by polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and target capture.

In some embodiments of the technology, methods are provided that comprise the following steps:

1) measuring an amount of at least one methylated marker gene in DNA from the sample, wherein the one or more genes is selected from one of the following groups:

(i) BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6;

(ii) MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, and MAX.chr11.68622869-68622968;

(iii) ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4;

(iv) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, and C10orf125;

(v) ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125;

(vi) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D; and

(vii) DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, ITPRIPL1;

2) measuring the amount of at least one reference marker in the DNA; and

3) calculating a value for the amount of the at least one methylated marker gene measured in the DNA as a percentage of the amount of the reference marker gene measured in the DNA, wherein the value indicates the amount of the at least one methylated marker DNA measured in the sample.

In some embodiments of the technology, methods are provided that comprise the following steps:

1) measuring a methylation level of a CpG site for one or more genes in a biological sample of a human individual through treating genomic DNA in the biological sample with bisulfite a reagent capable of modifying DNA in a methylation-specific manner (e.g., a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and a bisulfite reagent);

2) amplifying the modified genomic DNA using a set of primers for the selected one or more genes; and

3) determining the methylation level of the CpG site by methylation-specific PCR, quantitative methylation-specific PCR, methylation-sensitive DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, or bisulfite genomic sequencing PCR;

wherein the one or more genes is selected from one of the following groups:

(i) BHLHE23_C, CALN1_A, CD1D, HOXA7_A, LOC100132891, MAX.chr1.8277479-8277527, MAX.chr15.96889013-96889128, NACAD, ATP6V1B1, BANK1, C17orf64, DLX4, EMX1_A, FOXP4, GP5, ITPRIPL1, LMX1B_A, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, MAX.chr8.124173030-124173395, MPZ, PRKCB, STX16_B UBTF, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, ZSCAN12, BHLHE23_D, CXCL12, KCNK9, OTX1, RIC3, SCRT2_B, MAX.chr17.73073682-73073814, CDH4_E, HNF1B_B, TRH_A, MAX.chr20.1784209-1784461, MAX.chr5.145725410-145725459, MAX.chr5.77268672-77268725, BEST4, and DSCR6;

(ii) MAX.chr15.96889013-96889128, ATP6V1B1, C17orf64, ITPRIPL1, MAX.chr11.14926602-14927148, MAX.chr5.42994866-42994936, LOC100132891, ITPRIPL1, ABLIM1, MAX.chr19.46379903-46380197, COL23A1, LAYN, OTX1, TRH_A, MAX.chr5.145725410-145725459, and MAX.chr11.68622869-68622968;

(iii) ATP6V1B1, MAX.chr11.14926602-14927148, PRKCB, TRH_A, MPZ, GP5, TRIM67, MAX.chr12.4273906-4274012, CALN1_A, MAX.chr12.4273906-4274012, MAX.chr5.42994866-42994936, SCRT2_B, MAX.chr5.145725410-145725459, BHLHE23_D, MAX.chr5.77268672-77268725, EMX1_A, DSCR6, and DLX4;

(iv) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, GP5, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, DSCR6, ITPRIPL1, IGF2BP3_B, DLX4, ABLIM1, BHLHE23_D, ZSCAN12, GRASP, and C10orf125;

(v) ATP6V1B1, LMX1B_A, BANK1, OTX1, ST8SIA4, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_D, ALOX5, MAX.chr19.46379903-46380197, ODC1, CHST2_A, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, CHST2_B, ITPRIPL1, IGF2BP3_B, CDH4_E, ABLIM1, SLC30A10, C10orf125;

(vi) ATP6V1B1, LMX1B_A, BANK1, OTX1, MAX.chr11.14926602-14927148, UBTF, PRKCB, TRH_A, MPZ, DNM3_A, TRIM67, PLXNC1_A, MAX.chr12.4273906-4274012, CALN1_A, ITPRIPL1, MAX.chr12.4273906-4274012, GYPC_B, MAX.chr5.42994866-42994936, OSR2_A, SCRT2_B, MAX.chr5.145725410-145725459, MAX.chr11.68622869-68622968, MAX.chr8.124173030-124173395, MAX.chr20.1784209-1784461, LOC100132891, BHLHE23_C, ALOX5, MAX.chr19.46379903-46380197, CHST2_B, MAX.chr5.77268672-77268725, C17orf64, EMX1_A, DSCR6, ITPRIPL1, IGF2BP3_B, CDH4_E, DLX4, ABLIM1, BHLHE23_D; and

(vii) DSCR6, SCRT2_B, MPZ, MAX.chr8.124173030-124173395, OSR2_A, MAX.chr11.68622869-68622968, ITPRIPL1, MAX.chr5.145725410-145725459, BHLHE23_C, ITPRIPL1.

Preferably, the sensitivity for such methods is from about 70% to about 100%, or from about 80% to about 90%, or from about 80% to about 85%. Preferably, the specificity is from about 70% to about 100%, or from about 80% to about 90%, or from about 80% to about 85%.

Genomic DNA may be isolated by any means, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants, e.g., by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction, or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense, and required quantity of DNA. All clinical sample types comprising neoplastic matter or pre-neoplastic matter are suitable for use in the present method, e.g., cell lines, histological slides, biopsies, paraffin-embedded tissue, body fluids, stool, breast tissue, colonic effluent, urine, blood plasma, blood serum, whole blood, isolated blood cells, cells isolated from the blood, and combinations thereof.

The technology is not limited in the methods used to prepare the samples and provide a nucleic acid for testing. For example, in some embodiments, a DNA is isolated from a stool sample or from blood or from a plasma sample using direct gene capture, e.g., as detailed in U.S. Pat. Appl. Ser. No. 61/485386 or by a related method.

The genomic DNA sample is then treated with at least one reagent, or series of reagents, that distinguishes between methylated and non-methylated CpG dinucleotides within at least one marker comprising a DMR (e.g., DMR 1–375, e.g., as provided by Tables 2 and 18).

In some embodiments, the reagent converts cytosine bases which are unmethylated at the 5′-position to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. However in some embodiments, the reagent may be a methylation sensitive restriction enzyme.

In some embodiments, the genomic DNA sample is treated in such a manner that cytosine bases that are unmethylated at the 5′ position are converted to uracil, thymine, or another base that is dissimilar to cytosine in terms of hybridization behavior. In some embodiments, this treatment is carried out with bisulfite (hydrogen sulfite, disulfite) followed by alkaline hydrolysis.

The treated nucleic acid is then analyzed to determine the methylation state of the target gene sequences (at least one gene, genomic sequence, or nucleotide from a marker comprising a DMR, e.g., at least one DMR chosen from DMR 1–375, e.g., as provided in Tables 2 and 18). The method of analysis may be selected from those known in the art, including those listed herein, e.g., QuARTS and MSP as described herein.

Aberrant methylation, more specifically hypermethylation of a marker comprising a DMR (e.g., DMR 1–375, e.g., as provided by Tables 2 and 18) is associated with a breast cancer.

The technology relates to the analysis of any sample associated with a breast cancer. For example, in some embodiments the sample comprises a tissue and/or biological fluid obtained from a patient. In some embodiments, the sample comprises a secretion. In some embodiments, the sample comprises blood, serum, plasma, gastric secretions, pancreatic juice, a gastrointestinal biopsy sample, microdissected cells from a breast biopsy, and/or cells recovered from stool. In some embodiments, the sample comprises breast tissue. In some embodiments, the subject is human. The sample may include cells, secretions, or tissues from the breast, liver, bile ducts, pancreas, stomach, colon, rectum, esophagus, small intestine, appendix, duodenum, polyps, gall bladder, anus, and/or peritoneum. In some embodiments, the sample comprises cellular fluid, ascites, urine, feces, pancreatic fluid, fluid obtained during endoscopy, blood, mucus, or saliva. In some embodiments, the sample is a stool sample. In some embodiments, the sample is a breast tissue sample.

Such samples can be obtained by any number of means known in the art, such as will be apparent to the skilled person. For instance, urine and fecal samples are easily attainable, while blood, ascites, serum, or pancreatic fluid samples can be obtained parenterally by using a needle and syringe, for instance. Cell free or substantially cell free samples can be obtained by subjecting the sample to various techniques known to those of skill in the art which include, but are not limited to, centrifugation and filtration. Although it is generally preferred that no invasive techniques are used to obtain the sample, it still may be preferable to obtain samples such as tissue homogenates, tissue sections, and biopsy specimens

In some embodiments, the technology relates to a method for treating a patient (e.g., a patient with breast cancer, with early stage breast cancer, or who may develop breast cancer) (e.g., a patient with one or more of triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer), the method comprising determining the methylation state of one or more DMR as provided herein and administering a treatment to the patient based on the results of determining the methylation state. The treatment may be administration of a pharmaceutical compound, a vaccine, performing a surgery, imaging the patient, performing another test. Preferably, said use is in a method of clinical screening, a method of prognosis assessment, a method of monitoring the results of therapy, a method to identify patients most likely to respond to a particular therapeutic treatment, a method of imaging a patient or subject, and a method for drug screening and development.

In some embodiments of the technology, a method for diagnosing a breast cancer in a subject is provided. The terms "diagnosing" and "diagnosis" as used herein refer to methods by which the skilled artisan can estimate and even determine whether or not a subject is suffering from a given disease or condition or may develop a given disease or condition in the future. The skilled artisan often makes a diagnosis on the basis of one or more diagnostic indicators, such as for example a biomarker (e.g., a DMR as disclosed herein), the methylation state of which is indicative of the presence, severity, or absence of the condition.

Along with diagnosis, clinical cancer prognosis relates to determining the aggressiveness of the cancer and the likelihood of tumor recurrence to plan the most effective therapy. If a more accurate prognosis can be made or even a potential risk for developing the cancer can be assessed, appropriate therapy, and in some instances less severe therapy for the patient can be chosen. Assessment (e.g., determining methylation state) of cancer biomarkers is useful to separate subjects with good prognosis and/or low risk of developing cancer who will need no therapy or limited therapy from those more likely to develop cancer or suffer a recurrence of cancer who might benefit from more intensive treatments.

As such, "making a diagnosis" or "diagnosing", as used herein, is further inclusive of determining a risk of developing cancer or determining a prognosis, which can provide for predicting a clinical outcome (with or without medical treatment), selecting an appropriate treatment (or whether treatment would be effective), or monitoring a current treatment and potentially changing the treatment, based on the measure of the diagnostic biomarkers (e.g., DMR) disclosed herein. Further, in some embodiments of the presently disclosed subject

matter, multiple determination of the biomarkers over time can be made to facilitate diagnosis and/or prognosis. A temporal change in the biomarker can be used to predict a clinical outcome, monitor the progression of breast cancer, and/or monitor the efficacy of appropriate therapies directed against the cancer. In such an embodiment for example, one might expect to see a change in the methylation state of one or more biomarkers (e.g., DMR) disclosed herein (and potentially one or more additional biomarker(s), if monitored) in a biological sample over time during the course of an effective therapy.

The presently disclosed subject matter further provides in some embodiments a method for determining whether to initiate or continue prophylaxis or treatment of a cancer in a subject. In some embodiments, the method comprises providing a series of biological samples over a time period from the subject; analyzing the series of biological samples to determine a methylation state of at least one biomarker disclosed herein in each of the biological samples; and comparing any measurable change in the methylation states of one or more of the biomarkers in each of the biological samples. Any changes in the methylation states of biomarkers over the time period can be used to predict risk of developing cancer, predict clinical outcome, determine whether to initiate or continue the prophylaxis or therapy of the cancer, and whether a current therapy is effectively treating the cancer. For example, a first time point can be selected prior to initiation of a treatment and a second time point can be selected at some time after initiation of the treatment. Methylation states can be measured in each of the samples taken from different time points and qualitative and/or quantitative differences noted. A change in the methylation states of the biomarker levels from the different samples can be correlated with breast cancer risk, prognosis, determining treatment efficacy, and/or progression of the cancer in the subject.

In preferred embodiments, the methods and compositions of the invention are for treatment or diagnosis of disease at an early stage, for example, before symptoms of the disease appear. In some embodiments, the methods and compositions of the invention are for treatment or diagnosis of disease at a clinical stage.

As noted, in some embodiments, multiple determinations of one or more diagnostic or prognostic biomarkers can be made, and a temporal change in the marker can be used to determine a diagnosis or prognosis. For example, a diagnostic marker can be determined at an initial time, and again at a second time. In such embodiments, an increase in the marker from the initial time to the second time can be diagnostic of a particular type or severity of cancer, or a given prognosis. Likewise, a decrease in the marker from the initial time to the second

time can be indicative of a particular type or severity of cancer, or a given prognosis. Furthermore, the degree of change of one or more markers can be related to the severity of the cancer and future adverse events. The skilled artisan will understand that, while in certain embodiments comparative measurements can be made of the same biomarker at multiple time points, one can also measure a given biomarker at one time point, and a second biomarker at a second time point, and a comparison of these markers can provide diagnostic information.

As used herein, the phrase "determining the prognosis" refers to methods by which the skilled artisan can predict the course or outcome of a condition in a subject. The term "prognosis" does not refer to the ability to predict the course or outcome of a condition with 100% accuracy, or even that a given course or outcome is predictably more or less likely to occur based on the methylation state of a biomarker (e.g., a DMR). Instead, the skilled artisan will understand that the term "prognosis" refers to an increased probability that a certain course or outcome will occur; that is, that a course or outcome is more likely to occur in a subject exhibiting a given condition, when compared to those individuals not exhibiting the condition. For example, in individuals not exhibiting the condition (e.g., having a normal methylation state of one or more DMR), the chance of a given outcome (e.g., suffering from a breast cancer) may be very low.

In some embodiments, a statistical analysis associates a prognostic indicator with a predisposition to an adverse outcome. For example, in some embodiments, a methylation state different from that in a normal control sample obtained from a patient who does not have a cancer can signal that a subject is more likely to suffer from a cancer than subjects with a level that is more similar to the methylation state in the control sample, as determined by a level of statistical significance. Additionally, a change in methylation state from a baseline (e.g., "normal") level can be reflective of subject prognosis, and the degree of change in methylation state can be related to the severity of adverse events. Statistical significance is often determined by comparing two or more populations and determining a confidence interval and/or a *p* value. See, e.g., Dowdy and Wearden, Statistics for Research, John Wiley & Sons, New York, 1983, incorporated herein by reference in its entirety. Exemplary confidence intervals of the present subject matter are 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% and 99.99%, while exemplary *p* values are 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, and 0.0001.

In other embodiments, a threshold degree of change in the methylation state of a prognostic or diagnostic biomarker disclosed herein (e.g., a DMR) can be established, and the

degree of change in the methylation state of the biamarker in a biological sample is simply compared to the threshold degree of change in the methylation state. A preferred threshold change in the methylation state for biomarkers provided herein is about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 50%, about 75%, about 100%, and about 150%. In yet other embodiments, a "nomogram" can be established, by which a methylation state of a prognostic or diagnostic indicator (biomarker or combination of biomarkers) is directly related to an associated disposition towards a given outcome. The skilled artisan is acquainted with the use of such nomograms to relate two numeric values with the understanding that the uncertainty in this measurement is the same as the uncertainty in the marker concentration because individual sample measurements are referenced, not population averages.

In some embodiments, a control sample is analyzed concurrently with the biological sample, such that the results obtained from the biological sample can be compared to the results obtained from the control sample. Additionally, it is contemplated that standard curves can be provided, with which assay results for the biological sample may be compared. Such standard curves present methylation states of a biomarker as a function of assay units, e.g., fluorescent signal intensity, if a fluorescent label is used. Using samples taken from multiple donors, standard curves can be provided for control methylation states of the one or more biomarkers in normal tissue, as well as for "at-risk" levels of the one or more biomarkers in tissue taken from donors with metaplasia or from donors with a breast cancer. In certain embodiments of the method, a subject is identified as having metaplasia upon identifying an aberrant methylation state of one or more DMR provided herein in a biological sample obtained from the subject. In other embodiments of the method, the detection of an aberrant methylation state of one or more of such biomarkers in a biological sample obtained from the subject results in the subject being identified as having cancer.

The analysis of markers can be carried out separately or simultaneously with additional markers within one test sample. For example, several markers can be combined into one test for efficient processing of a multiple of samples and for potentially providing greater diagnostic and/or prognostic accuracy. In addition, one skilled in the art would recognize the value of testing multiple samples (for example, at successive time points) from the same subject. Such testing of serial samples can allow the identification of changes in marker methylation states over time. Changes in methylation state, as well as the absence of change in methylation state, can provide useful information about the disease status that

includes, but is not limited to, identifying the approximate time from onset of the event, the presence and amount of salvageable tissue, the appropriateness of drug therapies, the effectiveness of various therapies, and identification of the subject's outcome, including risk of future events.

The analysis of biomarkers can be carried out in a variety of physical formats. For example, the use of microtiter plates or automation can be used to facilitate the processing of large numbers of test samples. Alternatively, single sample formats could be developed to facilitate immediate treatment and diagnosis in a timely fashion, for example, in ambulatory transport or emergency room settings.

In some embodiments, the subject is diagnosed as having a breast cancer if, when compared to a control methylation state, there is a measurable difference in the methylation state of at least one biomarker in the sample. Conversely, when no change in methylation state is identified in the biological sample, the subject can be identified as not having breast cancer, not being at risk for the cancer, or as having a low risk of the cancer. In this regard, subjects having the cancer or risk thereof can be differentiated from subjects having low to substantially no cancer or risk thereof. Those subjects having a risk of developing a breast cancer can be placed on a more intensive and/or regular screening schedule, including endoscopic surveillance. On the other hand, those subjects having low to substantially no risk may avoid being subjected to additional testing for breast cancer (e.g., invasive procedure), until such time as a future screening, for example, a screening conducted in accordance with the present technology, indicates that a risk of breast cancer has appeared in those subjects.

As mentioned above, depending on the embodiment of the method of the present technology, detecting a change in methylation state of the one or more biomarkers can be a qualitative determination or it can be a quantitative determination. As such, the step of diagnosing a subject as having, or at risk of developing, a breast cancer indicates that certain threshold measurements are made, e.g., the methylation state of the one or more biomarkers in the biological sample varies from a predetermined control methylation state. In some embodiments of the method, the control methylation state is any detectable methylation state of the biomarker. In other embodiments of the method where a control sample is tested concurrently with the biological sample, the predetermined methylation state is the methylation state in the control sample. In other embodiments of the method, the predetermined methylation state is based upon and/or identified by a standard curve. In other embodiments of the method, the predetermined methylation state is a specifically state or

range of state. As such, the predetermined methylation state can be chosen, within acceptable limits that will be apparent to those skilled in the art, based in part on the embodiment of the method being practiced and the desired specificity, etc.

Further with respect to diagnostic methods, a preferred subject is a vertebrate subject. A preferred vertebrate is warm-blooded; a preferred warm-blooded vertebrate is a mammal. A preferred mammal is most preferably a human. As used herein, the term "subject' includes both human and animal subjects. Thus, veterinary therapeutic uses are provided herein. As such, the present technology provides for the diagnosis of mammals such as humans, as well as those mammals of importance due to being endangered, such as Siberian tigers; of economic importance, such as animals raised on farms for consumption by humans; and/or animals of social importance to humans, such as animals kept as pets or in zoos. Examples of such animals include but are not limited to: carnivores such as cats and dogs; swine, including pigs, hogs, and wild boars; ruminants and/or ungulates such as cattle, oxen, sheep, giraffes, deer, goats, bison, and camels; and horses. Thus, also provided is the diagnosis and treatment of livestock, including, but not limited to, domesticated swine, ruminants, ungulates, horses (including race horses), and the like.

The presently-disclosed subject matter further includes a system for diagnosing a breast cancer and/or a specific form of breast cancer (e.g., triple negative breast cancer, $HER2^+$ breast cancer, Luminal A breast cancer, Luminal B breast cancer, BRCA1 breast cancer, BRCA2 breast cancer) in a subject. The system can be provided, for example, as a commercial kit that can be used to screen for a risk of breast cancer or diagnose a breast cancer in a subject from whom a biological sample has been collected. An exemplary system provided in accordance with the present technology includes assessing the methylation state of a DMR as provided in Tables 2 and 18.

## EXAMPLES

### Example I.

This example describes the discovery and tissue validation of breast-cancer specific markers.

Table 1 shows the number of tissue samples for each subtype of breast cancer used in the discovery of breast cancer specific markers.

**Table 1.**

| Breast Cancer Subtype | Number of Subjects | Total |
|---|---|---|
| Basal-like / Triple Negative | 18 | 18 |
| $HER2^+$ | 18 | 18 |
| Luminal A | 18 | 18 |
| Luminal B | 18 | 18 |
| BRCA 1 | 6 | 15 |
| BRCA 2 | 9 | |
| Normal Breast | 18 | 45 |
| Normal Breast + BRCA | 9 | |
| Normal Buffy Coat | 18 | |

For discovery of methylation markers by RRBS, frozen tissue samples were obtained from 72 invasive breast cancer cases (18 luminal A, 18 luminal B, 18 basal-like/triple negative, and 18 HER2+), 15 invasive breast cancer from BRCA germline mutation patients (6 BRCA1, 9 BRCA2), and 45 controls (18 normal breast (reduction mammoplasty or prophylactic mastectomy, 9 histologically normal breast in germline BRCA carriers (prophylactic mastectomy), and 18 normal buffy coat)). Tumor and breast tissue sections were reviewed by an expert GI pathologist to confirm diagnosis and estimate abnormal cellularity. Sections were then macro-dissected. Genomic DNA was purified using the QiaAmp Mini kit (Qiagen, Valencia CA). DNA (300 ng) was fragmented by digestion with 10 Units of MspI. Digested fragments were end-repaired and A-tailed with 5 Units of Klenow fragment (3'-5' exo-), and ligated overnight to methylated TruSeq adapters (Illumina, San Diego CA) containing barcode sequences (to link each fragment to its sample ID.) Reactions were purified using AMPure XP SPRI beads/buffer (Beckman Coulter, Brea CA).

Tissue samples then underwent bisulfite conversion (twice) using a modified EpiTect protocol (Qiagen). qPCR (LightCycler 480 – Roche, Mannheim Germany) was used to determine the optimal enrichment Ct. The following conditions were used for final enrichment PCR: Each 50uL reaction contained 5uL of 10X buffer, 1.25uL of 10 mM each deoxyribonucleotide triphosphate (dNTP), 5uL primer cocktail (~5uM), 15uL template (sample), 1uL PfuTurbo Cx hotstart (Agilent, Santa Clara CA) and 22.75 water; temperatures and times were 95C-5min; 98C-30sec; 16 cycles of 98C-10sec, 65C-30sec, 72C-30sec, 72C-5min and 4C hold, respectively. Samples were SPRI bead purified and then tested on the Bioanalyzer 2100 (Agilent) to assess the DNA size distribution of the enrichment. Size

selection of 160-520bp fragments (40-400 bp inserts) was performed using AMPure XP SPRI beads/buffer (Beckman Coulter, Brea CA). Buffer cutoffs were 0.7X - 1.1X sample volumes. Samples were combined (equimolar) into 4-plex libraries based on the randomization scheme and tested with the bioanalyzer for final size and concentration verification, and with qPCR (KAPA Library Quantification Kit – KAPA Biosystems, Cape Town South Africa).

Tissue samples were loaded onto single read flow cells according to a randomized lane assignment and sequencing was performed by the Next Generation Sequencing Core at the Mayo Clinic Medical Genome Facility on the Illumina HiSeq 2000 platform. Reads were unidirectional for 101 cycles. The standard Illumina pipeline was run for the primary analysis. SAAP-RRBS (streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing) was used for quality scoring, sequence alignment, annotation, and methylation extraction.

Breast cancer tissue yielded large numbers of discriminate DMRs, many of which had not been identified before. Comparing the methylation of breast cancer tissue samples to normal breast tissue, 327 methylated regions were identified (see, Table 2) that distinguished breast cancer tissue from normal breast tissue (the genomic coordinates for the regions shown in Table 2 are based on the Human Feb. 2009 (GRCh37/hg19) Assembly). Table 3 shows 48 methylated regions that distinguished triple negative breast cancer tissue from normal breast tissue. Table 4 shows 122 methylated regions that distinguished $HER2^+$ breast cancer tissue from normal breast tissue. Table 5 shows 75 methylated regions that distinguished Luminal A breast cancer tissue from normal breast tissue. Table 6 shows 39 methylated regions that distinguished Luminal B breast cancer tissue from normal breast tissue. Table 7 shows 49 methylated regions that distinguished BRCA1 breast cancer tissue from normal breast tissue. Table 8 shows 45 methylated regions that distinguished BRCA2 breast cancer tissue from normal breast tissue. Table 9 shows 21 methylated regions that distinguished invasive breast cancer tissue from normal breast tissue.

**Table 2.** Identified methylated regions distinguishing breast cancer tissue from normal breast tissue.

| DMR No. | Gene Annotation | Region on Chromosome (starting base-ending base) |
|---|---|---|
| 1 | ZSCAN23 | chr6:28411152-28411272 |

| | | |
|---|---|---|
| 2 | AADAT.R | chr4:171010951-171010991 |
| 3 | ABLIM1 | chr10:116391588-116391793 |
| 4 | ACCN1 | chr17:31620207-31620314 |
| 5 | AFAP1L1 | chr5:148651161-148651242 |
| 6 | AJAP1_A | chr1:4715535-4715646 |
| 7 | AJAP1_B | chr1:4715931-4716021 |
| 8 | AKR1B1 | chr7:134143171-134143684 |
| 9 | ALOX5 | chr10:45914840-45914949 |
| 10 | AMN | chr14:103394920-103395019 |
| 11 | ANPEP | chr15:90358420-90358514 |
| 12 | ANTXR2 | chr4:80993475-80993634 |
| 13 | ARL5C | chr17:37321515-37321626 |
| 14 | ASCL2 | chr11:2292240-2292361 |
| 15 | ATP6V1B1 | chr2:71192354-71192453 |
| 16 | B3GNT5 | chr3:182971589-182971825 |
| 17 | BANK1 | chr4:102711871-102712076 |
| 18 | BCAT1 | chr12:25055906-25055975 |
| 19 | BEGAIN | chr14:101033665-101033813 |
| 20 | BEST4 | chr1:45251853-45252029 |
| 21 | BHLHE23_A | chr20:61637950-61637986 |
| 22 | BHLHE23_B | chr20:61638020-61638083 |
| 23 | BHLHE23_C | chr20:61638088-61638565 |
| 24 | BHLHE23_D | chr20:61638244-61638301 |
| 25 | BMP4 | chr14:54421578-54421916 |
| 26 | BMP6 | chr6:7727566-7727907 |
| 27 | C10orf125 | chr10:135171410-135171504 |
| 28 | C10orf93 | chr10:134756078-134756167 |
| 29 | C17orf64 | chr17:58499095-58499190 |
| 30 | C19orf35 | chr19:2282568-2282640 |
| 31 | C19orf66 | chr19:10197688-10197823 |
| 32 | C1QL2 | chr2:119916511-119916572 |
| 33 | C20orf195_A | chr20:62185293-62185364 |

| | | |
|---|---|---|
| 34 | C20orf195_B | chr20:62185418-62185546 |
| 35 | C7orf52 | chr7:100823483-100823514 |
| 36 | CALN1_A | chr7:71801486-71801594 |
| 37 | CALN1_B | chr7:71801741-71801800 |
| 38 | CAMKV | chr3:49907259-49907298 |
| 39 | CAPN2.FR | chr1:223900347-223900405 |
| 40 | CAV2 | chr7:116140205-116140342 |
| 41 | CBLN1_A | chr16:49315588-49315691 |
| 42 | CBLN1_B | chr16:49316198-49316258 |
| 43 | CCDC61 | chr19:46519467-46519536 |
| 44 | CCND2_A | chr12:4378317-4378375 |
| 45 | CCND2_B | chr12:4380560-4380681 |
| 46 | CCND2_C | chr12:4384096-4384146 |
| 47 | CD1D | chr1:158150864-158151129 |
| 48 | CD8A | chr2:87017780-87017917 |
| 49 | CDH4_A | chr20:59827230-59827285 |
| 50 | CDH4_B | chr20:59827762-59827776 |
| 51 | CDH4_C | chr20:59827794-59827868 |
| 52 | CDH4_D | chr20:59828193-59828258 |
| 53 | CDH4_E | chr20:59828479-59828729 |
| 54 | CDH4_F | chr20:59828778-59828814 |
| 55 | CHRNA7 | chr15:32322830-32322897 |
| 56 | CHST2_A | chr3:142838025-142838494 |
| 57 | CHST2_B | chr3:142839223-142839568 |
| 58 | CLIC6 | chr21:36042025-36042131 |
| 59 | CLIP4 | chr2:29338109-29338339 |
| 60 | COL23A1.R | chr5:178017669-178017854 |
| 61 | CR1 | chr1:207669481-207669639 |
| 62 | CRHBP | chr5:76249939-76249997 |
| 63 | CXCL12.F | chr10:44881210-44881300 |
| 64 | DBNDD1.FR | chr16:90085625-90085681 |
| 65 | DLK1 | chr14:101193295-101193318 |

| | | |
|---|---|---|
| 66 | DLX4 | chr17:48042562-48042606 |
| 67 | DLX6 | chr7:96635255-96635475 |
| 68 | DNAJC6 | chr1:65731412-65731507 |
| 69 | DNM3_A | chr1:171810393-171810575 |
| 70 | DNM3_B | chr1:171810648-171810702 |
| 71 | DNM3_C | chr1:171810806-171810920 |
| 72 | DSCR6 | chr21:38378540-38378601 |
| 73 | DTX1 | chr12:113515535-113515637 |
| 74 | EMX1_A | chr2:73151498-73151578 |
| 75 | EMX1_B | chr2:73151663-73151756 |
| 76 | EPHA4 | chr2:222436217-222436320 |
| 77 | ESPN | chr1:6508784-6509175 |
| 78 | ESYT3 | chr3:138153979-138154071 |
| 79 | ETS1_A | chr11:128391809-128391908 |
| 80 | ETS1_B | chr11:128392062-128392309 |
| 81 | FABP5 | chr8:82192605-82192921 |
| 82 | FAIM2 | chr12:50297863-50297988 |
| 83 | FAM126A | chr7:23053941-23054066 |
| 84 | FAM129C.F | chr19:17650551-17650610 |
| 85 | FAM150A | chr8:53478266-53478416 |
| 86 | FAM150B | chr2:287868-287919 |
| 87 | FAM171A1 | chr10:15412558-15412652 |
| 88 | FAM189A1 | chr15:29862130-29862169 |
| 89 | FAM20A | chr17:66597237-66597326 |
| 90 | FAM59B | chr2:26407713-26407972 |
| 91 | FBN1 | chr15:48937412-48937541 |
| 92 | FLJ42875 | chr1:2987037-2987116 |
| 93 | FLRT2 | chr14:85998469-85998535 |
| 94 | FMN2 | chr1:240255171-240255253 |
| 95 | FMNL2 | chr2:153192734-153192836 |
| 96 | FOXP4 | chr6:41528816-41528958 |
| 97 | FSCN1 | chr7:5633506-5633615 |

| | | |
|---|---|---|
| 98 | GAD2 | chr10:26505066-26505385 |
| 99 | GAS7 | chr17:10101325-10101397 |
| 100 | GCGR | chr17:79761970-79762088 |
| 101 | GLI3 | chr7:42267808-42267899 |
| 102 | GLP1R | chr6:39016381-39016421 |
| 103 | GNG4 | chr1:235813658-235813798 |
| 104 | GP5 | chr3:194118738-194118924 |
| 105 | GRASP | chr12:52400919-52401166 |
| 106 | GRM7 | chr3:6902873-6902931 |
| 107 | GSTP1 | chr11:67350986-67351055 |
| 108 | GYPC_A | chr2:127413505-127413678 |
| 109 | GYPC_B | chr2:127414096-127414189 |
| 110 | HAND2 | chr4:174450452-174450478 |
| 111 | HBM | chr16:216426-216451 |
| 112 | HES5 | chr1:2461823-2461915 |
| 113 | HHEX.F | chr10:94449486-94449597 |
| 114 | HMGA2 | chr12:66219385-66219487 |
| 115 | HNF1B_A | chr17:36103713-36103793 |
| 116 | HNF1B_B | chr17:36105390-36105448 |
| 117 | HOXA1_A | chr7:27135603-27135889 |
| 118 | HOXA1_B | chr7:27136191-27136244 |
| 119 | HOXA7_A | chr7:27195742-27195895 |
| 120 | HOXA7_B | chr7:27196032-27196190 |
| 121 | HOXA7_C | chr7:27196441-27196531 |
| 122 | HOXD9 | chr2:176987716-176987739 |
| 123 | IGF2BP3_A | chr7:23508901-23509225 |
| 124 | IGF2BP3_B | chr7:23513817-23514114 |
| 125 | IGFBP5 | chr2:217559103-217559244 |
| 126 | IGSF9B_A | chr11:133825409-133825476 |
| 127 | IGSF9B_B | chr11:133825491-133825530 |
| 128 | IL15RA | chr10:6018610-6018848 |
| 129 | IL17REL | chr22:50453462-50453555 |

| | | |
|---|---|---|
| 130 | INSM1 | chr20:20348140-20348182 |
| 131 | ITGA9 | chr3:37493895-37493994 |
| 132 | ITPKA_A | chr15:41787438-41787784 |
| 133 | ITPKA_B | chr15:41793928-41794003 |
| 134 | ITPRIPL1 | chr2:96990968-96991328 |
| 135 | JSRP1 | chr19:2253163-2253376 |
| 136 | KCNA1 | chr12:5019401-5019633 |
| 137 | KCNE3 | chr11:74178260-74178346 |
| 138 | KCNH8 | chr3:19189837-19189897 |
| 139 | KCNK17_A | chr6:39281195-39281282 |
| 140 | KCNK17_B | chr6:39281408-39281478 |
| 141 | KCNK9.FR | chr8:140715096-140715164 |
| 142 | KCNQ2 | chr20:62103558-62103625 |
| 143 | KIAA1949 | chr6:30646976-30647084 |
| 144 | KIRREL2 | chr19:36347825-36347863 |
| 145 | KLF16 | chr19:1857330-1857476 |
| 146 | KLHDC7B | chr22:50987219-50987304 |
| 147 | LAYN.R | chr11:111412023-111412074 |
| 148 | LIME1 | chr20:62369116-62369393 |
| 149 | LMX1B_A | chr9:129388175-129388223 |
| 150 | LMX1B_B | chr9:129388231-129388495 |
| 151 | LMX1B_C | chr9:129445588-129445603 |
| 152 | LOC100131176 | chr7:151106986-151107060 |
| 153 | LOC100132891 | chr8:72755897-72756295 |
| 154 | LOC100302401.R | chr1:178063509-178063567 |
| 155 | LOC283999 | chr17:76227905-76227960 |
| 156 | LRRC34 | chr3:169530006-169530139 |
| 157 | LSS.F | chr21:47649525-47649615 |
| 158 | LY6H | chr8:144241547-144241557 |
| 159 | MAGI2 | chr7:79083359-79083600 |
| 160 | MAST1 | chr19:12978399-12978642 |
| 161 | MAX.chr1.158083198-158083476 | chr1:158083198-158083476 |

| | | |
|---|---|---|
| 162 | MAX.chr1.228074764-228074977 | chr1:228074764-228074977 |
| 163 | MAX.chr1.239549742-239549886 | chr1:239549742-239549886 |
| 164 | MAX.chr1.46913931-46913950 | chr1:46913931-46913950 |
| 165 | MAX.chr1.8277285-8277316 | chr1:8277285-8277316 |
| 166 | MAX.chr1.8277479-8277527 | chr1:8277479-8277527 |
| 167 | MAX.chr10.130085265-130085312 | chr10:130085265-130085312 |
| 168 | MAX.chr11.14926602-14927148 | chr11:14926602-14927148 |
| 169 | MAX.chr11.68622869-68622968 | chr11:68622869-68622968 |
| 170 | MAX.chr12.4273906-4274012 | chr12:4273906-4274012 |
| 171 | MAX.chr12.59990591-59990895 | chr12:59990591-59990895 |
| 172 | MAX.chr14.101176106-101176260 | chr14:101176106-101176260 |
| 173 | MAX.chr15.96889013-96889128 | chr15:96889013-96889128 |
| 174 | MAX.chr17.73073682-73073814 | chr17:73073682-73073814 |
| 175 | MAX.chr17.8230197-8230314 | chr17:8230197-8230314 |
| 176 | MAX.chr18.5629721-5629791 | chr18:5629721-5629791 |
| 177 | MAX.chr18.76734362-76734476 | chr18:76734362-76734476 |
| 178 | MAX.chr19.30719261-30719354 | chr19:30719261-30719354 |
| 179 | MAX.chr19.46379903-46380197 | chr19:46379903-46380197 |
| 180 | MAX.chr2.223183057-223183114.FR | chr2:223183057-223183114 |
| 181 | MAX.chr2.238864674-238864735 | chr2:238864674-238864735 |
| 182 | MAX.chr2.97193163-97193287 | chr2:97193163-97193287 |
| 183 | MAX.chr2.97193478-97193562 | chr2:97193478-97193562 |
| 184 | MAX.chr20.1783841-1784054 | chr20:1783841-1784054 |
| 185 | MAX.chr20.1784209-1784461 | chr20:1784209-1784461 |
| 186 | MAX.chr21.44782441-44782498 | chr21:44782441-44782498 |
| 187 | MAX.chr21.47063802-47063851 | chr21:47063802-47063851 |
| 188 | MAX.chr22.23908718-23908782 | chr22:23908718-23908782 |
| 189 | MAX.chr22.42679578-42679917 | chr22:42679578-42679917 |
| 190 | MAX.chr4.8859253-8859329 | chr4:8859253-8859329 |
| 191 | MAX.chr4.8859602-8859669 | chr4:8859602-8859669 |
| 192 | MAX.chr4.8860002-8860038 | chr4:8860002-8860038 |
| 193 | MAX.chr5.145725410-145725459 | chr5:145725410-145725459 |

| | | |
|---|---|---|
| 194 | MAX.chr5.172234248-172234494 | chr5:172234248-172234494 |
| 195 | MAX.chr5.178957564-178957598 | chr5:178957564-178957598 |
| 196 | MAX.chr5.180101084-180101094 | chr5:180101084-180101094 |
| 197 | MAX.chr5.42952185-42952280 | chr5:42952185-42952280 |
| 198 | MAX.chr5.42994866-42994936 | chr5:42994866-42994936 |
| 199 | MAX.chr5.77268672-77268725 | chr5:77268672-77268725 |
| 200 | MAX.chr5.81148300-81148332 | chr5:81148300-81148332 |
| 201 | MAX.chr6.108440684-108440788 | chr6:108440684-108440788 |
| 202 | MAX.chr6.130686865-130686985 | chr6:130686865-130686985 |
| 203 | MAX.chr6.157556793-157556856 | chr6:157556793-157556856 |
| 204 | MAX.chr6.157557371-157557657 | chr6:157557371-157557657 |
| 205 | MAX.chr6.27064703-27064783 | chr6:27064703-27064783 |
| 206 | MAX.chr7.151145632-151145743 | chr7:151145632-151145743 |
| 207 | MAX.chr7.152622607-152622638 | chr7:152622607-152622638 |
| 208 | MAX.chr8.124173030-124173395 | chr8:124173030-124173395 |
| 209 | MAX.chr8.124173128-124173268 | chr8:124173128-124173268 |
| 210 | MAX.chr8.143533298-143533558 | chr8:143533298-143533558 |
| 211 | MAX.chr8.145104132-145104218 | chr8:145104132-145104218 |
| 212 | MAX.chr8.687688-687736 | chr8:687688-687736 |
| 213 | MAX.chr8.688863-688924 | chr8:688863-688924 |
| 214 | MAX.chr9.114010-114207 | chr9:114010-114207 |
| 215 | MAX.chr9.136474504-136474527 | chr9:136474504-136474527 |
| 216 | MCF2L2 | chr3:182896930-182897245 |
| 217 | MERTK | chr2:112656676-112656744 |
| 218 | MGAT1 | chr5:180230434-180230767 |
| 219 | MIB2 | chr1:1565891-1565987 |
| 220 | MN1 | chr22:28197962-28198388 |
| 221 | MPZ | chr1:161275561-161275996 |
| 222 | MSX2P1 | chr17:56234436-56234516 |
| 223 | NACAD | chr7:45128502-45128717 |
| 224 | NID2_A | chr14:52535260-52535353 |
| 225 | NID2_B | chr14:52535974-52536161 |

| | | |
|---|---|---|
| 226 | NID2_C | chr14:52536192-52536328 |
| 227 | NKX2-6 | chr8:23564115-23564146 |
| 228 | NR2F6 | chr19:17346428-17346459 |
| 229 | NTRK3 | chr15:88800287-88800414 |
| 230 | NXPH4 | chr12:57618904-57618944 |
| 231 | ODC1 | chr2:10589075-10589243 |
| 232 | OLIG3_A | chr6:137818896-137818917 |
| 233 | OLIG3_B | chr6:137818978-137818988 |
| 234 | OSR2_A | chr8:99952233-99952366 |
| 235 | OSR2_B | chr8:99952801-99952919 |
| 236 | OSR2_C | chr8:99960580-99960630 |
| 237 | OTX1.R | chr2:63281481-63281599 |
| 238 | PAQR6 | chr1:156215470-156215739 |
| 239 | PCDH8 | chr13:53421299-53421322 |
| 240 | PDX1 | chr13:28498503-28498544 |
| 241 | PDXK_A | chr21:45148429-45148556 |
| 242 | PDXK_B | chr21:45148575-45148681 |
| 243 | PEAR1 | chr1:156863318-156863493 |
| 244 | PIF1 | chr15:65116285-65116597 |
| 245 | PLXNC1_A | chr12:94544327-94544503 |
| 246 | PLXNC1_B | chr12:94544333-94544426 |
| 247 | POU4F1 | chr13:79177505-79177532 |
| 248 | PPARA | chr22:46545328-46545457 |
| 249 | PPARG | chr3:12330042-12330152 |
| 250 | PPP1R16B_A | chr20:37435507-37435716 |
| 251 | PPP1R16B_B | chr20:37435738-37435836 |
| 252 | PPP2R5C | chr14:102247681-102247929 |
| 253 | PRDM13_A | chr6:100061616-100061742 |
| 254 | PRDM13_B | chr6:100061748-100061792 |
| 255 | PRHOXNB | chr13:28552424-28552562 |
| 256 | PRKCB | chr16:23847575-23847699 |
| 257 | PRMT1 | chr19:50179501-50179635 |

| | | |
|---|---|---|
| 258 | PROM1 | chr4:16084793-16085112 |
| 259 | PTPRM | chr18:7568565-7568808 |
| 260 | PTPRN2 | chr7:157483341-157483429 |
| 261 | RASGRF2 | chr5:80256117-80256162 |
| 262 | RBFOX3_A | chr17:77179579-77179752 |
| 263 | RBFOX3_B | chr17:77179778-77180064 |
| 264 | RFX8 | chr2:102090934-102091130 |
| 265 | RGS17 | chr6:153452120-153452393 |
| 266 | RIC3.F | chr11:8190622-8190711 |
| 267 | RIPPLY2 | chr6:84563228-84563287 |
| 268 | RYR2_A | chr1:237205369-237205428 |
| 269 | RYR2_B | chr1:237205619-237205640 |
| 270 | SALL3 | chr18:76739321-76739404 |
| 271 | SBNO2 | chr19:1131795-1131992 |
| 272 | SCRT2_A | chr20:644533-644618 |
| 273 | SCRT2_B | chr20:644573-644618 |
| 274 | SERPINB9_A | chr6:2902941-2902998 |
| 275 | SERPINB9_B | chr6:2903031-2903143 |
| 276 | SLC16A3.F | chr17:80189895-80189962 |
| 277 | SLC22A20.FR | chr11:64993239-64993292 |
| 278 | SLC2A2 | chr3:170746149-170746208 |
| 279 | SLC30A10 | chr1:220101458-220101634 |
| 280 | SLC7A4 | chr22:21386780-21386831 |
| 281 | SLC8A3 | chr14:70654596-70654640 |
| 282 | SLITRK5.R | chr13:88329960-88330076 |
| 283 | SNCA | chr4:90758071-90758118 |
| 284 | SPHK2 | chr19:49127580-49127683 |
| 285 | ST8SIA4 | chr5:100240059-100240276 |
| 286 | STAC2_A | chr17:37381217-37381303 |
| 287 | STAC2_B | chr17:37381689-37381795 |
| 288 | STX16_A | chr20:57224798-57224975 |
| 289 | STX16_B | chr20:57225077-57225227 |

| | | |
|---|---|---|
| 290 | SYN2 | chr3:12045894-12045967 |
| 291 | SYNJ2 | chr6:158402213-158402536 |
| 292 | SYT5 | chr19:55690401-55690496 |
| 293 | TAL1 | chr1:47697702-47697882 |
| 294 | TBKBP1 | chr17:45772630-45772726 |
| 295 | TBX1 | chr22:19754257-19754550 |
| 296 | TEPP | chr16:58018790-58018831 |
| 297 | TIMP2 | chr17:76921762-76921779 |
| 298 | TLX1NB | chr10:102881178-102881198 |
| 299 | TMEFF2 | chr2:193060012-193060126 |
| 300 | TMEM176A | chr7:150497411-150497535 |
| 301 | TNFRSF10D | chr8:23020896-23021114 |
| 302 | TOX | chr8:60030723-60030754 |
| 303 | TRH_A | chr3:129693484-129693575 |
| 304 | TRH_B | chr3:129694457-129694501 |
| 305 | TRIM67 | chr1:231297047-231297159 |
| 306 | TRIM71_A | chr3:32858861-32858897 |
| 307 | TRIM71_B | chr3:32859445-32859559 |
| 308 | TRIM71_C | chr3:32860020-32860090 |
| 309 | TSHZ3 | chr19:31839809-31840038 |
| 310 | UBTF | chr17:42287924-42288018 |
| 311 | ULBP1 | chr6:150285563-150285661 |
| 312 | USP44_A | chr12:95942148-95942178 |
| 313 | USP44_B | chr12:95942519-95942558 |
| 314 | UTF1 | chr10:135044125-135044171 |
| 315 | UTS2R | chr17:80329497-80329534 |
| 316 | VIPR2 | chr7:158937370-158937481 |
| 317 | VN1R2 | chr19:53758121-53758147 |
| 318 | VSNL1 | chr2:17720216-17720257 |
| 319 | VSTM2B_A | chr19:30016283-30016357 |
| 320 | VSTM2B_B | chr19:30017789-30018165 |
| 321 | ZBTB16 | chr11:113929882-113930166 |

| | | |
|---|---|---|
| 322 | ZFP64 | chr20:50721057-50721235 |
| 323 | ZNF132 | chr19:58951402-58951775 |
| 324 | ZNF486 | chr19:20278004-20278145 |
| 325 | ZNF626 | chr19:20844070-20844199 |
| 326 | ZNF671 | chr19:58238810-58238955 |
| 327 | ZSCAN12 | chr6:28367128-28367509 |

**Table 3.** Table 3 shows 1) area under the curve for identified methylated regions distinguishing triple negative breast cancer tissue from normal breast tissue, 2) the Fold Change (FC) for triple negative breast cancer tissue vs. normal breast tissue, and 3) the Fold Change (FC) for triple negative breast cancer tissue vs. buffy coat (normal).

| Gene Annotation | Region on Chromosome (starting base-ending base) | AUC | FC Tissue | FC Buffy | DMR |
|---|---|---|---|---|---|
| ABLIM1 | chr10:116391588-116391793 | 0.821 | 9.187127 | 6.095762 | 3 |
| AJAP1_B | chr1:4715931-4716021 | 0.9358 | 33.64347 | 55.12195 | 7 |
| ASCL2 | chr11:2292240-2292361 | 0.9479 | 21.93763 | 13.66322 | 14 |
| ATP6V1B1 | chr2:71192354-71192453 | 1 | 2.711325 | 92.96954 | 15 |
| BANK1 | chr4:102711871-102712076 | 1 | 1.43525 | 61.96732 | 17 |
| CALN1_A | chr7:71801486-71801594 | 0.9346 | 17.12541 | 17.09291 | 36 |
| CALN1_B | chr7:71801741-71801800 | 0.8742 | 19.87087 | 36.73595 | 37 |
| CLIC6 | chr21:36042025-36042131 | 1 | 3.059621 | 93.40228 | 58 |
| DSCR6 | chr21:38378540-38378601 | 0.9641 | 16.80192 | 23.82779 | 72 |
| FOXP4 | chr6:41528816-41528958 | 1 | 1.645757 | $>1x10^6$ | 96 |
| GAD2 | chr10:26505066-26505385 | 0.9134 | 29.30865 | 44.14014 | 98 |
| GCGR | chr17:79761970-79762088 | 0.9506 | 15.57835 | 9.860312 | 100 |
| GP5 | chr3:194118738-194118924 | 0.9961 | 1.942734 | 122.3962 | 104 |
| GRASP | chr12:52400919-52401166 | 0.9506 | 34.60851 | 58.57791 | 105 |
| HBM | chr16:216426-216451 | 0.9389 | 28.44886 | 28.4872 | 111 |
| HNF1B_B | chr17:36105390-36105448 | 1 | 2.492725 | 20.57995 | 116 |
| KLF16 | chr19:1857330-1857476 | 0.8785 | 19.65243 | 148.6852 | 145 |
| MAGI2 | chr7:79083359-79083600 | 0.9306 | 16.79564 | 5.734084 | 159 |
| MAX.chr11.14926602-14927148 | chr11:14926602-14927148 | 1 | 3.891519 | 87.49446 | 168 |
| MAX.chr12.4273906-4274012 | chr12:4273906-4274012 | 0.9815 | 20.08783 | 149.5817 | 170 |
| MAX.chr17.73073 | chr17:73073682-73073814 | 0.9883 | 1.679173 | 72.85714 | 174 |

| | | | | | |
|---|---|---|---|---|---|
| 682-73073814 | | | | | |
| MAX.chr18.76734362-76734370 | chr18:76734362-76734370 | 0.9641 | 24.69328 | 77.26996 | 177 |
| MAX.chr2.97193478-97193562 | chr2:97193478-97193562 | 0.9167 | 22.01754 | 119.8408 | 183 |
| MAX.chr22.42679578-42679917 | chr22:42679578-42679917 | 0.9375 | 27.34823 | 20.78761 | 189 |
| MAX.chr4.8859253-8859329 | chr4:8859253-8859329 | 0.9346 | 13.7246 | 93.86646 | 190 |
| MAX.chr4.8859602-8859669 | chr4:8859602-8859669 | 0.9632 | 11.5 | 27.44798 | 191 |
| MAX.chr4.8860002-8860038 | chr4:8860002-8860038 | 0.9491 | 20.16179 | 84.79759 | 192 |
| MAX.chr5.145725410-145725459 | chr5:145725410-145725459 | 0.933 | 12.88169 | 25.65149 | 193 |
| MAX.chr6.157557371-157557657 | chr6:157557371-157557657 | 1 | 6.19614 | 35.10826 | 204 |
| MPZ | chr1:161275561-161275996 | 0.9504 | 20.73901 | 191.2216 | 221 |
| NKX2-6 | chr8:23564115-23564146 | 0.9583 | 18.63167 | 38.06928 | 227 |
| PDX1 | chr13:28498503-28498544 | 0.9657 | 18.77193 | 64.6598 | 240 |
| PLXNC1_A | chr12:94544327-94544503 | 0.9449 | 4.617089 | 38.86521 | 245 |
| PPARG | chr3:12330042-12330152 | 0.9259 | 30.22681 | 10.42603 | 249 |
| PRKCB | chr16:23847575-23847699 | 0.9281 | 20.45208 | 295.1076 | 256 |
| PTPRN2 | chr7:157483341-157483429 | 0.9281 | 12.35294 | 32.67167 | 260 |
| RBFOX3_A | chr17:77179579-77179752 | 0.9074 | 19.19924 | 18.24275 | 262 |
| SCRT2_A | chr20:644533-644618 | 0.9321 | 18.30644 | 7.92126 | 272 |
| SLC7A4 | chr22:21386780-21386831 | 0.9792 | 17.60673 | 23.649 | 280 |
| STAC2_B | chr17:37381689-37381795 | 0.9074 | 26.95157 | 73.07841 | 287 |
| STX16_A | chr20:57224798-57224975 | 1 | 1.36278 | 106.6599 | 288 |
| STX16_B | chr20:57225077-57225227 | 1 | 1.593456 | 198.3707 | 289 |
| TBX1 | chr22:19754257-19754550 | 0.8676 | 1.385844 | 35.45752 | 295 |
| TRH_A | chr3:129693484-129693575 | 1 | 3.188452 | 67.015 | 303 |
| VSTM2B_A | chr19:30016283-30016357 | 0.9246 | 27.51997 | 28.83311 | 319 |
| ZBTB16 | chr11:113929882-113930166 | 0.9003 | 18.82877 | 26.23126 | 321 |
| ZNF132 | chr19:58951402-58951775 | 0.9062 | 33.99015 | 85.56548 | 323 |
| ZSCAN23 | chr6:28411152-28411272 | 0.9163 | 20.33657 | 59.21927 | 1 |

**Table 4.** Table 4 shows 1) area under the curve for identified methylated regions distinguishing $HER2^+$ breast cancer tissue from normal breast tissue, 2) the Fold Change (FC) for $HER2^+$ breast cancer tissue vs. normal breast tissue, and 3) the Fold Change (FC) for $HER2^+$ breast cancer tissue vs. buffy coat (normal).

| Gene Annotation | Region on Chromosome (starting base-ending base) | AUC | FC Tissue | FC Buffy | DMR |
|---|---|---|---|---|---|
| ABLIM1 | chr10:116391588-116391793 | 0.9846 | 20.96881 | 13.91304 | 3 |
| AFAP1L1 | chr5:148651161-148651242 | 0.9902 | 19.53202 | 21.58802 | 5 |
| AKR1B1 | chr7:134143171-134143684 | 0.9537 | 31.02981 | 91.43744 | 8 |
| ALOX5 | chr10:45914840-45914949 | 0.9522 | 30.50987 | 110.198 | 9 |
| AMN | chr14:103394920-103395019 | 0.951 | 23.99824 | 172.1942 | 10 |
| ARL5C | chr17:37321515-37321626 | 0.9526 | 23.90438 | 76.38447 | 13 |
| BANK1 | chr4:102711871-102712076 | 1 | 1.178479 | 50.88113 | 17 |
| BCAT1 | chr12:25055906-25055975 | 0.9641 | 17.75806 | 73.18046 | 18 |
| BEGAIN | chr14:101033665-101033813 | 1 | 25.33593 | 29.59147 | 19 |
| BEST4 | chr1:45251853-45252029 | 0.9753 | 40.07491 | 76.68804 | 20 |
| BHLHE23_B | chr20:61638020-61638083 | 0.9765 | 33.01942 | 26.19353 | 22 |
| BHLHE23_C | chr20:61638088-61638565 | 0.9938 | 37.3359 | 51.54664 | 23 |
| C17orf64 | chr17:58499095-58499190 | 0.9583 | 19.9771 | 281.2989 | 29 |
| C1QL2 | chr2:119916511-119916572 | 1 | 20.92054 | 20.81967 | 32 |
| C7orf52 | chr7:100823483-100823514 | 0.9967 | 34.66759 | 20.82363 | 35 |

| | | | | | |
|---|---|---|---|---|---|
| CALN1_B | chr7:71801741-71801800 | 0.9444 | 17.5999 | 32.5375 | 37 |
| CAV2 | chr7:116140205-116140342 | 0.9506 | 21.22966 | 20.87111 | 40 |
| CD8A | chr2:87017780-87017917 | 0.9907 | 20 | 21.32184 | 48 |
| CDH4_A | chr20:59827230-59827285 | 1 | 37.02782 | 42.86441 | 49 |
| CDH4_B | chr20:59827762-59827776 | 0.9907 | 22.76012 | 30.46422 | 50 |
| CDH4_C | chr20:59827794-59827868 | 1 | 24.72984 | 23.55503 | 51 |
| CDH4_D | chr20:59828193-59828258 | 0.9958 | 25.09787 | 28.16619 | 52 |
| CDH4_E | chr20:59828479-59828729 | 1 | 28.97206 | 36.79341 | 53 |
| CDH4_F | chr20:59828778-59828814 | 0.9969 | 36.81109 | 34.35411 | 54 |
| CHST2_B | chr3:142839223-142839568 | 1 | 34.72482 | 117.3308 | 57 |
| CLIP4 | chr2:29338109-29338339 | 0.9739 | 20.94282 | 46.96947 | 59 |
| CR1 | chr1:207669481-207669639 | 0.9691 | 28.25359 | 42.1256 | 61 |
| DLK1 | chr14:101193295-101193318 | 0.9692 | 31.7083 | 30.33924 | 65 |
| DNAJC6 | chr1:65731412-65731507 | 0.9691 | 35.40474 | 85.64082 | 68 |
| DNM3_A | chr1:171810393-171810575 | 0.9506 | 16.78657 | 101.8429 | 69 |
| EMX1_A | chr2:73151498-73151578 | 0.9923 | 14.73071 | 31.60031 | 74 |
| ESPN | chr1:6508784-6509175 | 1 | 7.096692 | 53.37799 | 77 |
| FABP5 | chr8:82192605-82192921 | 0.9475 | 18.49851 | 297.5222 | 81 |
| FAM150A | chr8:53478266-53478416 | 1 | 26.83744 | 30.32598 | 85 |
| FLJ42875 | chr1:2987037-2987116 | 0.9667 | 40.47655 | 36.2069 | 92 |
| GLP1R | chr6:39016381- | 0.9725 | 22.49606 | 24.93019 | 102 |

| | | | | | |
|---|---|---|---|---|---|
| | 39016421 | | | | |
| GNG4 | chr1:235813658-235813798 | 0.9771 | 30.92768 | 28.97404 | 103 |
| GYPC_A | chr2:127413505-127413678 | 0.9568 | 20.44592 | 91.04351 | 108 |
| HAND2 | chr4:174450452-174450478 | 0.9804 | 15.73026 | 23.81474 | 110 |
| HES5 | chr1:2461823-2461915 | 0.9383 | 31.91815 | 23.13591 | 112 |
| HNF1B_A | chr17:36103713-36103793 | 0.963 | 29.18949 | 39.45301 | 115 |
| HNF1B_B | chr17:36105390-36105448 | 1 | 3.593578 | 29.6686 | 116 |
| HOXA1_A | chr7:27135603-27135889 | 0.966 | 38.04738 | 137.2168 | 117 |
| HOXA1_B | chr7:27136191-27136244 | 0.9522 | 33.78035 | 144.6796 | 118 |
| HOXA7_A | chr7:27195742-27195895 | 0.9784 | 22.8203 | 34.64696 | 119 |
| HOXA7_B | chr7:27196032-27196190 | 1 | 27.92413 | 23.54393 | 120 |
| HOXA7_C | chr7:27196441-27196531 | 0.9896 | 20.14606 | 27.05282 | 121 |
| HOXD9 | chr2:176987716-176987739 | 0.9926 | 21.14973 | 31.76069 | 122 |
| IGF2BP3_A | chr7:23508901-23509225 | 0.9599 | 22.75591 | 108.9025 | 123 |
| IGF2BP3_B | chr7:23513817-23514114 | 0.9853 | 8.970018 | 75.12555 | 124 |
| IGSF9B_A | chr11:133825409-133825476 | 0.9691 | 13.84201 | 22.99205 | 126 |
| IL15RA | chr10:6018610-6018848 | 0.941 | 6.854012 | 58.47407 | 128 |
| INSM1 | chr20:20348140-20348182 | 0.9542 | 25.90248 | 26.65219 | 130 |
| ITPKA_B | chr15:41793928-41794003 | 0.9686 | 21.34743 | 23.96879 | 133 |
| ITPRIPL1 | chr2:96990968-96991328 | 0.963 | 31.10465 | 280.3382 | 134 |
| KCNE3 | chr11:74178260- | 0.9529 | 37.65937 | 30.48685 | 137 |

| | | | | | |
|---|---|---|---|---|---|
| | 74178346 | | | | |
| KCNK17_B | chr6:39281408-39281478 | 0.966 | 31.5971 | 104.6458 | 140 |
| LIME1 | chr20:62369116-62369393 | 1 | 3.213465 | 75.53068 | 148 |
| LOC100132891 | chr8:72755897-72756295 | 0.9691 | 33.07259 | 53.92857 | 153 |
| LOC283999 | chr17:76227905-76227960 | 0.9837 | 14.82154 | 37.5134 | 155 |
| LY6H | chr8:144241547-144241557 | 0.9722 | 14.69706 | 28.21535 | 158 |
| MAST1 | chr19:12978399-12978642 | 0.9654 | 26.7166 | 37.34729 | 160 |
| MAX.chr1.158083198-158083476 | chr1:158083198-158083476 | 0.9907 | 35.99869 | 32.08705 | 161 |
| MAX.chr1.228074764-228074977 | chr1:228074764-228074977 | 0.9846 | 33.58852 | 37.24138 | 162 |
| MAX.chr1.46913931-46913950 | chr1:46913931-46913950 | 0.9784 | 27.23106 | 24.5654 | 164 |
| MAX.chr10.130085265-130085312 | chr10:130085265-130085312 | 1 | 23.65531 | 23.42432 | 167 |
| MAX.chr11.68622869-68622968 | chr11:68622869-68622968 | 1 | 72.19153 | 99.26843 | 169 |
| MAX.chr14.101176106-101176260 | chr14:101176106-101176260 | 0.9771 | 19.13125 | 42.66797 | 172 |
| MAX.chr15.96889013-96889128 | chr15:96889069-96889128 | 0.9882 | 16.95179 | 32.0494 | 173 |
| MAX.chr17.8230197-8230314 | chr17:8230197-8230314 | 0.966 | 17.19388 | 40.39153 | 175 |
| MAX.chr19.46379903-46380197 | chr19:46379903-46380197 | 0.9902 | 32.1749 | 31.74585 | 179 |
| MAX.chr2.97193163-97193287 | chr2:97193163-97193287 | 0.9522 | 25.05757 | 666.7396 | 182 |
| MAX.chr2.97193478-97193562 | chr2:97193478-97193562 | 0.9549 | 29.12281 | 158.5146 | 183 |
| MAX.chr20.1784209-1784461 | chr20:1784209-1784461 | 0.9784 | 60.31305 | 39.01045 | 185 |
| MAX.chr21.44782441-44782498 | chr21:44782441-44782498 | 0.9688 | 16.58956 | 71.97633 | 186 |

| | | | | | |
|---|---|---|---|---|---|
| MAX.chr22.23908718-23908782 | chr22:23908718-23908782 | 1 | 25.82947 | 20.84453 | 188 |
| MAX.chr5.145725410-145725459 | chr5:145725410-145725459 | 0.9969 | 14.69927 | 29.27086 | 193 |
| MAX.chr5.178957564-178957598 | chr5:178957564-178957598 | 0.9614 | 16.46627 | 42.22336 | 195 |
| MAX.chr5.180101084-180101094 | chr5:180101084-180101094 | 1 | 23.37255 | 25.00699 | 196 |
| MAX.chr5.42952185-42952280 | chr5:42952185-42952280 | 0.966 | 16.77837 | 67.63893 | 197 |
| MAX.chr5.42994866-42994936 | chr5:42994866-42994936 | 0.9112 | 4.703287 | 161.8831 | 198 |
| MAX.chr6.27064703-27064783 | chr6:27064703-27064783 | 0.9537 | 20.54983 | 23.77734 | 205 |
| MAX.chr7.152622607-152622638 | chr7:152622607-152622638 | 0.9522 | 24.6674 | 20.98723 | 207 |
| MAX.chr8.145104132-145104218 | chr8:145104132-145104218 | 0.9641 | 23.94389 | 106.2614 | 211 |
| MAX.chr9.136474504-136474527 | chr9:136474504-136474527 | 0.951 | 20.88926 | 25.01507 | 215 |
| MCF2L2 | chr3:182896930-182897245 | 0.9753 | 20.09711 | 22.94148 | 216 |
| MSX2P1 | chr17:56234436-56234516 | 0.9105 | 20.25101 | 185.2593 | 222 |
| NACAD | chr7:45128502-45128717 | 0.9583 | 24.13599 | 24.56509 | 223 |
| NID2_B | chr14:52535974-52536161 | 0.966 | 21.89118 | 30.61013 | 225 |
| NID2_C | chr14:52536192-52536328 | 0.9846 | 21.19688 | 35.70811 | 226 |
| ODC1 | chr2:10589075-10589243 | 0.9896 | 5.239957 | 199.2568 | 231 |
| OSR2_B | chr8:99952801-99952919 | 0.9599 | 24.39913 | 21.91589 | 235 |
| PAQR6 | chr1:156215470-156215739 | 0.9965 | 1.875785 | 35.09138 | 238 |
| PCDH8 | chr13:53421299-53421322 | 0.9907 | 14.32 | 28.05643 | 239 |
| PIF1 | chr15:65116285- | 0.9537 | 43.87855 | 44.78209 | 244 |

| | | | | | |
|---|---|---|---|---|---|
| | 65116597 | | | | |
| PPARA | chr22:46545328-46545457 | 0.9896 | 1.934821 | 27.81555 | 248 |
| PPP2R5C | chr14:102247681-102247929 | 0.9969 | 40.41616 | 21.95545 | 252 |
| PRDM13_A | chr6:100061616-100061742 | 0.9537 | 24.24062 | 61.61066 | 253 |
| PRHOXNB | chr13:28552424-28552562 | 1 | 32.97143 | 25.41024 | 255 |
| PRKCB | chr16:23847575-23847699 | 0.9537 | 30.71429 | 443.1833 | 256 |
| RBFOX3_A | chr17:77179579-77179752 | 0.9846 | 21.15348 | 20.09964 | 262 |
| RBFOX3_B | chr17:77179778-77180064 | 0.9784 | 22.97297 | 38.87734 | 263 |
| RFX8 | chr2:102090934-102091130 | 0.9475 | 14.08461 | 61.73279 | 264 |
| SNCA | chr4:90758071-90758118 | 0.9622 | 14.42541 | 42.52051 | 283 |
| STAC2_A | chr17:37381217-37381303 | 0.9815 | 43.97999 | 23.61791 | 286 |
| STAC2_B | chr17:37381689-37381795 | 0.9938 | 59.47293 | 161.2592 | 287 |
| STX16_B | chr20:57225077-57225227 | 0.989 | 1.467485 | 182.6884 | 289 |
| SYT5 | chr19:55690401-55690496 | 0.9938 | 16.49149 | 33.17451 | 292 |
| TIMP2 | chr17:76921762-76921779 | 0.9568 | 17.75848 | 42.58231 | 297 |
| TMEFF2 | chr2:193060012-193060126 | 0.9753 | 17.97114 | 35.24222 | 299 |
| TNFRSF10D | chr8:23020896-23021114 | 0.9475 | 22.13556 | 107.3874 | 301 |
| TRH_B | chr3:129694457-129694501 | 1 | 18.95629 | 21.0275 | 304 |
| TRIM67 | chr1:231297047-231297159 | 1 | 23.47643 | 15.57769 | 305 |
| TRIM71_C | chr3:32860020-32860090 | 0.9826 | 28.31276 | 43.84559 | 308 |

| | | | | | |
|---|---|---|---|---|---|
| USP44_A | chr12:95942148-95942178 | 0.9722 | 25.33383 | 22.23173 | 312 |
| USP44_B | chr12:95942519-95942558 | 0.9688 | 29.71223 | 20.72773 | 313 |
| UTF1 | chr10:135044125-135044171 | 0.9935 | 24.15274 | 23.83046 | 314 |
| UTS2R | chr17:80329497-80329534 | 0.9896 | 37.98289 | 25.32411 | 315 |
| VSTM2B_A | chr19:30016283-30016357 | 0.9654 | 57.09044 | 59.81456 | 319 |
| VSTM2B_B | chr19:30017789-30018165 | 0.9673 | 32.07169 | 27.33698 | 320 |
| ZFP64 | chr20:50721057-50721235 | 0.9506 | 27.53052 | 22.5886 | 322 |
| ZNF132 | chr19:58951402-58951775 | 0.9804 | 39.76355 | 100.0992 | 323 |

**Table 5.** Table 5 shows 1) area under the curve for identified methylated regions distinguishing Luminal A breast cancer tissue from normal breast tissue, 2) the Fold Change (FC) for Luminal A breast cancer tissue vs. normal breast tissue, and 3) the Fold Change (FC) for Luminal A breast cancer tissue vs. buffy coat (normal).

| Gene Annotation | Region on Chromosome (starting base-ending base) | AUC | FC Tissue | FC Buffy | DMR |
|---|---|---|---|---|---|
| ARL5C | chr17:37321515-37321626 | 0.9083 | 10.00664 | 31.97539 | 13 |
| BHLHE23_C | chr20:61638088-61638565 | 0.9184 | 31.17451 | 43.04012 | 23 |
| BMP6 | chr6:7727566-7727907 | 0.9248 | 33.44248 | 32.18487 | 26 |
| C10orf125 | chr10:135171410-135171661 | 0.9816 | 5.951195 | 52.52747 | 27 |
| C17orf64 | chr17:58499095-58499190 | 0.9414 | 9.129866 | 128.5583 | 29 |
| C19orf66 | chr19:10197688-10197823 | 0.9288 | 3.629997 | 23.26103 | 31 |
| CAMKV | chr3:49907259- | 0.9265 | 31.61795 | 34.87738 | 38 |

| | | | | | |
|---|---|---|---|---|---|
| | 49907298 | | | | |
| CD1D | chr1:158150864-158151129 | 0.9575 | 26.85386 | 35.71281 | 47 |
| CDH4_E | chr20:59828479-59828729 | 0.9575 | 19.38124 | 24.61343 | 53 |
| CDH4_F | chr20:59828778-59828814 | 0.9167 | 27.70653 | 25.85724 | 54 |
| CHST2_A | chr3:142838025-142838494 | 0.9167 | 47.12016 | 106.0335 | 56 |
| CRHBP | chr5:76249939-76249997 | 0.9294 | 14.22073 | 22.1281 | 62 |
| DLX6 | chr7:96635255-96635475 | 0.9622 | 13.78623 | 28.53928 | 67 |
| DNM3_B | chr1:171810648-171810702 | 0.9087 | 29.14931 | 295.5986 | 70 |
| DNM3_C | chr1:171810806-171810920 | 0.9753 | 23.67912 | 99.77376 | 71 |
| DNM3_A | chr1:171810393-171810575 | 0.9134 | 21.31894 | 129.3404 | 69 |
| ESYT3 | chr3:138153979-138154071 | 0.9479 | 39.19083 | 37.19512 | 78 |
| ETS1_A | chr11:128391809-128391908 | 0.9089 | 40.45139 | 159.0444 | 79 |
| ETS1_B | chr11:128392062-128392309 | 0.8872 | 34.63309 | 188.3098 | 80 |
| FAM126A | chr7:23053941-23054066 | 0.9706 | 57.86891 | 65.82935 | 83 |
| FAM189A1 | chr15:29862130-29862169 | 0.9757 | 18.04237 | 28.53505 | 88 |
| FAM20A | chr17:66597237-66597326 | 0.9019 | 35.24514 | 24.36451 | 89 |
| FAM59B | chr2:26407713-26407972 | 0.9479 | 1.945513 | 103.9384 | 90 |
| FBN1 | chr15:48937412-48937541 | 0.9599 | 31.33933 | 27.92071 | 91 |
| FLRT2 | chr14:85998469-85998535 | 0.9428 | 15.80425 | 20.40157 | 93 |
| FMN2 | chr1:240255171-240255253 | 0.9294 | 27.79887 | 61.08723 | 94 |

| | | | | | |
|---|---|---|---|---|---|
| FOXP4 | chr6:41528816-41528958 | 1 | 1.388687 | #DIV/0! | 96 |
| GAS7 | chr17:10101325-10101397 | 0.9282 | 39.97585 | 23.28643 | 99 |
| GYPC_A | chr2:127413505-127413678 | 0.9379 | 16.91651 | 75.32742 | 108 |
| GYPC_B | chr2:127414096-127414189 | 0.9727 | 15.16704 | 832.1792 | 109 |
| HAND2 | chr4:174450452-174450478 | 0.9583 | 13.64474 | 20.65737 | 110 |
| HES5 | chr1:2461823-2461915 | 0.9111 | 21.96548 | 15.9217 | 112 |
| HMGA2 | chr12:66219385-66219487 | 0.9314 | 46.53533 | 21.43751 | 114 |
| HNF1B_B | chr17:36105390-36105448 | 0.9926 | 2.464626 | 20.34797 | 116 |
| IGF2BP3_B | chr7:23513817-2351411 | 0.969 | 15.15625 | 99.58003 | 124 |
| IGF2BP3_A | chr7:23508901-23509225 | 0.9167 | 18.96654 | 90.76778 | 123 |
| KCNH8 | chr3:19189837-19189897 | 0.9821 | 26.86423 | 11.12219 | 138 |
| KCNK17_A | chr6:39281195-39281282 | 0.9111 | 64.74638 | 44.94467 | 139 |
| KCNQ2 | chr20:62103558-62103625 | 0.9379 | 15.9322 | 58.35214 | 142 |
| KLHDC7B | chr22:50987219-50987304 | 1 | 1.458785 | 126.5684 | 146 |
| LOC100132891 | chr8:72755897-72756295 | 0.9477 | 21.07843 | 34.37075 | 153 |
| MAX.chr1.46913931-46913950 | chr1:46913931-46913950 | 0.9074 | 23.06829 | 20.81013 | 164 |
| MAX.chr11.68622869-68622968 | chr11:68622869-68622968 | 0.9395 | 46.67485 | 64.1812 | 169 |
| MAX.chr12.4273906-4274012 | chr12:4273906-4274012 | 0.9379 | 20.87418 | 155.4373 | 170 |
| MAX.chr12.59990591-59990895 | chr12:59990591-59990895 | 0.8807 | 14.01947 | 21.10553 | 171 |
| MAX.chr17.73073682- | chr17:73073682- | 0.9449 | 1.052067 | 45.64784 | 174 |

| | | | | | |
|---|---|---|---|---|---|
| 73073814 | 73073814 | | | | |
| MAX.chr20.1783841-1784054 | chr20:1783841-1784054 | 0.9074 | 27.09573 | 22.06724 | 184 |
| MAX.chr21.47063802-47063851 | chr21:47063802-47063851 | 0.9757 | 16.51515 | 79.7561 | 187 |
| MAX.chr4.8860002-8860038 | chr4:8860002-8860038 | 0.9363 | 16.17858 | 68.04479 | 192 |
| MAX.chr5.172234248-172234494 | chr5:172234248-172234494 | 0.9201 | 1.531023 | 83.07827 | 194 |
| MAX.chr5.178957564-178957598 | chr5:178957564-178957598 | 0.9392 | 11.40949 | 29.25659 | 195 |
| MAX.chr6.130686865-130686985 | chr6:130686865-130686985 | 0.9583 | 39.03866 | 37.31522 | 202 |
| MAX.chr8.687688-687736 | chr8:687688-687736 | 0.9286 | 24.48762 | 22.46817 | 212 |
| MAX.chr8.688863-688924 | chr8:688863-688924 | 0.9303 | 15.25862 | 30.30423 | 213 |
| MAX.chr9.114010-114207 | chr9:114010-114207 | 0.9085 | 25.1809 | 34.53142 | 214 |
| MPZ | chr1:161275561-161275996 | 0.933 | 36.3026 | 503.8832 | 221 |
| NID2_A | chr14:52535260-52535353 | 0.9316 | 29.32631 | 35.83691 | 224 |
| NKX2-6 | chr8:23564115-23564146 | 0.908 | 15.67986 | 32.03798 | 227 |
| ODC1 | chr2:10589075-10589243 | 1 | 5.298588 | 201.4864 | 231 |
| OSR2_A | chr8:99952233-99952366 | 0.951 | 17.65456 | 23.40924 | 234 |
| POU4F1 | chr13:79177505-79177532 | 0.9241 | 14.6281 | 25.83187 | 247 |
| PRDM13_B | chr6:100061748-100061792 | 0.9549 | 22.67697 | 52.7912 | 254 |
| PRKCB | chr16:23847575-23847699 | 0.9248 | 26.98915 | 389.4325 | 256 |
| RASGRF2 | chr5:80256117-80256162 | 0.9327 | 24.29321 | 45.671 | 261 |
| RIPPLY2 | chr6:84563228-84563287 | 0.9216 | 20.4497 | 24 | 267 |

| | | | | | |
|---|---|---|---|---|---|
| SLC30A10 | chr1:220101458-220101634 | 0.9346 | 21.20187 | 19.7307 | 279 |
| ST8SIA4 | chr5:100240059-100240276 | 1 | 1.754394 | 257.6766 | 285 |
| SYN2 | chr3:12045894-12045967 | 0.9232 | 22.95533 | 31.86263 | 290 |
| TRIM71_A | chr3:32858861-32858897 | 0.9184 | 15.38071 | 50.65283 | 306 |
| TRIM71_B | chr3:32859445-32859559 | 0.9375 | 15.41597 | 43.92036 | 307 |
| TRIM71_C | chr3:32860020-32860090 | 0.9115 | 25.64374 | 39.71231 | 308 |
| UBTF | chr17:42287924-42288018 | 1 | 2.648869 | 421.8795 | 310 |
| ULBP1 | chr6:150285563-150285661 | 0.902 | 16.53425 | 26.75089 | 311 |
| USP44_B | chr12:95942519-95942558 | 0.975 | 29.4964 | 20.57716 | 313 |
| VSTM2B_A | chr19:30016283-30016357 | 0.9283 | 34.40535 | 36.04704 | 319 |

**Table 6.** Table 6 shows 1) area under the curve for identified methylated regions distinguishing Luminal B breast cancer tissue from normal breast tissue, 2) the Fold Change (FC) for Luminal B breast cancer tissue vs. normal breast tissue, and 3) the Fold Change (FC) for Luminal B breast cancer tissue vs. buffy coat (normal).

| **Gene Annotation** | **Region on Chromosome (starting base-ending base)** | **AUC** | **FC Tissue** | **FC Buffy** | **DMR** |
|---|---|---|---|---|---|
| ACCN1 | chr17:31620207-31620314 | 0.9198 | 23.85808 | 9.167347 | 4 |
| AJAP1_A | chr1:4715535-4715646 | 0.9815 | 24.85037 | 23.66546 | 6 |
| AJAP1_B | chr1:4715931-4716021 | 0.9491 | 33.32713 | 54.60366 | 7 |
| BEST4 | chr1:45251853-45252029 | 0.9059 | 32.89966 | 62.95737 | 20 |

| | | | | | |
|---|---|---|---|---|---|
| CALN1_B | chr7:71801741-71801800 | 0.9059 | 16.99878 | 31.42622 | 37 |
| CBLN1_B | chr16:49316198-49316258 | 0.933 | 15.66904 | 39.58407 | 42 |
| CDH4_E | chr20:59828479-59828729 | 0.9259 | 17.19561 | 21.83777 | 53 |
| DLX4 | chr17:48042562-48042606 | 1 | 3.346919 | 60.11236 | 66 |
| FOXP4 | chr6:41528816-41528958 | 1 | 1.056007 | #DIV/0! | 96 |
| IGSF9B_B | chr11:133825491-133825530 | 0.9815 | 21.1913 | 21.56637 | 127 |
| ITPRIPL1 | chr2:96990968-96991328 | 0.9074 | 21.92125 | 197.5706 | 134 |
| KCNA1 | chr12:5019401-5019633 | 0.9414 | 53.02013 | 39.91732 | 136 |
| KLF16 | chr19:1857330-1857476 | 0.8791 | 12.18471 | 92.18633 | 145 |
| LMX1B_A | chr9:129388175-129388223 | 0.9965 | 2.639923 | 62.01749 | 149 |
| MAST1 | chr19:12978399-12978642 | 0.9706 | 16.13892 | 22.56069 | 160 |
| MAX.chr11.14926602-14927148 | chr11:14926602-14927148 | 1 | 3.646943 | 81.99557 | 168 |
| MAX.chr17.73073682-73073814 | chr17:73073682-73073814 | 0.9514 | 1.236217 | 53.63787 | 174 |
| MAX.chr18.76734362-76734476 | chr18:76734362-76734476 | 0.9414 | 15.62804 | 48.90311 | 177 |
| MAX.chr19.30719261-30719354 | chr19:30719261-30719354 | 0.9101 | 23.15574 | 21.34761 | 178 |
| MAX.chr22.42679578-42679917 | chr22:42679578-42679917 | 0.963 | 28.63358 | 21.76462 | 189 |
| MAX.chr4.8860002-8860038 | chr4:8860002-8860038 | 0.9259 | 17.90907 | 75.323 | 192 |
| MAX.chr5.145725410-145725459 | chr5:145725410-145725459 | 0.9012 | 10.81956 | 21.54514 | 193 |
| MAX.chr5.178957564-178957598 | chr5:178957564-178957598 | 0.9028 | 14.08818 | 36.12539 | 195 |
| MAX.chr5.77268672- | chr5:77268672- | 0.9228 | 16.4233 | 39.91228 | 199 |

| | | | | | |
|---|---|---|---|---|---|
| 77268725 | 77268725 | | | | |
| MAX.chr8.124173128-124173268 | chr8:124173128-124173268 | 0.9105 | 12.93676 | 45.59879 | 209 |
| MPZ | chr1:161275561-161275996 | 0.9653 | 19.98003 | 184.2234 | 221 |
| PPARA | chr22:46545328-46545457 | 0.9931 | 1.592475 | 22.89388 | 248 |
| PRMT1 | chr19:50179501-50179635 | 0.8837 | 11.53981 | 25.86275 | 257 |
| RBFOX3_B | chr17:77179778-77180064 | 0.9012 | 18.327 | 31.01493 | 263 |
| RYR2_A | chr1:237205369-237205428 | 0.9392 | 21.32044 | 25 | 268 |
| SALL3 | chr18:76739321-76739404 | 0.96 | 58.85028 | 60.07958 | 270 |
| SCRT2_A | chr20:644533-644618 | 0.9871 | 19.11925 | 8.272966 | 272 |
| SPHK2 | chr19:49127580-49127683 | 0.9753 | 38.67547 | 42.87091 | 284 |
| STX16_B | chr20:57225077-57225227 | 1 | 1.503476 | 187.169 | 289 |
| SYNJ2 | chr6:158402213-158402536 | 1 | 1.81213 | 79.15141 | 291 |
| TMEM176A | chr7:150497411-150497535 | 0.8719 | 18.02734 | 13.07736 | 300 |
| TSHZ3 | chr19:31839809-31840038 | 0.9475 | 19.63569 | 29.13422 | 309 |
| VIPR2 | chr7:158937370-158937481 | 0.9537 | 28.49829 | 22.56321 | 316 |

**Table 7.** Table 7 shows 1) area under the curve for identified methylated regions distinguishing BRCA1 breast cancer tissue from normal breast tissue, 2) the Fold Change (FC) for BRCA1 breast cancer tissue vs. normal breast tissue, and 3) the Fold Change (FC) for BRCA1 breast cancer tissue vs. buffy coat (normal).

| Gene Annotation | Region on Chromosome (starting base-ending base) | AUC | FC Tissue | FC Buffy | DMR No. |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| C10orf93 | chr10:134756078-134756167 | 1 | 35.64082278 | 21.3938 | 28 |
| C20orf195_A | chr20:62185293-62185364 | 0.9537 | 25.6624628 | 88.34146 | 33 |
| C20orf195_B | chr20:62185418-62185546 | 0.9537 | 31.08894431 | 47.34177 | 34 |
| CALN1_B | chr7:71801741-71801800 | 1 | 22.4757876 | 41.55176 | 37 |
| CBLN1_A | chr16:49315588-49315691 | 1 | 23.38948327 | 22.60229 | 41 |
| CBLN1_B | chr16:49316198-49316258 | 0.9815 | 27.353707 | 71.4388 | 42 |
| CCDC61 | chr19:46519467-46519536 | 0.9667 | 48.80498092 | 67.25713 | 43 |
| CCND2 | chr12:4378317-4378375 | 0.9333 | 16.28123545 | 100.7685 | 44 |
| CCND2 | chr12:4380560-4380681 | 0.951 | 10.56487202 | 70.74468 | 45 |
| CCND2 | chr12:4384096-4384146 | 0.9907 | 25.60667341 | 76.22272 | 46 |
| EMX1_B | chr2:73151663-73151756 | 0.9833 | 13.75989446 | 30.57754 | 75 |
| FAM150B | chr2:287868-287919 | 0.9306 | 32.67264761 | 21.96353 | 86 |
| GRASP | chr12:52400919-52401166 | 0.9259 | 28.3875581 | 48.04841 | 105 |
| HBM | chr16:216426-216451 | 0.9706 | 35.04374159 | 35.09097 | 111 |
| ITPRIPL1 | chr2:96990968-96991328 | 1 | 17.39816032 | 163.0944 | 134 |
| KCNK17_A | chr6:39281195-39281282 | 0.9583 | 36.55797101 | 25.37726 | 139 |
| KIAA1949 | chr6:30646976-30647084 | 0.9556 | 30.04064322 | 173.3102 | 143 |
| LOC100131176 | chr7:151106986-151107060 | 1 | 16.94187139 | 29.40354 | 152 |
| MAST1 | chr19:12978399-12978642 | 1 | 19.30541369 | 26.98715 | 160 |
| MAX.chr1.8277285-8277316 | chr1:8277285-8277316 | 0.9815 | 31.30790191 | 52.33035 | 165 |

| | | | | | |
|---|---|---|---|---|---|
| MAX.chr1.8277479-8277527 | chr1:8277479-8277527 | 1 | 18.61607143 | 45.48146 | 166 |
| MAX.chr11.14926602-14927148 | chr11:14926602-14927148 | 1 | 4.590639238 | 107.8495 | 168 |
| MAX.chr15.96889013-96889128 | chr15:96889013-96889128 | 0.9778 | 18.09917355 | 35.80772 | 173 |
| MAX.chr18.5629721-5629791 | chr18:5629721-5629791 | 0.9375 | 17.83216783 | 20.53691 | 177 |
| MAX.chr19.30719261-30719354 | chr19:30719261-30719354 | 1 | 33.50409836 | 30.88791 | 178 |
| MAX.chr22.42679578-42679917 | chr22:42679578-42679917 | 0.9778 | 37.74834437 | 39.34049 | 189 |
| MAX.chr5.178957564-178957598 | chr5:178957564-178957598 | 1 | 22.21108884 | 56.95444 | 195 |
| MAX.chr5.77268672-77268725 | chr5:77268672-77268725 | 0.9074 | 15.49630845 | 37.65949 | 199 |
| MAX.chr6.157556793-157556856 | chr6:157556793-157556856 | 0.9778 | 33.75787815 | 33.87352 | 203 |
| MAX.chr8.124173030-124173395 | chr8:124173030-124173395 | 1 | 23.47893058 | 63.5876 | 208 |
| MN1 | chr22:28197962-28198388 | 0.9352 | 23.36568394 | 26.69456 | 220 |
| MPZ | chr1:161275561-161275996 | 0.8519 | 49.15590864 | 856.6978 | 221 |
| NR2F6 | chr19:17346428-17346459 | 1 | 13.02466029 | 353.7936 | 228 |
| PDXK_A | chr21:45148429-45148556 | 0.9722 | 75.55254849 | 229.9245 | 241 |
| PDXK_B | chr21:45148575-45148681 | 0.9778 | 24.6031746 | 68.40522 | 242 |
| PTPRM | chr18:7568565-7568808 | 1 | 27.52463054 | 20.36446 | 259 |
| RYR2_B | chr1:237205619-237205640 | 0.95 | 21.12877583 | 25.85603 | 269 |
| SERPINB9_A | chr6:2902941-2902998 | 0.9907 | 28.33433917 | 27.88833 | 274 |
| SERPINB9_B | chr6:2903031-2903143 | 0.9769 | 25.91687042 | 40.06479 | 275 |
| SLC8A3 | chr14:70654596- | 0.9706 | 16.90022757 | 46.84595 | 281 |

| | | | | | |
|---|---|---|---|---|---|
| | 70654640 | | | | |
| STX16_B | chr20:57225077-57225227 | 1 | 1.678527607 | 208.9613 | 289 |
| TEPP | chr16:58018790-58018831 | 0.9222 | 14.38988095 | 44.1351 | 296 |
| TOX | chr8:60030723-60030754 | 0.9537 | 13.484375 | 86.64659 | 302 |
| VIPR2 | chr7:158937370-158937481 | 0.9074 | 30.22915651 | 23.9336 | 316 |
| VSTM2B_A | chr19:30016283-30016357 | 0.9853 | 42.3267861 | 44.34645 | 319 |
| ZNF486 | chr19:20278004-20278145 | 1 | 28.7755102 | 40.02498 | 324 |
| ZNF626 | chr19:20844070-20844199 | 1 | 72.64705882 | 47.34274 | 325 |
| ZNF671 | chr19:58238810-58238955 | 1 | 30.69748581 | 235.6787 | 326 |

**Table 8.** Table 8 shows 1) area under the curve for identified methylated regions distinguishing BRCA2 breast cancer tissue from normal breast tissue, 2) the Fold Change (FC) for BRCA2 breast cancer tissue vs. normal breast tissue, and 3) the Fold Change (FC) for BRCA2 breast cancer tissue vs. buffy coat (normal).

| **Gene Annotation** | **Region on Chromosome (starting base-ending base)** | **AUC** | **FC Tissue** | **FC Buffy** | **DMR No.** |
|---|---|---|---|---|---|
| ANTXR2 | chr4:80993475-80993634 | 0.9074 | 28.18115 | 48.41438 | 12 |
| B3GNT5 | chr3:182971589-182971825 | 0.9136 | 118.1266 | 122.9242 | 16 |
| BHLHE23_A | chr20:61637950-61637986 | 0.9948 | 21.22272 | 21.07755 | 21 |
| BMP4 | chr14:54421578-54421916 | 0.9815 | 45.77028 | 21.74194 | 25 |
| CHRNA7 | chr15:32322830-32322897 | 1 | 349.4748 | 29.49189 | 55 |
| EPHA4 | chr2:222436217-222436320 | 0.9236 | 58.94207 | 21.52714 | 76 |

| | | | | | |
|---|---|---|---|---|---|
| FAM171A1 | chr10:15412558-15412652 | 0.9087 | 51.57005 | 26.28231 | 87 |
| FAM20A | chr17:66597237-66597326 | 0.996 | 30.89732 | 21.35891 | 89 |
| FMNL2 | chr2:153192734-153192836 | 0.9028 | 53.3376 | 24.66469 | 95 |
| FSCN1 | chr7:5633506-5633615 | 0.9028 | 25.15063 | 30.87569 | 97 |
| GSTP1 | chr11:67350986-67351055 | 0.9 | $>1x10^6$ | $>1x10^6$ | 107 |
| HBM | chr16:216426-216451 | 0.9 | 22.7961 | 22.82682 | 111 |
| IGFBP5 | chr2:217559103-217559244 | 0.9722 | 59.11994 | 22.99517 | 125 |
| IL17REL | chr22:50453462-50453555 | 1 | 28.30452 | 28.99172 | 129 |
| ITGA9 | chr3:37493895-37493994 | 0.9706 | 43.32188 | 35.86874 | 131 |
| ITPRIPL1 | chr2:96990968-96991328 | 0.9583 | 14.7331 | 147.0693 | 134 |
| KIRREL2 | chr19:36347825-36347863 | 0.9853 | 45.39026 | 35.4729 | 144 |
| LRRC34 | chr3:169530006-169530139 | 0.9306 | 22.69192 | 30.27401 | 156 |
| MAX.chr1.239549742-239549886 | chr1:239549742-239549886 | 0.916 | 20.67734 | 21.47568 | 163 |
| MAX.chr1.8277479-8277527 | chr1:8277479-8277527 | 0.8333 | 13.25255 | 32.37769 | 166 |
| MAX.chr11.14926602-14927148 | chr11:14926602-14927148 | 0.9922 | 3.97073 | 93.28576 | 168 |
| MAX.chr15.96889013-96889128 | chr15:96889013-96889128 | 0.9514 | 8.454545 | 16.72662 | 173 |
| MAX.chr2.238864674-238864735 | chr2:238864674-238864735 | 0.9762 | 27.89736 | 28.99259 | 181 |
| MAX.chr5.81148300-81148332 | chr5:81148300-81148332 | 0.9583 | 12.50391 | 24.59992 | 200 |
| MAX.chr7.151145632-151145743 | chr7:151145632-151145743 | 0.9444 | 8.972603 | 58.5148 | 206 |
| MAX.chr8.124173030-124173395 | chr8:124173030-124173395 | 0.9306 | 7.530176 | 56.77946 | 208 |

| | | | | | |
|---|---|---|---|---|---|
| MAX.chr8.143533298-143533558 | chr8:143533298-143533558 | 0.9097 | 32.741 | 20.5064 | 210 |
| MERTK | chr2:112656676-112656744 | 0.9222 | 27.51721 | 62.09376 | 217 |
| MPZ | chr1:161275561-161275996 | 0.9236 | 33.56504 | 584.9775 | 221 |
| NID2_C | chr14:52536192-52536328 | 0.9236 | 13.44693 | 22.6526 | 226 |
| NTRK3 | chr15:88800287-88800414 | 0.9167 | 20.89983 | 28.86352 | 229 |
| OLIG3_A | chr6:137818896-137818917 | 0.9198 | 15.98856 | 20.54162 | 232 |
| OLIG3_B | chr6:137818978-137818988 | 0.9012 | 11.04806 | 20.26448 | 233 |
| OSR2_C | chr8:99960580-99960630 | 0.9136 | 19.58805 | 32.89474 | 236 |
| PROM1 | chr4:16084793-16085112 | 0.9583 | 32.17623 | 41.64147 | 258 |
| RGS17 | chr6:153452120-153452393 | 0.9028 | 24.55645 | 20.63008 | 265 |
| SBNO2 | chr19:1131795-1131992 | 0.9012 | 58.01495 | 69.1242 | 271 |
| STX16_B | chr20:57225077-57225227 | 1 | 1.597137 | 198.8289 | 289 |
| TBKBP1 | chr17:45772630-45772726 | 0.9559 | 21.05769 | 41.61125 | 294 |
| TLX1NB | chr10:102881178-102881198 | 0.9074 | 22.34146 | 128.8689 | 298 |
| VIPR2 | chr7:158937370-158937481 | 0.9074 | 26.0117 | 20.59448 | 316 |
| VN1R2 | chr19:53758121-53758147 | 0.9583 | 17.79366 | 22.4549 | 317 |
| VSNL1 | chr2:17720216-17720257 | 0.9485 | 59.26645 | 43.69099 | 318 |
| ZFP64 | chr20:50721057-50721235 | 0.9167 | 25.93427 | 21.27889 | 322 |

**Table 9.** Table 9 shows 1) area under the curve for identified methylated regions distinguishing invasive breast cancer tissue from normal breast tissue, 2) the Fold Change

(FC) for invasive breast cancer tissue vs. normal breast tissue, and 3) the Fold Change (FC) for invasive breast cancer tissue vs. buffy coat (normal).

| Gene | Region on Chromosome (starting base-ending base) | AUC | FC Tissue | FC Buffy | DMR No. |
|---|---|---|---|---|---|
| CDH4_E | chr20:59828479-59828729 | 0.9319 | 24.19 | 24.91762 | 53 |
| FLJ42875 | chr1:2987037-2987116 | 0.9012 | 36.28 | 26.33723 | 92 |
| GAD2 | chr10:26505066-26505385 | 0.9016 | 25.3 | 33.18529 | 98 |
| GRASP | chr12:52400919-52401166 | 0.9311 | 40.47 | 56.12708 | 105 |
| ITPRIPL1 | chr2:96990968-96991328 | 0.91 | 32.57 | 236.8703 | 134 |
| KCNA1 | chr12:5019401-5019633 | 0.9147 | 55.3 | 35.34681 | 136 |
| MAX.chr12.4273906-4274012 | chr12:4273906-4274012 | 0.939 | 25.47 | 153.0038 | 170 |
| MAX.chr18.76734362-76734476 | chr18:76734362-76734476 | 0.9304 | 21.29 | 55.1493 | 177 |
| MAX.chr19.30719261-30719354 | chr19:30719261-30719354 | 0.9174 | 28.37 | 22.54408 | 178 |
| MAX.chr4.8859602-8859669 | chr4:8859602-8859669 | 0.9211 | 11.78 | 24.06671 | 191 |
| MAX.chr4.8860002-8860038 | chr4:8860002-8860038 | 0.9401 | 24.35 | 83.41947 | 192 |
| MAX.chr5.145725410-145725459 | chr5:145725410-145725459 | 0.9266 | 14.46 | 23.92735 | 193 |
| MAX.chr5.178957564-178957598 | chr5:178957564-178957598 | 0.9022 | 17.01 | 35.18328 | 195 |
| MAX.chr5.77268672-77268725 | chr5:77268672-77268725 | 0.9044 | 16.79 | 34.23046 | 199 |
| MPZ | chr1:161275561-161275996 | 0.9007 | 56.97 | 527.027 | 221 |
| NKX2-6 | chr8:23564115-23564146 | 0.9056 | 19.22 | 32.40724 | 227 |

| | | | | | |
|---|---|---|---|---|---|
| PRKCB | chr16:23847575-23847699 | 0.9032 | 35.63 | 371.6895 | 256 |
| RBFOX3_B | chr17:77179778-77180064 | 0.9241 | 22.81 | 32.30013 | 263 |
| SALL3 | chr18:76739321-76739404 | 0.9136 | 66.21 | 56.29973 | 270 |
| VSTM2B_A | chr19:30016283-30016357 | 0.9278 | 43.07 | 37.76572 | 319 |

Next, SYBR Green Methylation-specific PCR (qMSP) was performed on the discovery samples to confirm the accuracy and reproducibility of the candidate DMR's shown in Table 2. In addition, a 16 marker subset was run on frozen low grade and high grade DCIS samples to test applicability (22 high grade/CIS/P3 DCIS (ductal carcinoma in situ); 11 low grade/P1 DCIS).

qMSP primers were designed for each of the marker regions using Methprimer software (Li LC and Dahiya R. Bioinformatics. 2002 Nov;18(11):1427-31) They were synthesized by IDT (Integrated DNA Technologies). Assays were tested and optimized (using the Roche LightCycler 480) on dilutions of bisulfite converted universally methylated DNA, along with converted unmethylated DNA and converted and unconverted leukocyte DNA negative controls (10ng/ea). Assays taken forward needed to demonstrate linear regression curves and negative control values less than 5-fold below the lowest standard (1.6 genomic copies). Some of the more promising DMRs which had assay or control failures were re-designed. Of the 127 total designs (Table 10 shows the forward and reverse primer sequence information for the 127 total designs), 80 high performing MSP assays met QC criteria and were applied to the samples. The MSP primer sequences, each of which include 2-8 CpGs, were designed to provide a quick means of assessing methylation in the samples, and as such, were biased for amplification efficiency over trying to target the most discriminate CpGs - which would have required lengthy optimization timeframes.

DNA was purified as described in the discovery RRBS section and quantified using picogreen absorbance (Tecan/Invitrogen). 2ug of sample DNA was then treated with sodium bisulfite and purified using the Zymo EZ-96 Methylation kit (Zymo Research). Eluted material was amplified on Roche 480 LightCyclers using 384-well blocks. Each plate was able to accommodate 2 markers (and standards and controls) for a total of 40 plates. The 80 MSP assays had differing optimal amplification profiles (Tm = 60, 65, or 70°C) and were

grouped accordingly. The 20uL reactions were run using LightCycler 480 SYBR I Master mix (Roche) and 0.5umoles of primer for 50 cycles and analyzed, generally, by the Fit Point 18% absolute quantification method. All parameters (noise band, threshold, etc.) were pre-specified in an automated macro to avoid user subjectivity. The raw data, expressed in genomic copy number, was normalized to the amount of input DNA (β-actin). Results were analyzed logistically using JMP and displayed as AUC values. Twelve comparisons were run: each breast cancer subtype vs normal breast, and each subtype vs buffy coat. In addition, the methylation fold change ratio (mFCR) was calculated for each comparison using both average and median fractional methylation (FCR = cancer(methylated copies/β-actin copies)/normal(methylated copies/β-actin copies)). Both of these performance metrics were critical for assessing the potential of a marker in a clinical blood-based test.

>90% of the markers tested yielded superior performance in both AUC and FCR categories, with numerous AUCs in excess of 0.90, cancer vs normal tissue FCRs >10, and cancer vs buffy coat FCRs >50.

Table 11 shows area under the curve for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal breast tissue.

Table 12 shows area under the curve for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal buffy coat.

Table 13 shows methylation fold change for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal breast tissue.

Table 14 shows methylation fold change for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal buffy coat.

In the DCIS high grade vs low grade comparison, AUCs of the 16 markers tested ranged from 0.57 to 0.92. Several combinations of two markers achieved 95% sensitivity at

91% specificity (only 1 false positive) (Table 15). A 3 marker combination (SCRT2_B, ITPRIPL1, MAX.chr8.124173030-124173395) was 100% sensitive at 91% specificity.

**Table 10.**

| Gene Annotation | DMR No. | Forward Primer 5'-3' | SEQ ID NO: | Reverse Primer 5'-3' | SEQ ID NO: |
|---|---|---|---|---|---|
| AADAT-RS | 2 | GAG TTT CGG CGG CGT TTT TCG | 1 | CGC TAC GTC TAA CTT CCC GCG C | 2 |
| ABLIM1-FS | 3 | TTT TCG ACG AGT AGG ATT GAA GAA GGA AC | 3 | GCG AAT CTA TCT ACC GAA ACG CGC T | 4 |
| AJAP1_A | 6 | TTT TGA TTT GTA ATA TAG AGG AAA GCG TCG T | 5 | GTA TAA ACG CGT AAA TAC CAA ACT AAA CGA A | 6 |
| AJAP1_B | 7 | GTT TCG AGA AAG GAG AAG GGG GAG C | 7 | ACT CCC AAC GAA AAC TTC GCA AAC G | 8 |
| ALOX5-RS | 9 | GTT TTT TGT CGG GAG TTA TTC GT | 9 | CCA AAA ATT AAA TTA AAA ACG CTA CGC A | 10 |
| ASCL2-RS | 14 | GTT TTA GGA GGG TGG GGC GT | 11 | AAC ACG ACT ATT CGA AAA ACG CGC A | 12 |
| ATP6V1B1-RS | 15 | TTC GTA GTA TCG GGA GTC GA | 13 | GAA ATA ATA AAA ACG CCG CAC GCT | 14 |
| BANK1-FS | 17 | GTC GTA GTT TTC GCG GGT GGT AAG C | 15 | CGA ACG CTA CCT AAA CTC TCC CGA C | 16 |
| BEST4-RS | 20 | GGA ATC GCG AGT TTT GGG ATA GTC G | 17 | AAA TAC AAT TAC ACC CTC TAC CGC C | 18 |
| BHLHE23_C | 23 | GAG GCG TTC GGT GGG ATT TC | 19 | CCC CGA CCT ATA AAC CTA CGA CGC T | 20 |
| BHLHE23_D | 24 | GAG GAG GTA GCG GGC GTC GA | 21 | CGC GTC GAT CTA ACT TAC CTA CGA A | 22 |
| C10orf125-FS | 27 | TTG CGT TTA TCG ATT TCG TTT TCG T | 23 | GCA CTA CTA TCC CCC GAA CTA CTC TAC GC | 24 |
| C17orf64-RS | 29 | TTA TTA GGC GGG GAG TCG GGT GTC | 25 | CTC GAA TCC CTA AAA AAC TCG CGA A | 26 |
| C19orf66-FS | 31 | AGG AAA TTC GGT AGC GAT TAT ACG G | 27 | AAA CCC CTA CAA CCT CAC CGT ACA CGA T | 28 |
| CALN1_A | 36 | CGG AGT TAA TAG GTA CGG GAG GCG T | 29 | CAA ACC CCC GAA CTA TCG CGA A | 30 |
| CAPN2-FS | 39 | CGG GTA TCG CGG TTA AGT TGG C | 31 | TAT CGT AAA AAC CCA ACC CCT CGA C | 32 |
| CD1D-FS | 47 | GGG ATT GGT GAG ATT CGG GAC GT | 33 | CTC CCC GAA ACC AAA AAA CAA CGA | 34 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | A | |
| CDH4_E | 53 | GTT TTA AAT CGT ATT CGT AGT TCG G | 35 | ACG AAC GAA AAC TTT CCT AAA CGA A | 36 |
| CHST2_A | 56 | GCG TTT TTT TAT CGT TTT AGG GCG T | 37 | ACC GAC ACT ACC AAC CTC TCC GAA | 38 |
| CHST2_B | 57 | TGC GGG GAT TTT TAG CGG AAG C | 39 | CCG ACG AAC TAT CCG ACT ATC ACT CGT T | 40 |
| CLIC6-FS | 58 | GTA GTA GGT GGA GGG GGC GAG TTC | 41 | CTC TCG AAA ACC GCA AAA TCC TCG | 42 |
| CLIP4-FS | 59 | GGT AAT ATT GCG ATA TTT CGT AGA CGT | 43 | AAC AAT CAA ATA ATC GAA CGC ACG C | 44 |
| COL23A1-RS | 60 | GTC GTT TTT CGT TAC GAA GCG GC | 45 | AAA ACT AAA TAA ATC TAT CCT CGA T | 46 |
| CXCL12-FS | 63 | GCG TCG GCG GTT TTT AGT AAA AGC | 47 | AAC GAA TCT CAT TAA ATC TCC CGT C | 48 |
| DBNDD1R-FS | 64 | GAT TTT CGG GAG CGG CGA | 49 | CTT CCC CGC AAC GAA CCG | 50 |
| DLX4-FS | 66 | TTC GTT GGT ATA TTC GCG TAG GTG C | 51 | CGA ATA CCG AAA TCT ATA ACC CCG AA | 52 |
| DLX6-FS | 67 | ATT ATG ATT ACG ATG GTT GAC GG | 53 | CTC CAT AAA AAC GAA TTT AAA CGA A | 54 |
| DNM3_A | 69 | TTT GGT TAT AGA ACG TAG AGG TCG T | 55 | ATC GAA CCA CCA AAC CAA ACG C | 56 |
| DSCR6-FS | 72 | GGG AAG TTT AGT AGG TGA GCG T | 57 | ACT AAA AAC GTT TCC GTC GAA CGC A | 58 |
| DTX1-RS | 73 | GTT GGT AGG AGT AGG GTT GGT TCG A | 59 | ATC GCA ATC GTA ACC CGT AAA CGC | 60 |
| EMX1_A | 74 | ATT CGT ACG GTT TTT TCG TTT TCG T | 61 | GAC CAA CTA CTT CCG CTC GAC GC | 62 |
| ETS1_B | 80 | CGG ATT TAG CGG TCG AGA CG | 63 | TTT AAA ACG TTT CTC GCG ACG CC | 64 |
| FAM126A-FS | 83 | TCG TTA GGC GAT GAT AAT TAG CGA | 65 | TAA AAA AAC CAT AAA CCC TAA CGA C | 66 |
| FAM129C-FS | 84 | GTT GGA GAA GAC GAT TCG TTC GGA C | 67 | CCA AAA CCT CAC TCC TCA ACC GC | 68 |
| FBN1-FS | 91 | CGC GAT GCG CGT TTT GAA C | 69 | GAC GCG ACT AAC TTC CAA CCT AAC GAA | 70 |
| FMN2-RS | 94 | TTT TCG TGG TTG TCG TCG TTG C | 71 | GCC GCG CTC TAC ACT AAA CAT ATT CGC | 72 |
| FOXP4-FS | 96 | CGG GGA AGT GGG AGT TTT TAG CG | 73 | AAA AAA ACT AAA TCA AAA CCG CGA C | 74 |
| GAS7-FS | 99 | GCG AGT TCG CGT TGT TTA CGT TTC | 75 | ACC GAC GCT ACC TAT AAC TCC ACG CT | 76 |

| | | | | | |
|---|---|---|---|---|---|
| GP5-RS | 104 | TTA GGT TTG TTT ATT AAT TTT ACG T | 77 | TCT ACA AAA CGC CGC GAC | 78 |
| GRM7-FS | 106 | GTT AAT TCG AGA GCG CGA GGC GT | 79 | GAC CAA AAA AAA TAA AAA ATC CCG CGA C | 80 |
| GYPC_B | 109 | TAA AGA AAT AGA AAG CGG GCG ATA CGT | 81 | CGA ACT AAA AAA ACC GCC AAC CCG | 82 |
| HHEX-RS | 113 | GGG TTT TGC GGT TAA TGG CG | 83 | AAT AAC AAA CGC GTC CCG AAA ACG A | 84 |
| HNF1B_B | 116 | TTA GTT TTT TTT GGT TTT TAT TTG AAT TTC GA | 85 | AAC TTT TCC ACC GAT TCT CAA TTC CG | 86 |
| HOXA1_A | 117 | ATT TAA ATT TTC GGC GTT TCG TCG T | 87 | ACA CTC CAA ATC GAC CTT TAC AAT CGC | 88 |
| HOXA7_A | 119 | AGT TTG GTT CGT TTA GCG ATT GCG T | 89 | AAC GCG ACT AAA ACC AAT TTC CGC A | 90 |
| IGF2BP3_A | 123 | TTT ATT TGT TTT TAT CGT TCG TCG G | 91 | AAA TAT ATA CCC GAT TTC CCC GTT | 92 |
| IGF2BP3_B | 124 | TAA TCG GCG TCG AGA GAG ATA TCG T | 93 | CCG TCA ACC AAT CGA AAA CGA A | 94 |
| IL15RA-FS | 128 | TCG TTT ATT TCG TTT TTT TTG TCG A | 95 | AAC CAA CCT AAA ATC TAC ACT CGC A | 96 |
| ITPRIPL1-FS | 134 | GGG TCG TAG GGG TTT ATC GC | 97 | CAT ACT TAT CCG AAC GTC TAA ACG TC | 98 |
| ITPRIPL1-FS | 134 | GGT TTT AGC GAT GAA TCG GAC GT | 99 | CAC GAT CTT AAA AAA ACA ACG CGA C | 100 |
| KCNH8-RS | 138 | CGT ATT TTT AGG TTT AGT TCG GCG T | 101 | ACA CTA TTA CCC GCG AAA AAA CGA T | 102 |
| KCNK17_B | 140 | GAG TTT GTT TGG GGG TTG GTC GTA TTC | 103 | CCA AAT ATA ACG TTT AAC TCT TTA CCA CGA A | 104 |
| KCNK9-FS | 141 | TTT TTT TTG ATT CGG ATT TTT TCG G | 105 | CTA ATA AAC GCC GCC GTA TTC GAC G | 106 |
| KLF16-FS | 145 | TTT TCG CGT TGT TTT TAT TTA TCG T | 107 | TAC ACA ACC ACC CAA CTA CTC CGC G | 108 |
| KLHDC7B-RS | 146 | TGT TGT TGG GTA AAG GTT AGT ACG T | 109 | CGA AAA CCC AAC TCC CGA A | 110 |
| LAYN-RS | 147 | TTT TTG CGG TCG TTT TTC GGA GC | 111 | CTT ACC AAC TAA CCC CCG CCT ACC G | 112 |
| LIME1-RS | 148 | CGT TTT AGT AGG GAT TGG GGG CGA | 113 | CCC GAA AAC CAA AAT AAA ATC CGC A | 114 |
| LMX1B_A | 149 | CGG AAT AGC GCG GTC GTT TTT TC | 115 | TTT AAC CGT AAC GCT CGC CTC GAC | 116 |
| LOC100132891-FS | 153 | GTC GGT TGT GTT TAG AGC GTA GCG T | 117 | AAA AAA AAC CCC GAC GAC GAA | 118 |

| | | | | | |
|---|---|---|---|---|---|
| LOC100132891-FS | 153 | GTT GCG ATT GTT TGT ATT TTG CGG | 119 | ATA ATA ACA AAA AAC CCC TCC CGA C | 120 |
| LSS-FS | 157 | AGT TTC GTT AGG GAA GGG TTG CGT C | 121 | CAA CTA AAA CTC TAC CGC GCT CGA T | 122 |
| MAGI2-RS | 159 | AGG AAG GGT TTC GAG TTT AGT GCG G | 123 | AAA AAA ATC AAC GCG TCC TCC TCG C | 124 |
| MAST1-RS | 160 | TTT CGA TTT CGT TTT TAA ATT TCG T | 125 | AAA CTA AAC GAC CTA ACC CTA CGT A | 126 |
| MAX.chr1.8277479-8277527-RS | 166 | AAG TTT ACG CGC GAG TTT GAT CGT C | 127 | CGA AAC GAC TTC TCT CCC CGC A | 128 |
| MAX.chr11.14926602-14927148-FS | 168 | TTT AGT TCG CGG AAG TTA GGT TCG G | 129 | GAA AAC ACA ATA AAC CCC GCC GTC | 130 |
| MAX.chr11.68622869-68622968-FS | 169 | GTT AGA TTG TAG GAG GGA TTA GCG G | 131 | AAA AAA CGA CTA AAA AAT TCA CGC C | 132 |
| MAX.chr12.4273906-4274012-FS | 170 | TTT GGA GTT TGG GGG ATC GAT AGT C | 133 | CGA CGA AAC TAA AAC CGC GTA CGT A | 134 |
| MAX.chr12.4273906-4274012-FS | 170 | TTT GGA GTT TGG GGG ATC GAT AGT C | 135 | CGA CGA AAC TAA AAC CGC GTA CGT A | 136 |
| MAX.chr12.59990671-59990859-FS | 171 | ATT ATA TTG GGG GCG TTA GGT TCG G | 137 | AAC AAA CAA TTC GCA CGT AAA CGA A | 138 |
| MAX.chr15.96889013-96889128-FS | 173 | GGG CGG TTT ACG TGG ATT TTT ATA GAT TTT C | 139 | GCG TCT CGA ACC GTA CCC TAA CGT A | 140 |
| MAX.chr17.73073682-73073814-RS | 174 | CGT CGT TGT TGA TTA TGA TCG CGG | 141 | CGC TTC CTA ACA ACC TTC CTC GAA | 142 |
| MAX.chr18.76734362-76734476-RS | 177 | TTA ACG GTA TTT TTT GTT TTT TCG T | 143 | AAA AAA AAC TCG TCC CCG CGC T | 144 |
| MAX.chr19.46379903-46380197-FS | 179 | TCG GTT AGT TCG AGG TAG GAA GTT TTG C | 145 | TAT TAA CCG AAA AAC GAA AAC CAA ATC CGA | 146 |
| MAX.chr19.46379903-46380197-FS | 179 | AGT TTT GTT GTT TTG GGT AGG TCG G | 147 | AAA AAC TAA AAA CCT TTC TCT CGA C | 148 |
| MAX.chr2.223183057-223183114-RS | 180 | GCG TTG AGA GTG ACG GAT ATT TTT CGT C | 149 | ACT ACC TAA ACT CCG AAC ACG CCC G | 150 |
| MAX.chr20.1784209-1784461-FS | 185 | TTA GCG TAT CGG GAA TTA GGG GGA C | 151 | GAA AAC GAA AAA ACG ACG CGC A | 152 |
| MAX.chr20.1784209-1784461-RS | 185 | TCG TTT TTT AGG TGG GGA AGA AGC G | 153 | GAA CCG TAT TTA AAA CCA ATC CCC GC | 154 |
| MAX.chr4.8859602-8859669-RS | 191 | AAT TGG GGT TCG GGG TTC GGT AC | 155 | TTA CCC CTA CCC AAA AAA ATA CGC T | 156 |
| MAX.chr5.145725410-145725459-RS | 193 | GGG GTT AGA GTT TCG CGT TCG C | 157 | CGC GTC TCC CGT CCT ATC TAT ATA CGT C | 158 |
| MAX.chr5.42994866-42994936-FS | 198 | TAG GAA TTT TTT AAA TTC GTT TTA | 159 | CAC AAA AAC TCG ATA CAA TTA CCG | 160 |

| | | | | | |
|---|---|---|---|---|---|
| | | CGG | | TT | |
| MAX.chr5.77268672-77268725-FS | 199 | TAT TTT ATA GTC GCG TTA AAA GCG T | 161 | GTC GAT AAA AAA CCT ACG CGA CGA A | 162 |
| MAX.chr6.157557371-157557657-FS | 204 | GAT TTA GTT TTT CGG GTT TAT AGC GG | 163 | TAT TAA AAA CGA CCA AAC CTC CGC A | 164 |
| MAX.chr8.124173030-124173395-FS | 208 | TGG TTG TAG GCG TTT TGT TGG AGT TC | 165 | AAA AAC GAC CCT AAC CAC CCT CGT T | 166 |
| MCF2L2-FS | 216 | TTT TGC GTA GTT GGG TAG GGT TCG G | 167 | CCC GCA TTC CCG AAA AAA ACG AT | 168 |
| MCF2L2-RS | 216 | TTA GGG TTT TTT TCG AGG AGT TCG A | 169 | ATC CCC CGT ACG AAA CTA AAC GCG | 170 |
| MCF2L2-RS | 216 | GCG TTC GTA TTT TCG GGA GAG GC | 171 | TCT ACG TAA CTA AAC AAA ACC CGA A | 172 |
| MIB2-FS | 219 | CGT TTT GTG TTT TAT AAA AAG AAA GAT TTT CGG | 173 | AAA ACC CCA AAA ACG CCC GAT | 174 |
| MPZ-FS | 221 | GGG GCG TAT ATA TTA GTT ATC GAG CGA | 175 | AAA AAA AAC CCT AAA AAC CGC CGA A | 176 |
| MSX2P1-FS | 222 | TTC GTT TAA TGA GAA GGG GTT AGC GG | 177 | TAA AAC AAA CTA AAA ACC TTA ACG CGA CGC T | 178 |
| NACAD-RS | 223 | GGG GAG GGA GTT TTT TTT AC | 179 | GTA CGC GAA CTC GCC AAA CAC TAC G | 180 |
| ODC1-FS | 231 | GTA GGG TTG GTA GTC GTT TTT ACG T | 181 | AAC CCA TCT AAT TAC AAA ATA CCT CGA T | 182 |
| ODC1-RS | 231 | GGT TTT ATA GGG GAA ATT ATT TTC GT | 183 | AAA ACC TCG TCT TTA TAA CAT CGA A | 184 |
| ODC1-RS | 231 | TAG GAT ATT TCG ATG TTA TAA AGA CGA | 185 | AAC AAA ACT AAC AAC CGC CTC CAC G | 186 |
| OSR2_A | 234 | TTT GGA GTT ATC GGA AGG CGA AAG TAC | 187 | GCA CGC CGA AAA AAT AAA AAC GAA | 188 |
| OTX1-RS | 237 | TTT TCG ATA TCG ATA TCG AAG GCG T | 189 | ATA ACT TAA AAC CCT AAA TTC CGC C | 190 |
| PAQR6-FS | 238 | GCG GGT AGT AGG AAG ATT AGT AGC GG | 191 | CCG ACT TCC GTA CGA AAC CGT A | 192 |
| PLXNC1_A | 245 | TAA TAG AGG TTT GCG TTG GAA TCG A | 193 | AAC GCA CCC TAA ACA AAA CCA CGA C | 194 |
| PLXNC1_B | 246 | TGA AGA GTT GTT AGT TCG TTT AGC GT | 195 | GCC AAA AAT TCG ATT CCA ACG CA | 196 |
| PPARA-FS | 248 | TAG TGG TAG GTA TAG TTG GTA GCG G | 197 | ATC AAA ACT CCC CTC CTC GAA AAC G | 198 |

| | | | | | |
|---|---|---|---|---|---|
| PPARG-RS | 249 | GTT TTT AAG CGG CGG TCG T | 199 | AAA AAA AAT CCC GTT CGC T | 200 |
| PRKCB-RS | 256 | GCG CGC GTT TAT TAG ATG AAG TCG | 201 | AAA ATC AAA AAC CAC AAA TTC ACC GCC | 202 |
| PRMT1-FS | 257 | CGG GGA GAG GAG GGG TAG GAT TTA C | 203 | CAA CTT AAA CAC CAC TTC CTC CGA A | 204 |
| RBFOX3_A | 262 | TGT TTT TTT TGT TCG GGC GG | 205 | AAA TAA CTA ACT CCT ACT CTC GCC CGC T | 206 |
| RFX8-FS | 264 | ATA GTT TTT TAA TTT TCG CGT TTC GTC GA | 207 | AAA AAC AAC TCC AAC CCA CAC CGC | 208 |
| RIC3-RS | 266 | GCG GGA GGA GTA GGT TAA TTT TCG A | 209 | AAA AAC AAA ATA CGC GAA ACG CAC G | 210 |
| SCRT2_B | 273 | CGA GAA GGT TTT GTC GTA GAC GTC GT | 211 | TAC GTA TCC ATA CCC GCG CTC G | 212 |
| SLC16A3-FS | 276 | TTT GTT TGT ATA ATA GGG GTT GCG G | 213 | CGC CTA ACT ACC GAA AAA TAC CGA A | 214 |
| SLC22A20-FS | 277 | GGT GGG GTT ATT TTT TTA TGG AGT CGA TTC | 215 | CGA ACC AAA CCT ACG ATT CCC GAA | 216 |
| SLC2A2-RS | 278 | GGG AGA AGA GAA TGG TTT TTT GTC GTC | 217 | TCT TAT ACT CAA CCC CGA CCT ACC GAC | 218 |
| SLC30A10-FS | 279 | GTT TTA TTC GGG GTT TTA GCG TTA TTT ACG G | 219 | AAA AAA CCG CGT TAC TCA ACG CGC | 220 |
| SLC7A4-RS | 280 | GTT TAG AGC GGA GGT AGC GGT TGC | 221 | CGC CTA TTC TTA AAC CTA AAC CCG TC | 222 |
| SLITRK5-FS | 282 | CGT AGA GGA TTA TAA AGA TTT GTA CGA | 223 | TAC TAT AAC TAC TAC GAT AAC GAC GAC GAC | 224 |
| SPHK2-RS | 284 | AGA TTT CGG TTT TTG TTT CGA TTT TCG T | 225 | ATT AAT ACT AAC TTA CGA AAC CGC C | 226 |
| ST8SIA4-RS | 285 | ATT ATT TTT GAG CGT GAA AAA TCG T | 227 | AAA TTT CTC TCC AAT TAA ATT CCG TA | 228 |
| STAC2_B | 287 | GTG GGT TTG TCG TCG GAT TTC G | 229 | AAA TAA CCG CGT CAT CCG ATT CGT T | 230 |
| STX16_A | 288 | TGG ATG TTT TAT ATT AAT TTT TAG TTG TAT AAC G | 231 | GTA CTT TTT CTC TCA CGA AAA ATA TTC CCG C | 232 |
| STX16_B | 289 | TGC GTG GAA TAA ATT TTA TAT ACG T | 233 | GCT CAA CAC ACG AAA AAC CCT CGA A | 234 |
| STX16_B | 289 | CGG TGC GGG GTT TTA ATA AAG GAT C | 235 | TCC ACG CAA AAA CAA AAA ACG CGT A | 236 |
| SYNJ2-FS | 291 | GGC GTA GTT ATG ATT TCG TTT TTT | 237 | ATC CTT TCG ACC CTA CGT ACC TCG | 238 |

| | | | | | |
|---|---|---|---|---|---|
| | | CGT | | AT | |
| TBX1-FS | 295 | TTT ACG ATT ATT GTT TTA GAT AAT ACG G | 239 | GAA CCC GAC GAA CTT CGA A | 240 |
| TMEM176A-FS | 300 | GGG AAA TCG CGT AGT TTG GGC | 241 | AAA ACG ACG AAA AAA CGA AAA CGA C | 242 |
| TNFRSF10D-FS | 301 | AGT TAT CGC GAT CGG TTT GGG TTA AC | 243 | AAA CGA TTA CCT CTT TCG TTC GTT CGT T | 244 |
| TRH_A | 303 | CGG CGG TTT ATT TGA AGA GGG TTC | 245 | CGA CAA ATC AAA AAT CTA CAA CGC T | 246 |
| TRIM67-RS | 305 | TTT TAA CGT TAG TTA CGA GTT GCG G | 247 | CGA ACA AAC CAA ACA ACC GAA | 248 |
| UBTF-RS | 310 | GTA GAT TAG GCG GGG GCG A | 249 | GAA CAA AAA CAT AAA CTA ATA CAA ATA TCT CCC G | 250 |
| ZSCAN12-FS | 327 | GGA GGG AGA GTT TTT CGC GGA TTC | 251 | CTA AAC CCC TCA AAC CCT AAC CGA T | 252 |
| GRASP | 105 | TGT TTT CGG ATA CGG CGA GC | 253 | ACG AAC GAA CTA TAC GCG ACG CT | 254 |

**Table 11.** Table 11 shows area under the curve for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal breast tissue.

| Gene Annotation | Basal-like / Triple Negative | HER2+ | Luminal A | Luminal B | BRCA1 | BRCA2 | DMR No. |
|---|---|---|---|---|---|---|---|
| BHLHE23_C | 0.75 | 0.93567 | 0.80392 | 0.74728 | 0.82716 | 0.70782 | 23 |
| CALN1_A | 0.89699 | 0.86842 | 0.73638 | 0.98039 | 0.95679 | 0.68724 | 36 |
| CD1D | 0.66204 | 0.82066 | 0.91503 | 0.78431 | 0.83951 | 0.66667 | 47 |
| CHST2_A | 0.65972 | 0.8655 | 0.94118 | 0.66231 | 0.69753 | 0.63374 | 56 |
| FAM126A | 0.36806 | 0.64717 | 0.86928 | 0.53377 | 0.59877 | 0.72428 | 83 |
| FMN2 | 0.55324 | 0.93762 | 0.8976 | 0.7037 | 0.75926 | 0.65432 | 94 |
| HOXA1_A | 0.60417 | 0.94152 | 0.81264 | 0.57734 | 0.56173 | 0.58848 | 117 |
| HOXA7_A | 0.52315 | 0.95906 | 0.84423 | 0.76144 | 0.82716 | 0.73251 | 119 |
| KCNH8 | 0.5463 | 0.96881 | 0.85839 | 0.72985 | 0.73457 | 0.78189 | 138 |
| LOC100132891 | 0.81019 | 0.9883 | 0.94771 | 0.80174 | 0.8642 | 0.75309 | 153 |
| MAX.chr1.8277479-8277527 | 0.68981 | 0.66569 | 0.73203 | 0.69172 | 0.99383 | 0.607 | 166 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MAX.chr15.96889013-96889128 | 0.68287 | 0.93372 | 0.82353 | 0.88235 | 0.98765 | 0.80247 | 173 |
| NACAD | 0.86111 | 0.8577 | 0.7146 | 0.70806 | 0.84568 | 0.7284 | 223 |
| SLC30A10 | 0.64352 | 0.77388 | 0.94989 | 0.57516 | 0.58333 | 0.77366 | 279 |
| TRIM67 | 0.81134 | 0.97856 | 0.88126 | 0.77015 | 0.73457 | 0.71811 | 305 |
| ATP6V1B1 | 0.8912 | 0.8616 | 0.83007 | 0.82789 | 1 | 0.81893 | 15 |
| BANK1 | 0.75231 | 0.70175 | 0.83878 | 0.66231 | 0.91975 | 0.79835 | 17 |
| C10orf125 | 0.39352 | 0.76706 | 0.94444 | 0.70806 | 0.73765 | 0.7037 | 27 |
| C17orf64 | 0.58333 | 0.97856 | 0.85185 | 0.81699 | 0.81173 | 0.86831 | 29 |
| CHST2_B | 0.59722 | 0.91228 | 0.89325 | 0.68192 | 0.64506 | 0.69547 | 57 |
| DLX4 | 0.86574 | 0.85965 | 0.61547 | 0.83333 | 0.84877 | 0.79424 | 66 |
| DNM3_A | 0.35185 | 0.92203 | 0.92375 | 0.58388 | 0.62346 | 0.65021 | 69 |
| EMX1_A | 0.74537 | 0.91813 | 0.86275 | 0.68627 | 0.94444 | 0.67901 | 74 |
| FOXP4 | 0.7037 | 0.61209 | 0.64488 | 0.58388 | 0.91358 | 0.53086 | 96 |
| GP5 | 0.87731 | 0.77388 | 0.82353 | 0.72113 | 1 | 0.65844 | 104 |
| IGF2BP3_A | 0.64583 | 0.92008 | 0.87364 | 0.75599 | 0.69753 | 0.74897 | 123 |
| ITPRIPL1 | 0.94676 | 1 | 0.91285 | 0.94336 | 1 | 0.95473 | 134 |
| KLHDC7B | 0.63194 | 0.5614 | 0.58606 | 0.66449 | 0.42593 | 0.55144 | 146 |
| LMX1B_A | 0.77083 | 0.81871 | 0.878 | 0.76253 | 0.80864 | 0.79012 | 149 |
| MAX.chr11.14926602-14927148 | 0.98611 | 0.93567 | 0.94553 | 0.94553 | 1 | 0.98354 | 168 |
| MAX.chr5.42994866-42994936 | 0.84259 | 0.91618 | 0.90196 | 0.93682 | 0.93827 | 0.95885 | 198 |
| MAX.chr8.124173030-124173395 | 0.87963 | 0.8577 | 0.87146 | 0.89542 | 0.96914 | 0.77778 | 208 |
| MPZ | 0.9294 | 0.98246 | 0.93682 | 0.88344 | 0.86111 | 0.79835 | 221 |
| ODC1 | 0.3588 | 0.89474 | 0.83442 | 0.60349 | 0.46914 | 0.55967 | 231 |
| PLXNC1_A | 0.61806 | 0.83626 | 0.8976 | 0.57952 | 0.67901 | 0.65844 | 245 |
| PRKCB | 0.91204 | 0.96491 | 0.99782 | 0.97603 | 0.88272 | 0.78189 | 256 |
| ST8SIA4 | 0.84722 | 0.47173 | 0.80392 | 0.66885 | 0.57407 | 0.56379 | 285 |
| STX16_B | 0.84259 | 0.7115 | 0.80174 | 0.71895 | 0.98765 | 0.59259 | 289 |
| UBTF | 0.69676 | 0.67836 | 0.91939 | 0.68192 | 0.83333 | 0.76132 | 310 |
| LOC100132891 | 0.66898 | 0.94542 | 0.93682 | 0.79303 | 0.96914 | 0.83951 | 153 |
| ITPRIPL1 | 0.88657 | 0.98051 | 0.90741 | 0.87364 | 0.99383 | 0.86008 | 134 |
| ABLIM1 | 0.7662 | 0.91618 | 0.79303 | 0.67756 | 0.83951 | 0.83128 | 3 |
| KLF16 | 0.91898 | 0.75049 | 0.64924 | 0.9085 | 0.76543 | 0.66255 | 145 |
| MAX.chr12.4273906-4274012 | 0.83796 | 0.88499 | 0.86928 | 0.94336 | 0.72222 | 0.74074 | 170 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MAX.chr12.59990671-59990859 | 0.68056 | 0.89084 | 0.77342 | 0.53595 | 0.54938 | 0.60494 | 171 |
| MAX.chr19.46379903-46380197 | 0.72801 | 0.96296 | 0.79085 | 0.83987 | 0.91975 | 0.82099 | 179 |
| ZSCAN12 | 0.76852 | 0.88694 | 0.74946 | 0.78214 | 0.82099 | 0.70782 | 327 |
| AADAT | 0.49537 | 0.7193 | 0.58606 | 0.56427 | 0.69136 | 0.61317 | 2 |
| BHLHE23_D | 0.65972 | 0.85185 | 0.86492 | 0.84314 | 0.85802 | 0.70782 | 24 |
| COL23A1 | 0.66898 | 0.89669 | 0.82353 | 0.68954 | 0.49074 | 0.92181 | 60 |
| CXCL12 | 0.56019 | 0.70565 | 0.66231 | 0.59695 | 0.85802 | 0.72428 | 63 |
| KCNK9 | 0.80556 | 0.88499 | 0.82789 | 0.67756 | 0.84568 | 0.70782 | 141 |
| LAYN | 0.55787 | 0.96686 | 0.76471 | 0.63834 | 0.62963 | 0.84774 | 147 |
| OTX1 | 0.60648 | 0.78363 | 0.84749 | 0.69717 | 0.98765 | 0.81481 | 237 |
| PLXNC1_A | 0.70718 | 0.85673 | 0.91068 | 0.62854 | 0.67284 | 0.68519 | 245 |
| RIC3 | 0.78009 | 0.90643 | 0.74292 | 0.7756 | 0.83951 | 0.69136 | 266 |
| SCRT2_B | 0.91319 | 0.95517 | 0.73638 | 0.7658 | 0.91975 | 0.47119 | 273 |
| IGF2BP3_B | 0.62037 | 0.96296 | 0.87582 | 0.66885 | 0.73457 | 0.7572 | 124 |
| MAX.chr17.73073682-73073814 | 0.78009 | 0.67836 | 0.59913 | 0.52505 | 0.88889 | 0.73251 | 174 |
| TBX1 | 0.45139 | 0.49708 | 0.75163 | 0.69499 | 0.48765 | 0.46914 | 295 |
| ALOX5 | 0.44676 | 0.91618 | 0.76688 | 0.60784 | 0.47531 | 0.74486 | 9 |
| ASCL2 | 0.82899 | 0.92271 | 0.77101 | 0.48913 | 0.82609 | 0.6087 | 14 |
| CDH4_E | 0.81597 | 0.94639 | 0.84205 | 0.82789 | 0.80556 | 0.70782 | 53 |
| MAST1 | 0.91304 | 0.95411 | 0.7971 | 0.84511 | 0.87681 | 0.82298 | 160 |
| MAX.chr20.1784209-1784461 | 0.57101 | 0.9686 | 0.90145 | 0.6481 | 0.49275 | 0.65839 | 185 |
| RBFOX3_A | 0.75652 | 0.92271 | 0.86957 | 0.83967 | 0.7029 | 0.58385 | 262 |
| TRH_A | 0.97222 | 0.94347 | 0.97168 | 0.86275 | 1 | 0.79012 | 303 |
| HNF1B_B | 0.78472 | 0.88109 | 0.83007 | 0.71895 | 0.65432 | 0.73663 | 116 |
| MAX.chr12.4273906-4274012 | 0.89855 | 0.92271 | 0.90725 | 0.89258 | 0.69565 | 0.63354 | 170 |
| GAS7 | 0.77391 | 0.89614 | 0.82319 | 0.74936 | 0.58696 | 0.46584 | 99 |
| MAX.chr5.145725410-145725459 | 0.90725 | 0.99758 | 0.92319 | 0.85166 | 0.85145 | 0.85714 | 193 |
| MAX.chr5.77268672-77268725 | 0.85797 | 0.98551 | 0.91304 | 0.90537 | 0.86232 | 0.73292 | 199 |
| GYPC_B | 0.68986 | 0.93478 | 0.96232 | 0.72379 | 0.32609 | 0.50932 | 109 |
| DLX6 | 0.56667 | 0.91063 | 0.87246 | 0.58951 | 0.70652 | 0.76708 | 67 |
| FBN1 | 0.61739 | 0.87077 | 0.94058 | 0.65473 | 0.58333 | 0.79814 | 91 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| OSR2_A | 0.71304 | 0.89614 | 0.90145 | 0.75703 | 0.63043 | 0.74534 | 234 |
| BEST4 | 0.69275 | 0.96377 | 0.84058 | 0.91049 | 0.81884 | 0.69565 | 20 |
| AJAP1_B | 0.76232 | 0.84058 | 0.76232 | 0.91816 | 0.73188 | 0.73913 | 7 |
| DSCR6 | 0.98261 | 0.92512 | 0.87246 | 0.86445 | 0.85507 | 0.76398 | 72 |
| MAX.chr11.68622869-68622968 | 0.50145 | 0.98792 | 0.95362 | 0.7289 | 0.53623 | 0.81366 | 169 |

**Table 12.** Table 12 shows area under the curve for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal buffy coat.

| Gene Annotation | Basal-like | HER2+ | Luminal A | Luminal B | BRCA1 | BRCA2 | DMR No. |
|---|---|---|---|---|---|---|---|
| BHLHE23_C | 0.83594 | 0.98026 | 0.89338 | 0.82721 | 0.875 | 0.80556 | 23 |
| CALN1_A | 0.93555 | 0.92599 | 0.84926 | 1 | 1 | 0.88194 | 36 |
| CD1D | 0.74609 | 0.86184 | 0.93015 | 0.86397 | 0.92708 | 0.76389 | 47 |
| CHST2_A | 0.73828 | 0.88651 | 0.95772 | 0.73162 | 0.76563 | 0.68056 | 56 |
| FAM126A | 0.75781 | 0.89474 | 0.93199 | 0.84743 | 0.91667 | 0.86458 | 83 |
| FMN2 | 0.77344 | 0.96053 | 0.95221 | 0.78676 | 0.9375 | 0.79861 | 94 |
| HOXA1_A | 0.74609 | 0.97039 | 0.93015 | 0.64706 | 0.64583 | 0.66667 | 117 |
| HOXA7_A | 0.54297 | 0.98684 | 0.87868 | 0.79044 | 0.91667 | 0.84028 | 119 |
| KCNH8 | 0.57031 | 0.98684 | 0.87132 | 0.76287 | 0.77083 | 0.77778 | 138 |
| LOC100132891 | 0.80859 | 0.99342 | 0.95221 | 0.77941 | 0.875 | 0.70139 | 153 |
| MAX.chr1.8277479-8277527 | 0.82031 | 0.80592 | 0.86397 | 0.87132 | 1 | 0.83333 | 166 |
| MAX.chr15.96889013-96889128 | 0.87109 | 0.99013 | 0.95956 | 0.95588 | 1 | 0.97917 | 173 |
| NACAD | 0.89453 | 0.91118 | 0.80147 | 0.73529 | 0.90625 | 0.75 | 223 |
| SLC30A10 | 0.75391 | 0.91118 | 0.98897 | 0.65074 | 0.59375 | 0.91667 | 279 |
| TRIM67 | 0.80664 | 1 | 0.8989 | 0.76654 | 0.69792 | 0.78125 | 305 |
| ATP6V1B1 | 1 | 1 | 1 | 1 | 1 | 1 | 15 |
| BANK1 | 0.99609 | 0.99671 | 1 | 0.98529 | 1 | 1 | 17 |
| C10orf125 | 0.70313 | 0.88816 | 0.95588 | 0.81618 | 0.875 | 0.81944 | 27 |
| C17orf64 | 0.78906 | 0.99671 | 0.94853 | 0.93015 | 0.84375 | 0.95833 | 29 |
| CHST2_B | 0.64453 | 0.92105 | 0.91176 | 0.72426 | 0.67188 | 0.70833 | 57 |
| DLX4 | 0.93945 | 0.9227 | 0.79596 | 0.94853 | 0.92188 | 0.90278 | 66 |
| DNM3_A | 0.71875 | 0.98684 | 0.96324 | 0.79779 | 0.79167 | 0.85417 | 69 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| EMX1_A | 0.80469 | 0.92763 | 0.90441 | 0.75735 | 0.98958 | 0.72917 | 74 |
| FOXP4 | 1 | 1 | 1 | 1 | 1 | 1 | 96 |
| GP5 | 1 | 1 | 1 | 0.99265 | 1 | 1 | 104 |
| IGF2BP3_A | 0.65625 | 0.92105 | 0.86949 | 0.75919 | 0.71354 | 0.75694 | 123 |
| ITPRIPL1 | 0.94922 | 1 | 0.90809 | 0.95956 | 1 | 0.95833 | 134 |
| KLHDC7B | 1 | 1 | 1 | 0.98897 | 1 | 0.98611 | 146 |
| LMX1B_A | 1 | 0.99671 | 1 | 1 | 1 | 1 | 149 |
| MAX.chr11.14926602-14927148 | 1 | 0.99342 | 1 | 0.99632 | 1 | 1 | 168 |
| MAX.chr5.42994866-42994936 | 0.94922 | 0.98026 | 0.98529 | 0.99265 | 1 | 1 | 198 |
| MAX.chr8.124173030-124173395 | 0.98438 | 0.91118 | 0.98162 | 0.98897 | 1 | 0.98611 | 208 |
| MPZ | 0.94922 | 0.99342 | 0.96691 | 0.94485 | 0.90625 | 0.86111 | 221 |
| ODC1 | 0.89453 | 0.99013 | 0.97794 | 0.90074 | 0.84375 | 0.875 | 231 |
| PLXNC1_A | 0.94531 | 0.94408 | 0.97794 | 0.88235 | 0.89583 | 0.92361 | 245 |
| PRKCB | 0.97266 | 0.98684 | 1 | 1 | 0.9375 | 0.86806 | 256 |
| ST8SIA4 | 0.99219 | 1 | 1 | 0.98162 | 1 | 0.99306 | 285 |
| STX16_B | 1 | 1 | 1 | 1 | 1 | 1 | 289 |
| UBTF | 0.99609 | 0.99671 | 1 | 0.99632 | 1 | 0.92361 | 310 |
| LOC100132891 | 0.8125 | 0.97697 | 1 | 0.84926 | 1 | 0.90278 | 153 |
| ITPRIPL1 | 0.90234 | 1 | 0.95404 | 0.95221 | 1 | 0.9375 | 134 |
| ABLIM1 | 0.83398 | 0.9523 | 0.8511 | 0.75919 | 0.92708 | 0.88542 | 3 |
| KLF16 | 1 | 0.87171 | 0.83456 | 1 | 0.84375 | 0.8125 | 145 |
| MAX.chr12.4273906-4274012 | 0.90625 | 0.94737 | 0.97059 | 0.97059 | 0.83333 | 0.94444 | 170 |
| MAX.chr12.59990671-59990859 | 0.78516 | 0.92434 | 0.85662 | 0.63235 | 0.69792 | 0.6875 | 171 |
| MAX.chr19.46379903-46380197 | 0.78125 | 0.97697 | 0.82353 | 0.88603 | 0.94792 | 0.86111 | 179 |
| ZSCAN12 | 0.76758 | 0.88487 | 0.75 | 0.78125 | 0.82292 | 0.70486 | 327 |
| AADAT | 0.76172 | 0.85855 | 0.80147 | 0.71691 | 0.89583 | 0.81944 | 2 |
| BHLHE23_D | 0.71875 | 0.86184 | 0.89706 | 0.87868 | 0.85417 | 0.77778 | 24 |
| COL23A1 | 0.67969 | 0.90461 | 0.8125 | 0.69118 | 0.47917 | 0.93056 | 60 |
| CXCL12 | 0.96875 | 0.97697 | 0.95221 | 0.87132 | 0.98958 | 0.99306 | 63 |
| KCNK9 | 0.92188 | 0.92434 | 0.91912 | 0.71691 | 0.9375 | 0.70833 | 141 |
| LAYN | 0.56641 | 0.97039 | 0.75735 | 0.65441 | 0.625 | 0.84722 | 147 |
| OTX1 | 0.99219 | 1 | 1 | 0.99632 | 1 | 1 | 237 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| PLXNC1_A | 0.81445 | 0.90625 | 0.95956 | 0.76287 | 0.71875 | 0.80208 | 245 |
| RIC3 | 0.85352 | 0.96711 | 0.82353 | 0.83824 | 0.89583 | 0.77778 | 266 |
| SCRT2_B | 0.93359 | 0.98684 | 0.83088 | 0.8364 | 0.97917 | 0.61806 | 273 |
| IGF2BP3_B | 0.72656 | 0.97204 | 0.90257 | 0.74632 | 0.8125 | 0.81597 | 124 |
| MAX.chr17.73073682-73073814 | 1 | 1 | 0.93934 | 0.98529 | 1 | 1 | 174 |
| TBX1 | 0.99609 | 1 | 1 | 1 | 1 | 1 | 295 |
| ALOX5 | 0.77539 | 0.99671 | 0.87316 | 0.7739 | 0.73958 | 0.81944 | 9 |
| ASCL2 | 0.85778 | 0.92593 | 0.79556 | 0.575 | 0.82222 | 0.65714 | 14 |
| CDH4_E | 0.87891 | 0.95724 | 0.85294 | 0.85294 | 0.84375 | 0.75694 | 53 |
| MAST1 | 0.90667 | 0.96296 | 0.76444 | 0.85833 | 0.86667 | 0.83333 | 160 |
| MAX.chr20.1784209-1784461 | 0.71556 | 0.98889 | 0.96 | 0.73125 | 0.63333 | 0.74286 | 185 |
| RBFOX3_A | 0.72 | 0.90741 | 0.8 | 0.7625 | 0.64444 | 0.51429 | 262 |
| TRH_A | 1 | 0.99342 | 1 | 0.90809 | 1 | 0.97222 | 303 |
| HNF1B_B | 1 | 1 | 1 | 1 | 1 | 0.95139 | 116 |
| MAX.chr12.4273906-4274012 | 0.9875 | 0.98611 | 0.95 | 0.92647 | 0.79167 | 0.88393 | 170 |
| GAS7 | 0.93333 | 0.97917 | 0.91458 | 0.86213 | 0.75 | 0.79911 | 99 |
| MAX.chr5.145725410-145725459 | 0.94583 | 0.98264 | 0.92708 | 0.87868 | 0.86979 | 0.91071 | 193 |
| MAX.chr5.77268672-77268725 | 0.95417 | 0.99306 | 0.96667 | 0.95588 | 0.92708 | 0.875 | 199 |
| GYPC_B | 0.90833 | 0.99306 | 1 | 0.87132 | 0.76042 | 0.85714 | 109 |
| DLX6 | 0.70208 | 0.96528 | 0.95 | 0.7261 | 0.83854 | 0.85268 | 67 |
| FBN1 | 0.41667 | 0.85417 | 0.94583 | 0.65074 | 0.5625 | 0.79464 | 91 |
| OSR2_A | 0.90417 | 1 | 1 | 0.96507 | 0.95833 | 0.91518 | 234 |
| BEST4 | 0.69167 | 0.93403 | 0.825 | 0.87132 | 0.80208 | 0.74107 | 20 |
| AJAP1_B | 0.675 | 0.71875 | 0.675 | 0.82353 | 0.65625 | 0.39286 | 7 |
| DSCR6 | 0.97917 | 0.93056 | 0.8625 | 0.89706 | 0.875 | 0.80357 | 72 |
| MAX.chr11.68622869-68622968 | 0.71458 | 0.99653 | 0.97917 | 0.88603 | 0.70313 | 0.88393 | 169 |

**Table 13.** Table 13 shows methylation fold change for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal breast tissue.

| Gene Annotation | Basal-like | HER2+ | Luminal A | Luminal B | BRCA1 | BRCA2 | DMR No. |
|---|---|---|---|---|---|---|---|
| BHLHE23_C | 17.39 | 28.60 | 15.89 | 8.07 | 21.97 | 5.13 | 23 |
| CALN1_A | 28.82 | 16.81 | 15.57 | 15.63 | 22.24 | 9.44 | 36 |
| CD1D | 10.77 | 16.99 | 21.18 | 9.33 | 13.48 | 10.65 | 47 |
| CHST2_A | 15.19 | 82.95 | 80.41 | 28.63 | 13.89 | 57.16 | 56 |
| FAM126A | 4.45 | 10.43 | 31.13 | 10.07 | 14.02 | 30.56 | 83 |
| FMN2 | 2.71 | 27.06 | 22.58 | 16.21 | 24.59 | 7.57 | 94 |
| HOXA1_A | 20.01 | 42.46 | 26.87 | 14.45 | 12.63 | 18.71 | 117 |
| HOXA7_A | 9.72 | 23.28 | 15.61 | 8.19 | 6.23 | 7.65 | 119 |
| KCNH8 | 1.78 | 39.32 | 35.88 | 14.75 | 14.13 | 45.90 | 138 |
| LOC100132891 | 185.19 | 312.06 | 194.65 | 143.63 | 186.82 | 180.80 | 153 |
| MAX.chr1.8277479-8277527 | 9.79 | 5.38 | 9.20 | 7.31 | 18.72 | 6.42 | 166 |
| MAX.chr15.96889013-96889128 | 6.87 | 7.87 | 5.80 | 5.54 | 9.78 | 5.97 | 173 |
| NACAD | 59.72 | 97.55 | 3.38 | 20.96 | 21.53 | 22.18 | 223 |
| SLC30A10 | 6.66 | 87.99 | 105.28 | 31.45 | 49.17 | 88.74 | 279 |
| TRIM67 | 86.70 | 69.09 | 53.28 | 43.64 | 83.52 | 32.16 | 305 |
| ATP6V1B1 | 3.95 | 2.43 | 2.52 | 2.18 | 3.75 | 2.52 | 15 |
| BANK1 | 2.21 | 1.49 | 1.81 | 1.48 | 2.17 | 2.14 | 17 |
| C10orf125 | 3.79 | 18.09 | 19.60 | 13.34 | 16.57 | 12.82 | 27 |
| C17orf64 | 9.30 | 42.58 | 15.72 | 11.05 | 11.18 | 24.94 | 29 |
| CHST2_B | 48.81 | 281.00 | 298.32 | 116.70 | 135.51 | 191.68 | 57 |
| DLX4 | 45.64 | 60.68 | 16.25 | 29.97 | 41.18 | 29.75 | 66 |
| DNM3_A | 8.86 | 23.78 | 30.08 | 11.49 | 18.30 | 21.56 | 69 |
| EMX1_A | 43.60 | 122.29 | 82.84 | 25.57 | 49.07 | 59.72 | 74 |
| FOXP4 | 1.88 | 0.87 | 1.24 | 0.91 | 1.66 | 1.15 | 96 |
| GP5 | 4.39 | 2.30 | 2.78 | 2.43 | 6.32 | 2.58 | 104 |
| IGF2BP3_A | 20.99 | 38.55 | 29.02 | 33.09 | 12.02 | 26.13 | 123 |
| ITPRIPL1 | 68.12 | 53.72 | 43.47 | 51.51 | 63.79 | 47.41 | 134 |
| KLHDC7B | 1.31 | 0.91 | 1.09 | 0.79 | 1.00 | 0.91 | 146 |
| LMX1B_A | 2.06 | 2.17 | 2.14 | 1.44 | 2.08 | 2.32 | 149 |
| MAX.chr11.14926602-14927148 | 36.24 | 22.85 | 19.10 | 23.87 | 28.89 | 17.02 | 168 |
| MAX.chr5.42994866-42994936 | 19.79 | 13.50 | 9.94 | 11.94 | 21.43 | 12.21 | 198 |
| MAX.chr8.124173030- | 21.90 | 24.50 | 10.92 | 14.77 | 28.87 | 24.86 | 208 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 124173395 | | | | | | | |
| MPZ | 123.49 | 79.48 | 48.93 | 74.77 | 131.51 | 63.10 | 221 |
| ODC1 | 1.78 | 8.69 | 8.19 | 4.66 | 1.97 | 7.56 | 231 |
| PLXNC1_A | 7.89 | 13.76 | 9.63 | 8.37 | 12.31 | 18.75 | 245 |
| PRKCB | 26.21 | 38.24 | 34.42 | 25.44 | 26.52 | 11.10 | 256 |
| ST8SIA4 | 0.48 | 1.05 | 1.65 | 0.65 | 1.13 | 1.44 | 285 |
| STX16_B | 3.01 | 1.93 | 1.96 | 2.16 | 3.90 | 2.88 | 289 |
| UBTF | 1.79 | 1.77 | 2.71 | 1.50 | 1.93 | 2.92 | 310 |
| LOC100132891 | 9.99 | 14.86 | 9.71 | 8.33 | 11.87 | 10.56 | 153 |
| ITPRIPL1 | 61.84 | 47.86 | 33.92 | 39.52 | 46.55 | 32.55 | 134 |
| ABLIM1 | 126.47 | 140.31 | 63.32 | 44.29 | 132.06 | 117.33 | 3 |
| KLF16 | 25.67 | 7.34 | 4.47 | 7.38 | 10.97 | 6.36 | 145 |
| MAX.chr12.4273906-4274012 | 301.23 | 105.51 | 111.39 | 153.60 | 171.56 | 6.10 | 170 |
| MAX.chr12.59990671-59990859 | 28.29 | 52.80 | 32.14 | 14.31 | 24.76 | 57.36 | 171 |
| MAX.chr19.46379903-46380197 | 22.21 | 55.50 | 38.80 | 19.14 | 43.26 | 63.39 | 179 |
| ZSCAN12 | 10284.41 | 4154.53 | 78.06 | 4637.92 | 2760.20 | 188.89 | 327 |
| AADAT | 1.31 | 6.34 | 21.38 | 14.40 | 1.47 | 1.94 | 2 |
| BHLHE23_D | 158.47 | 264.51 | 122.98 | 34.78 | 124.83 | 48.69 | 24 |
| COL23A1 | 3.98 | 24.50 | 27.84 | 17.76 | 1.63 | 29.07 | 60 |
| CXCL12 | 2.86 | 10.36 | 6.15 | 6.61 | 11.71 | 26.51 | 63 |
| KCNK9 | 48.27 | 85.89 | 94.77 | 21.69 | 46.39 | 130.44 | 141 |
| LAYN | 16.27 | 55.56 | 25.44 | 23.04 | 21.03 | 38.03 | 147 |
| OTX1 | 2.80 | 2.55 | 2.75 | 2.26 | 6.13 | 4.22 | 237 |
| PLXNC1_A | 22.95 | 41.20 | 28.28 | 29.16 | 37.81 | 53.43 | 245 |
| RIC3 | 75.70 | 47.95 | 37.52 | 32.73 | 77.58 | 16.49 | 266 |
| SCRT2_B | 164.74 | 109.65 | 63.97 | 81.04 | 101.97 | 5.69 | 273 |
| IGF2BP3_B | 185.90 | 412.90 | 282.64 | 212.43 | 385.38 | 225.97 | 124 |
| MAX.chr17.73073682-73073814 | 1.73 | 0.83 | 0.85 | 1.08 | 1.60 | 1.84 | 174 |
| TBX1 | 1.84 | 1.42 | 0.69 | 0.79 | 1.41 | 2.45 | 295 |
| ALOX5 | 5.56 | 22.45 | 13.11 | 12.76 | 8.52 | 23.22 | 9 |
| ASCL2 | 36.06 | 18.95 | 24.95 | 8.63 | 3.72 | 22.63 | 14 |
| CDH4_E | 85.74 | 88.80 | 89.15 | 55.48 | 180.60 | 66.56 | 53 |
| MAST1 | 143.54 | 82.72 | 21.57 | 16.93 | 93.73 | 52.01 | 160 |
| MAX.chr20.1784209- | 34.65 | 76.68 | 61.27 | 31.39 | 29.19 | 34.67 | 185 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1784461 | | | | | | | |
| RBFOX3_A | 83.13 | 51.12 | 56.20 | 30.74 | 29.32 | 41.81 | 262 |
| TRH_A | 13.50 | 11.59 | 8.32 | 9.56 | 17.69 | 9.88 | 303 |
| HNF1B_B | 3.02 | 4.34 | 2.55 | 2.40 | 2.62 | 4.08 | 116 |
| MAX.chr12.4273906-4274012 | 114.75 | 55.89 | 51.99 | 73.21 | 59.98 | 9.32 | 170 |
| GAS7 | 60.65 | 32.36 | 32.26 | 48.04 | 33.41 | 9.47 | 99 |
| MAX.chr5.145725410-145725459 | 90.31 | 118.51 | 79.19 | 78.52 | 136.00 | 88.88 | 193 |
| MAX.chr5.77268672-77268725 | 41.72 | 50.96 | 29.95 | 37.74 | 65.41 | 40.86 | 199 |
| GYPC_B | 16.66 | 22.32 | 18.77 | 17.73 | 4.48 | 5.49 | 109 |
| DLX6 | 91.12 | 105.35 | 81.17 | 31.54 | 38.38 | 24.90 | 67 |
| FBN1 | 1.41 | 92.19 | 132.70 | 56.35 | 26.88 | 122.22 | 91 |
| OSR2_A | 42.34 | 72.55 | 32.82 | 45.53 | 77.54 | 58.19 | 234 |
| BEST4 | 71.08 | 73.96 | 61.71 | 99.72 | 85.57 | 64.51 | 20 |
| AJAP1_B | 28.08 | 13.72 | 10.00 | 20.22 | 16.87 | 1.71 | 7 |
| DSCR6 | 53.05 | 52.43 | 20.76 | 33.33 | 36.39 | 35.57 | 72 |
| MAX.chr11.68622869-68622968 | 20.35 | 111.34 | 116.50 | 58.10 | 30.52 | 70.46 | 169 |

**Table 14.** Table 14 shows methylation fold change for the identified 80 methylated regions distinguishing basal / triple negative breast tissue, HER2+ breast tissue, Luminal A breast tissue, Luminal B breast tissue, BRCA1 breast tissue, and BRCA2 breast tissue in comparison with normal buffy coat.

| Gene Annotation | Basal-like | HER2+ | Luminal A | Luminal B | BRCA1 | BRCA2 | DMR No. |
|---|---|---|---|---|---|---|---|
| BHLHE23_C | 128.20 | 210.78 | 117.13 | 59.52 | 161.94 | 37.79 | 23 |
| CALN1_A | 106.73 | 62.24 | 57.65 | 57.89 | 82.38 | 34.97 | 36 |
| CD1D | 20.42 | 32.22 | 40.18 | 17.70 | 25.57 | 20.20 | 47 |
| CHST2_A | 39.29 | 214.65 | 208.06 | 74.08 | 35.94 | 147.92 | 56 |
| FAM126A | 55.82 | 131.03 | 391.00 | 126.44 | 176.02 | 383.76 | 83 |
| FMN2 | 7.46 | 74.48 | 62.15 | 44.63 | 67.68 | 20.83 | 94 |
| HOXA1_A | 65.29 | 138.57 | 87.67 | 47.15 | 41.21 | 61.06 | 117 |
| HOXA7_A | 33.10 | 79.29 | 53.17 | 27.90 | 21.23 | 26.06 | 119 |
| KCNH8 | 2.80 | 61.92 | 56.51 | 23.23 | 22.25 | 72.29 | 138 |
| LOC100132891 | 315.30 | 531.29 | 331.39 | 244.53 | 318.07 | 307.81 | 153 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MAX.chr1.8277479-8277527 | 54.62 | 30.04 | 51.31 | 40.77 | 104.42 | 35.83 | 166 |
| MAX.chr15.96889013-96889128 | 31.27 | 35.82 | 26.39 | 25.21 | 44.51 | 27.19 | 173 |
| NACAD | 7734.43 | 12633.27 | 437.58 | 2714.09 | 2788.77 | 2872.49 | 223 |
| SLC30A10 | 35.40 | 467.99 | 559.93 | 167.25 | 261.52 | 471.97 | 279 |
| TRIM67 | 184.28 | 146.87 | 113.25 | 92.76 | 177.52 | 68.37 | 305 |
| ATP6V1B1 | 135.04 | 83.03 | 86.23 | 74.40 | 128.07 | 86.05 | 15 |
| BANK1 | 104.87 | 70.72 | 85.77 | 70.16 | 102.93 | 101.27 | 17 |
| C10orf125 | 18.08 | 86.39 | 93.58 | 63.71 | 79.10 | 61.20 | 27 |
| C17orf64 | 44.89 | 205.57 | 75.89 | 53.34 | 53.97 | 120.40 | 29 |
| CHST2_B | 188.62 | 1085.94 | 1152.89 | 451.00 | 523.69 | 740.74 | 57 |
| DLX4 | 3920.54 | 5213.08 | 1395.77 | 2574.90 | 3537.32 | 2556.10 | 66 |
| DNM3_A | 61.22 | 164.42 | 207.92 | 79.42 | 126.48 | 149.04 | 69 |
| EMX1_A | 144.48 | 405.29 | 274.54 | 84.73 | 162.63 | 197.93 | 74 |
| FOXP4 | 381.51 | 176.16 | 251.88 | 183.64 | 336.40 | 232.30 | 96 |
| GP5 | 280.65 | 146.74 | 178.00 | 155.65 | 404.01 | 164.80 | 104 |
| IGF2BP3_A | 24.63 | 45.23 | 34.05 | 38.83 | 14.10 | 30.66 | 123 |
| ITPRIPL1 | 84.73 | 66.82 | 54.07 | 64.07 | 79.34 | 58.97 | 134 |
| KLHDC7B | 186.31 | 129.54 | 154.64 | 111.99 | 142.69 | 129.65 | 146 |
| LMX1B_A | 203.26 | 213.63 | 211.34 | 142.05 | 204.63 | 229.02 | 149 |
| MAX.chr11.14926602-14927148 | 497.03 | 313.40 | 261.90 | 327.34 | 396.19 | 233.42 | 168 |
| MAX.chr5.42994866-42994936 | 73.59 | 50.18 | 36.95 | 44.37 | 79.69 | 45.39 | 198 |
| MAX.chr8.124173030-124173395 | 64.77 | 72.44 | 32.29 | 43.69 | 85.38 | 73.51 | 208 |
| MPZ | 249.93 | 160.87 | 99.03 | 151.32 | 266.15 | 127.71 | 221 |
| ODC1 | 30.64 | 149.64 | 141.00 | 80.25 | 33.86 | 130.18 | 231 |
| PLXNC1_A | 106.94 | 186.43 | 130.47 | 113.37 | 166.70 | 253.90 | 245 |
| PRKCB | 75.54 | 110.24 | 99.23 | 73.32 | 76.45 | 31.99 | 256 |
| ST8SIA4 | 38.54 | 83.88 | 132.23 | 51.67 | 90.22 | 115.10 | 285 |
| STX16_B | 475.75 | 305.35 | 309.74 | 340.45 | 615.98 | 455.10 | 289 |
| UBTF | 125.97 | 125.03 | 190.93 | 105.76 | 136.15 | 205.77 | 310 |
| LOC100132891 | 25.07 | 37.29 | 24.36 | 20.90 | 29.79 | 26.49 | 153 |
| ITPRIPL1 | 234.04 | 181.13 | 128.39 | 149.56 | 176.16 | 123.19 | 134 |
| ABLIM1 | 1223.97 | 1357.94 | 612.81 | 428.63 | 1278.01 | 1135.46 | 3 |
| KLF16 | 75.80 | 21.67 | 13.19 | 21.78 | 32.40 | 18.78 | 145 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MAX.chr12.4273906-4274012 | >1x10$^6$ | >1x10$^6$ | >1x10$^6$ | >1x10$^6$ | >1x10$^6$ | >1x10$^6$ | 170 |
| MAX.chr12.59990671-59990859 | 82.48 | 153.92 | 93.68 | 41.71 | 72.17 | 167.20 | 171 |
| MAX.chr19.46379903-46380197 | 52.82 | 132.00 | 92.28 | 45.54 | 102.88 | 150.76 | 179 |
| ZSCAN12 | 2693238 | 1087971 | 20443.01 | 1214559 | 722829 | 49466.69 | 327 |
| AADAT | 2.91 | 14.15 | 47.72 | 32.13 | 3.28 | 4.33 | 2 |
| BHLHE23_D | 272.07 | 454.11 | 211.13 | 59.71 | 214.31 | 83.59 | 24 |
| COL23A1 | 4.68 | 28.80 | 32.73 | 20.87 | 1.91 | 34.18 | 60 |
| CXCL12 | 73.68 | 266.57 | 158.26 | 170.11 | 301.27 | 681.91 | 63 |
| KCNK9 | 88.62 | 157.70 | 174.01 | 39.83 | 85.17 | 239.50 | 141 |
| LAYN | 17.15 | 58.55 | 26.81 | 24.28 | 22.17 | 40.08 | 147 |
| OTX1 | 118.61 | 107.72 | 116.27 | 95.75 | 259.48 | 178.56 | 237 |
| PLXNC1_A | 52.27 | 93.86 | 64.43 | 66.42 | 86.12 | 121.71 | 245 |
| RIC3 | 187.45 | 118.73 | 92.91 | 81.04 | 192.10 | 40.83 | 266 |
| SCRT2_B | 617.81 | 411.22 | 239.92 | 303.91 | 382.41 | 21.34 | 273 |
| IGF2BP3_B | 137.82 | 306.10 | 209.53 | 157.48 | 285.70 | 167.52 | 124 |
| MAX.chr17.73073682-73073814 | 150.82 | 72.51 | 74.12 | 93.99 | 139.52 | 159.80 | 174 |
| TBX1 | 715.28 | 553.15 | 266.62 | 307.86 | 548.03 | 950.84 | 295 |
| ALOX5 | 71.55 | 288.98 | 168.70 | 164.22 | 109.6464 | 298.90 | 9 |
| ASCL2 | 40.30 | 21.18 | 27.89 | 9.65 | 4.1572 | 25.29 | 14 |
| CDH4_E | 314.87 | 326.08 | 327.36 | 203.74 | 663.182 | 244.42 | 53 |
| MAST1 | 232.79 | 134.16 | 34.98 | 27.45 | 152.0212 | 84.36 | 160 |
| MAX.chr20.1784209-1784461 | 88.90 | 196.72 | 157.17 | 80.52 | 74.88755 | 88.94 | 185 |
| RBFOX3_A | 52.90 | 32.53 | 35.76 | 19.56 | 18.6575 | 26.60 | 262 |
| TRH_A | 175.90 | 151.01 | 108.45 | 124.56 | 230.4582 | 128.70 | 303 |
| HNF1B_B | 31.72 | 45.52 | 26.74 | 25.24 | 27.46255 | 42.86 | 116 |
| MAX.chr12.4273906-4274012 | 487.31 | 237.37 | 220.79 | 310.90 | 254.73 | 39.59 | 170 |
| GAS7 | 235.17 | 125.48 | 125.07 | 186.25 | 129.54 | 36.70 | 99 |
| MAX.chr5.145725410-145725459 | 96.45 | 126.57 | 84.58 | 83.86 | 145.26 | 94.93 | 193 |
| MAX.chr5.77268672-77268725 | 96.32 | 117.66 | 69.15 | 87.14 | 151.02 | 94.34 | 199 |
| GYPC_B | 371.40 | 497.60 | 418.40 | 395.42 | 99.91 | 122.49 | 109 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DLX6 | 364.67 | 421.64 | 324.85 | 126.23 | 153.59 | 99.67 | 67 |
| FBN1 | 0.76 | 49.47 | 71.20 | 30.24 | 14.42 | 65.58 | 91 |
| OSR2_A | 2655.22 | 4549.99 | 2058.28 | 2855.38 | 4862.64 | 3648.91 | 234 |
| BEST4 | 13.74 | 14.29 | 11.93 | 19.27 | 16.54 | 12.47 | 20 |
| AJAP1_B | 3.21 | 1.57 | 1.14 | 2.31 | 1.93 | 0.20 | 7 |
| DSCR6 | 51.96 | 51.35 | 20.34 | 32.65 | 35.65 | 34.84 | 72 |
| MAX.chr11.68622869-68622968 | 99.98 | 547.01 | 572.37 | 285.43 | 149.92 | 346.17 | 169 |

**Table 15.** Table 15 shows AUC, average FC, median FC, and p-value distinguishing DCIS high grade and DCIS low grade.

| Gene Annotation | AUC | Average FC | Median FC | p-value | DMR No. |
|---|---|---|---|---|---|
| SCRT2_B | 0.92149 | 43.36 | 49.89 | <.0001 | 273 |
| MPZ | 0.90083 | 11.89 | 23.24 | 0.0001 | 221 |
| MAX.chr8.124173030-124173395 | 0.90083 | 5.62 | 8.24 | 0.0102 | 208 |
| ITPRIPL1 | 0.88017 | 2.47 | 25.81 | 0.0903 | 134 |
| ITPRIPL1 | 0.87603 | 3.04 | 26.71 | 0.036 | 134 |
| DLX4 | 0.85124 | 5.03 | 3.99 | 0.0017 | 66 |
| CALN1_A | 0.82231 | 5.90 | 26.11 | 0.0105 | 36 |
| IGF2BP3_B | 0.81405 | 5.69 | 48.23 | 0.0127 | 124 |
| LOC100132891 | 0.78512 | 2.95 | 4.82 | 0.0495 | 153 |
| MAX.chr5.42994866-42994936 | 0.7562 | 4.85 | 2.10 | 0.0051 | 198 |
| MAX.chr11.14926602-14927148 fp | 0.73554 | 1.57 | 2.51 | 0.1458 | 168 |
| PRKCB | 0.72314 | 1.83 | 2.83 | 0.1702 | 256 |
| EMX1_A | 0.66529 | 1.23 | 21.26 | 0.7382 | 74 |
| DNM3_A | 0.66116 | 1.77 | 2.43 | 0.1205 | 69 |
| CHST2_B | 0.64876 | 3.84 | 5.48 | 0.0531 | 57 |
| C10orf125 | 0.57025 | 1.24 | 1.77 | 0.7431 | 27 |

**Example II.**

This example describes the tissue validation of breast-cancer specific markers. Independent tissue samples (fresh frozen) were selected from institutional cancer registries at Mayo Clinic Rochester and were reviewed by an expert pathologist to confirm correct classification and to guide macro-dissection. Cases comprised 29 triple negative/basal-like, 34 HER2 type, 36 luminal A, and 25 luminal B invasive breast cancers. Also included were 5

BRCA1 and 6 BRCA2 cancers, 21 DCIS w/HGD and 12 DCIS w/LGD. Controls included 27 age matched normal breast tissues and 18 buffy coat samples from normal females.

55 methylated DNA markers (MDMs) were chosen from the list of 80 MDMs (see, Example I and Tables 11-15) which were tested on the discovery samples.

Genomic DNA was prepared using QIAamp DNA Mini Kits (Qiagen, Valencia CA) and bisulfite converted using the EZ-96 DNA Methylation kit (Zymo Research, Irvine CA). Amplification primers were designed from marker sequences using Methprimer software (University of California, San Francisco CA) and synthesized commercially (IDT, Coralville IA). Assays were rigorously tested and optimized by SYBR Green qPCR (Roche) on bisulfite converted (methylated and unmethylated genomic DNA) and unconverted controls. Assays which cross reacted with negative controls were either redesigned or discarded. Melting curve analysis was utilized to ensure specific amplification was occurring.

qMSP was performed using the LightCycler 480 instrument on 2uL of converted DNA in a total reaction volume of 25uL. Standards were derived from serially diluted universal methylated DNA (Zymo Research). Raw marker copies were standardized to CpG-agnostic β-actin, a marker for total genomic DNA.

Results were analyzed logistically using JMP10 (SAS, Cary NC). Cases were compared separately to normal breast controls and normal buffy coat samples. Methylation ratios and absolute differentials were calculated for each of the MDMs.

MDM performance in the independent samples was excellent with many AUCs and methylation fold change ratios (FCs) greater than 0.90 and 50, respectively. Results are provided in Table 16A (Triple Negative), 16B ($HER2^+$), 16C (Luminal A), 16D (Luminal D), and 16E (Overall). Here, the MDMs are ranked by AUC (comparing overall cases to buffy coat samples). This is a critical metric for potential application in plasma as the majority of cell-free DNA (cfDNA) originates with leukocytes. Any MDM which does not highly discriminate epithelial-derived cancers from leukocyte DNA will fail in a blood test format, no matter its performance in tissues. 41 of 55 MDMs had cancer v buffy coat AUCs in excess of 0.9, with 3 achieving perfect discrimination (AUC=1). Tables 16A, 16B, 16C, 16D, and 16E also list AUCs, FCs, p-values, and % cancer methylation as other critical metrics in evaluating and demonstrating the excellence of these MDMs.

Table 17 highlights the top 10 MDMs for discriminating DCIS HGD from DCIS LGD.

**Table 16A.**

| Gene Annotation | Triple Negative AUC | p-value | %meth | FC | DMR No. |
|---|---|---|---|---|---|
| ATP6V1B1 | 0.91799 | <.0001 | 27.64 | 3.27 | 15 |
| FOXP4 | 0.63653 | 0.0061 | 52.73 | 1.53 | 96 |
| LMX1B_A | 0.77446 | <.0001 | 26.28 | 3.32 | 149 |
| BANK1 | 0.72693 | 0.0007 | 23.39 | 1.84 | 17 |
| OTX1 | 0.7987 | <.0001 | 25.43 | 3.46 | 237 |
| ST8SIA4 | 0.69711 | 0.075 | 8.34 | 0.68 | 285 |
| MAX.chr11.14926602-14927148 | 0.99441 | <.0001 | 25.03 | 44.82 | 168 |
| UBTF | 0.64026 | 0.0065 | 25.66 | 1.87 | 310 |
| STX16_B | 0.59459 | 0.0021 | 37.06 | 2.48 | 289 |
| KLHDC7B | 0.58155 | 0.1761 | 23.62 | 1.25 | 146 |
| PRKCB | 0.88816 | <.0001 | 14.86 | 32.90 | 256 |
| TBX1 | 0.44548 | 0.7784 | 14.24 | 1.06 | 295 |
| TRH_A | 0.95433 | <.0001 | 25.50 | 9.67 | 303 |
| MPZ | 0.96319 | <.0001 | 21.79 | 75.65 | 221 |
| GP5 | 0.81081 | <.0001 | 26.05 | 3.53 | 104 |
| DNM3_A | 0.53588 | 0.0003 | 9.43 | 11.61 | 69 |
| MAX.chr17.73073682-73073814 | 0.61696 | 0.0223 | 26.25 | 1.58 | 174 |
| TRIM67 | 0.88164 | <.0001 | 11.53 | 44.19 | 305 |
| PLXNC1_A | 0.56757 | <.0001 | 6.94 | 10.77 | 245 |
| MAX.chr12.4273906-4274012 | 0.95433 | <.0001 | 14.01 | 64.62 | 170 |
| CALN1_A | 0.92637 | <.0001 | 14.36 | 34.56 | 36 |
| ITPRIPL1 | 0.73066 | <.0001 | 12.14 | 26.28 | 134 |
| MAX.chr12.4273906-4274012 | 0.91053 | <.0001 | 10.16 | 299.72 | 170 |
| GYPC_B | 0.69152 | <.0001 | 9.77 | 10.05 | 109 |
| MAX.chr5.42994866-42994936 | 0.86952 | <.0001 | 12.97 | 18.90 | 198 |
| OSR2_A | 0.65284 | <.0001 | 14.65 | 35.06 | 234 |
| SCRT2 | 0.85927 | <.0001 | 10.26 | 78.29 | 273 |
| MAX.chr5.145725410-145725459 | 0.87372 | <.0001 | 8.50 | 43.77 | 193 |
| MAX.chr11.68622869- | 0.67102 | 0.0003 | 9.03 | 10.44 | 169 |

| | | | | | |
|---|---|---|---|---|---|
| 68622968 | | | | | |
| MAX.chr8.124173030-124173395 | 0.79776 | <.0001 | 20.24 | 2.84 | 208 |
| CXCL12 | 0.47717 | 0.0347 | 22.81 | 3.67 | 63 |
| MAX.chr20.1784209-1784461 | 0.64865 | <.0001 | 7.57 | 23.22 | 185 |
| LOC100132891 | 0.77074 | <.0001 | 10.31 | 33.46 | 153 |
| BHLHE23_D | 0.80429 | <.0001 | 6.09 | 94.32 | 24 |
| ALOX5 | 0.60951 | 0.0006 | 8.79 | 8.24 | 9 |
| MAX.chr19.46379903-46380197 | 0.72693 | <.0001 | 7.50 | 18.54 | 179 |
| ODC1 | 0.50326 | 0.0009 | 5.35 | 11.28 | 231 |
| CHST2_B | 0.59226 | 0.0003 | 6.21 | 115.56 | 57 |
| MAX.chr5.77268672-77268725 | 0.87698 | <.0001 | 11.21 | 43.30 | 199 |
| C17orf64 | 0.64678 | <.0001 | 19.80 | 21.95 | 29 |
| EMX1_A | 0.81454 | <.0001 | 8.05 | 61.16 | 74 |
| CHST2_A | 0.4986 | 0.0014 | 4.33 | 52.73 | 56 |
| DSCR6 | 0.83504 | <.0001 | 8.99 | 92.33 | 72 |
| ITPRIPL1 | 0.75676 | <.0001 | 11.13 | 26.84 | 134 |
| IGF2BP3_B | 0.62302 | 0.0099 | 11.34 | 28.74 | 124 |
| CDH4_E | 0.71016 | <.0001 | 4.40 | 8.46 | 53 |
| NACAD | 0.77213 | <.0001 | 4.51 | 40.68 | 223 |
| DLX4 | 0.94874 | <.0001 | 26.56 | 11.40 | 66 |
| ABLIM1 | 0.7931 | <.0001 | 6.12 | 309.82 | 3 |
| BHLHE23_C | 0.71109 | <.0001 | 7.21 | 64.86 | 23 |
| MAST1 | 0.66356 | <.0001 | 11.84 | 39.32 | 160 |
| ZSCAN12 | 0.7754 | <.0001 | 7.32 | 130.97 | 327 |
| SLC30A10 | 0.64585 | 0.003 | 4.45 | 28.19 | 279 |
| GRASP | 0.75862 | <.0001 | 7.85 | 48.03 | 105 |
| C10orf125 | 0.59972 | 0.0007 | 7.38 | 6.46 | 27 |

**Table 16B.**

| Gene Annotation | HER2+ AUC | p-value | %meth | FC | DMR No. |
|---|---|---|---|---|---|
| ATP6V1B1 | 0.90143 | <.0001 | 29.30 | 3.47 | 15 |
| FOXP4 | 0.55008 | 0.1059 | 43.50 | 1.26 | 96 |
| LMX1B_A | 0.86725 | <.0001 | 26.35 | 3.33 | 149 |

| | | | | | |
|---|---|---|---|---|---|
| BANK1 | 0.80843 | <.0001 | 29.09 | 2.29 | 17 |
| OTX1 | 0.84261 | <.0001 | 27.67 | 3.76 | 237 |
| ST8SIA4 | 0.63275 | 0.0135 | 18.88 | 1.54 | 285 |
| MAX.chr11.14926602-14927148 | 0.94913 | <.0001 | 20.82 | 37.28 | 168 |
| UBTF | 0.82989 | <.0001 | 45.57 | 3.32 | 310 |
| STX16_B | 0.62401 | 0.0042 | 36.51 | 2.45 | 289 |
| KLHDC7B | 0.61447 | 0.0227 | 28.31 | 1.50 | 146 |
| PRKCB | 0.9221 | <.0001 | 17.64 | 39.07 | 256 |
| TBX1 | 0.36486 | 0.8886 | 14.01 | 1.04 | 295 |
| TRH_A | 0.95946 | <.0001 | 32.64 | 12.38 | 303 |
| MPZ | 0.95588 | <.0001 | 26.60 | 92.35 | 221 |
| GP5 | 0.86169 | <.0001 | 40.00 | 5.43 | 104 |
| DNM3_A | 0.96502 | <.0001 | 29.87 | 36.79 | 69 |
| MAX.chr17.73073682-73073814 | 0.42687 | 0.7884 | 17.65 | 1.06 | 174 |
| TRIM67 | 0.91335 | <.0001 | 10.00 | 38.35 | 305 |
| PLXNC1_A | 0.86963 | <.0001 | 11.11 | 17.25 | 245 |
| MAX.chr12.4273906-4274012 | 0.84976 | <.0001 | 10.53 | 48.55 | 170 |
| CALN1_A | 0.87917 | <.0001 | 12.21 | 29.38 | 36 |
| ITPRIPL1 | 0.95866 | <.0001 | 23.28 | 50.38 | 134 |
| MAX.chr12.4273906-4274012 | 0.8752 | <.0001 | 5.88 | 173.48 | 170 |
| GYPC_B | 0.98569 | <.0001 | 20.00 | 20.57 | 109 |
| MAX.chr5.42994866-42994936 | 0.94436 | <.0001 | 12.87 | 18.76 | 198 |
| OSR2_A | 0.8283 | <.0001 | 22.48 | 53.78 | 234 |
| SCRT2 | 0.88116 | <.0001 | 7.61 | 58.12 | 273 |
| MAX.chr5.145725410-145725459 | 0.96343 | <.0001 | 13.01 | 67.01 | 193 |
| MAX.chr11.68622869-68622968 | 0.95151 | <.0001 | 21.77 | 25.18 | 169 |
| MAX.chr8.124173030-124173395 | 0.81558 | <.0001 | 22.14 | 3.10 | 208 |
| CXCL12 | 0.60413 | 0.0012 | 34.62 | 5.57 | 63 |
| MAX.chr20.1784209-1784461 | 0.96105 | <.0001 | 13.80 | 42.33 | 185 |

| | | | | | |
|---|---|---|---|---|---|
| LOC100132891 | 0.91773 | <.0001 | 22.35 | 72.51 | 153 |
| BHLHE23_D | 0.84499 | <.0001 | 5.47 | 84.74 | 24 |
| ALOX5 | 0.89587 | <.0001 | 20.49 | 19.21 | 9 |
| MAX.chr19.46379903-46380197 | 0.88394 | <.0001 | 15.20 | 37.60 | 179 |
| ODC1 | 0.86248 | <.0001 | 10.09 | 21.29 | 231 |
| CHST2_B | 0.92806 | <.0001 | 15.73 | 292.49 | 57 |
| MAX.chr5.77268672-77268725 | 0.93084 | <.0001 | 12.93 | 49.95 | 199 |
| C17orf64 | 0.95469 | <.0001 | 32.32 | 35.82 | 29 |
| EMX1_A | 0.88474 | <.0001 | 11.68 | 88.73 | 74 |
| CHST2_A | 0.83863 | <.0001 | 8.85 | 107.82 | 56 |
| DSCR6 | 0.94118 | <.0001 | 7.61 | 78.17 | 72 |
| ITPRIPL1 | 0.9531 | <.0001 | 23.40 | 56.42 | 134 |
| IGF2BP3_B | 0.86606 | <.0001 | 27.94 | 70.85 | 124 |
| CDH4_E | 0.77424 | <.0001 | 6.06 | 11.65 | 53 |
| NACAD | 0.78219 | <.0001 | 4.02 | 36.27 | 223 |
| DLX4 | 0.83227 | <.0001 | 32.28 | 13.86 | 66 |
| ABLIM1 | 0.83148 | <.0001 | 6.60 | 333.82 | 3 |
| BHLHE23_C | 0.84579 | <.0001 | 7.29 | 65.61 | 23 |
| MAST1 | 0.79571 | <.0001 | 14.23 | 47.24 | 160 |
| ZSCAN12 | 0.86248 | <.0001 | 8.11 | 145.16 | 327 |
| SLC30A10 | 0.79849 | <.0001 | 9.87 | 62.56 | 279 |
| GRASP | 0.8124 | <.0001 | 9.75 | 59.63 | 105 |
| C10orf125 | 0.82273 | <.0001 | 10.54 | 9.22 | 27 |

**Table 16C.**

| Gene Annotation | Luminal A AUC | p-value | %meth | FC | DMR No. |
|---|---|---|---|---|---|
| ATP6V1B1 | 0.86937 | <.0001 | 22.81 | 2.70 | 15 |
| FOXP4 | 0.72222 | 0.0004 | 47.80 | 1.39 | 96 |
| LMX1B_A | 0.89489 | <.0001 | 26.33 | 3.32 | 149 |
| BANK1 | 0.88363 | <.0001 | 30.59 | 2.41 | 17 |
| OTX1 | 0.85736 | <.0001 | 28.06 | 3.81 | 237 |
| ST8SIA4 | 0.82583 | <.0001 | 29.43 | 2.40 | 285 |
| MAX.chr11.14926602-14927148 | 0.84159 | <.0001 | 9.80 | 17.54 | 168 |
| UBTF | 0.89865 | <.0001 | 45.16 | 3.29 | 310 |

| | | | | | |
|---|---|---|---|---|---|
| STX16_B | 0.72523 | <.0001 | 42.32 | 2.84 | 289 |
| KLHDC7B | 0.76426 | <.0001 | 33.45 | 1.77 | 146 |
| PRKCB | 0.9542 | <.0001 | 24.47 | 54.19 | 256 |
| TBX1 | 0.72297 | 0.0688 | 9.74 | 0.72 | 295 |
| TRH_A | 0.93168 | <.0001 | 27.00 | 10.24 | 303 |
| MPZ | 0.87725 | <.0001 | 9.40 | 32.62 | 221 |
| GP5 | 0.72823 | <.0001 | 20.40 | 2.77 | 104 |
| DNM3_A | 0.97748 | <.0001 | 31.67 | 39.00 | 69 |
| MAX.chr17.73073682-73073814 | 0.52177 | 0.3713 | 19.47 | 1.17 | 174 |
| TRIM67 | 0.93619 | <.0001 | 10.08 | 38.66 | 305 |
| PLXNC1_A | 0.81081 | <.0001 | 14.22 | 22.09 | 245 |
| MAX.chr12.4273906-4274012 | 0.91216 | <.0001 | 11.65 | 53.72 | 170 |
| CALN1_A | 0.87012 | <.0001 | 9.36 | 22.52 | 36 |
| ITPRIPL1 | 0.90841 | <.0001 | 12.19 | 26.37 | 134 |
| MAX.chr12.4273906-4274012 | 0.94482 | <.0001 | 5.19 | 153.20 | 170 |
| GYPC_B | 0.91742 | <.0001 | 16.59 | 17.06 | 109 |
| MAX.chr5.42994866-42994936 | 0.83859 | <.0001 | 7.69 | 11.21 | 198 |
| OSR2_A | 0.82995 | <.0001 | 13.90 | 33.26 | 234 |
| SCRT2 | 0.80143 | <.0001 | 4.15 | 31.68 | 273 |
| MAX.chr5.145725410-145725459 | 0.91066 | <.0001 | 7.10 | 36.56 | 193 |
| MAX.chr11.68622869-68622968 | 0.94219 | <.0001 | 14.20 | 16.42 | 169 |
| MAX.chr8.124173030-124173395 | 0.88589 | <.0001 | 19.02 | 2.66 | 208 |
| CXCL12 | 0.76201 | <.0001 | 46.83 | 7.54 | 63 |
| MAX.chr20.1784209-1784461 | 0.89189 | <.0001 | 9.92 | 30.44 | 185 |
| LOC100132891 | 0.93956 | <.0001 | 15.50 | 50.30 | 153 |
| BHLHE23_D | 0.82808 | <.0001 | 4.39 | 67.94 | 24 |
| ALOX5 | 0.83033 | <.0001 | 14.53 | 13.62 | 9 |
| MAX.chr19.46379903-46380197 | 0.83408 | <.0001 | 11.15 | 27.58 | 179 |
| ODC1 | 0.91967 | <.0001 | 9.20 | 19.40 | 231 |

| | | | | | |
|---|---|---|---|---|---|
| CHST2_B | 0.94557 | <.0001 | 15.06 | 280.03 | 57 |
| MAX.chr5.77268672-77268725 | 0.86186 | <.0001 | 9.93 | 38.36 | 199 |
| C17orf64 | 0.95495 | <.0001 | 27.43 | 30.41 | 29 |
| EMX1_A | 0.91366 | <.0001 | 9.89 | 75.12 | 74 |
| CHST2_A | 0.95646 | <.0001 | 11.32 | 137.95 | 56 |
| DSCR6 | 0.77928 | <.0001 | 3.50 | 35.98 | 72 |
| ITPRIPL1 | 0.89414 | <.0001 | 7.98 | 19.25 | 134 |
| IGF2BP3_B | 0.9223 | <.0001 | 27.68 | 70.19 | 124 |
| CDH4_E | 0.81757 | <.0001 | 7.40 | 14.23 | 53 |
| NACAD | 0.70833 | 0.0006 | 2.88 | 25.92 | 223 |
| DLX4 | 0.76877 | <.0001 | 9.95 | 4.27 | 66 |
| ABLIM1 | 0.8217 | <.0001 | 3.92 | 198.44 | 3 |
| BHLHE23_C | 0.79992 | <.0001 | 5.53 | 49.76 | 23 |
| MAST1 | 0.71096 | 0.0006 | 5.15 | 17.10 | 160 |
| ZSCAN12 | 0.67042 | 0.0114 | 2.65 | 47.50 | 327 |
| SLC30A10 | 0.90053 | <.0001 | 11.36 | 71.97 | 279 |
| GRASP | 0.72598 | <.0001 | 4.84 | 29.64 | 105 |
| C10orf125 | 0.88964 | <.0001 | 18.44 | 16.13 | 27 |

**Table 16D.**

| Gene Annotation | Luminal B AUC | p-value | %meth | FC | DMR No. |
|---|---|---|---|---|---|
| ATP6V1B1 | 0.85838 | <.0001 | 27.93 | 3.31 | 15 |
| FOXP4 | 0.59676 | 0.023 | 48.67 | 1.41 | 96 |
| LMX1B_A | 0.90811 | <.0001 | 27.31 | 3.45 | 149 |
| BANK1 | 0.80865 | <.0001 | 31.07 | 2.45 | 17 |
| OTX1 | 0.89838 | <.0001 | 32.49 | 4.42 | 237 |
| ST8SIA4 | 0.62811 | 0.0286 | 19.04 | 1.55 | 285 |
| MAX.chr11.14926602-14927148 | 0.99351 | <.0001 | 21.39 | 38.30 | 168 |
| UBTF | 0.87784 | <.0001 | 52.44 | 3.82 | 310 |
| STX16_B | 0.71892 | 0.0002 | 39.48 | 2.65 | 289 |
| KLHDC7B | 0.72432 | 0.001 | 34.42 | 1.82 | 146 |
| PRKCB | 0.91243 | <.0001 | 20.33 | 45.01 | 256 |
| TBX1 | 0.38054 | 0.372 | 18.89 | 1.40 | 295 |
| TRH_A | 0.92649 | <.0001 | 31.24 | 11.85 | 303 |

| | | | | | |
|---|---|---|---|---|---|
| MPZ | 0.95189 | <.0001 | 18.90 | 65.63 | 221 |
| GP5 | 0.77189 | <.0001 | 35.26 | 4.78 | 104 |
| DNM3_A | 0.89514 | <.0001 | 25.61 | 31.54 | 69 |
| MAX.chr17.73073682-73073814 | 0.58595 | 0.0507 | 25.18 | 1.51 | 174 |
| TRIM67 | 0.92 | <.0001 | 12.08 | 46.32 | 305 |
| PLXNC1_A | 0.80973 | <.0001 | 8.38 | 13.02 | 245 |
| MAX.chr12.4273906-4274012 | 0.89622 | <.0001 | 12.62 | 58.22 | 170 |
| CALN1_A | 0.80541 | <.0001 | 10.14 | 24.39 | 36 |
| ITPRIPL1 | 0.95135 | <.0001 | 21.99 | 47.60 | 134 |
| MAX.chr12.4273906-4274012 | 0.87135 | <.0001 | 5.91 | 174.39 | 170 |
| GYPC_B | 0.8973 | <.0001 | 15.82 | 16.27 | 109 |
| MAX.chr5.42994866-42994936 | 0.92973 | <.0001 | 11.46 | 16.71 | 198 |
| OSR2_A | 0.92216 | <.0001 | 24.46 | 58.54 | 234 |
| SCRT2 | 0.82216 | <.0001 | 10.65 | 81.27 | 273 |
| MAX.chr5.145725410-145725459 | 0.89351 | <.0001 | 12.41 | 63.94 | 193 |
| MAX.chr11.68622869-68622968 | 0.93297 | <.0001 | 39.01 | 45.12 | 169 |
| MAX.chr8.124173030-124173395 | 0.9373 | <.0001 | 27.88 | 3.91 | 208 |
| CXCL12 | 0.53405 | 0.0003 | 64.30 | 10.35 | 63 |
| MAX.chr20.1784209-1784461 | 0.87784 | <.0001 | 17.84 | 54.72 | 185 |
| LOC100132891 | 0.92541 | <.0001 | 34.07 | 110.53 | 153 |
| BHLHE23_D | 0.8 | <.0001 | 6.95 | 107.58 | 24 |
| ALOX5 | 0.84432 | <.0001 | 20.19 | 18.93 | 9 |
| MAX.chr19.46379903-46380197 | 0.94054 | <.0001 | 18.28 | 45.21 | 179 |
| ODC1 | 0.68973 | <.0001 | 5.35 | 11.29 | 231 |
| CHST2_B | 0.86324 | <.0001 | 10.01 | 186.17 | 57 |
| MAX.chr5.77268672-77268725 | 0.96541 | <.0001 | 15.27 | 58.97 | 199 |
| C17orf64 | 0.90595 | <.0001 | 32.60 | 36.14 | 29 |
| EMX1_A | 0.8973 | <.0001 | 15.12 | 114.84 | 74 |

| | | | | | |
|---|---|---|---|---|---|
| CHST2_A | 0.72865 | <.0001 | 6.30 | 76.74 | 56 |
| DSCR6 | 0.92432 | <.0001 | 9.59 | 98.53 | 72 |
| ITPRIPL1 | 0.91135 | <.0001 | 19.45 | 46.90 | 134 |
| IGF2BP3_B | 0.82973 | <.0001 | 45.42 | 115.16 | 124 |
| CDH4_E | 0.81838 | <.0001 | 9.05 | 17.39 | 53 |
| NACAD | 0.75081 | 0.0002 | 6.43 | 57.97 | 223 |
| DLX4 | 0.94595 | <.0001 | 21.61 | 9.28 | 66 |
| ABLIM1 | 0.88541 | <.0001 | 4.31 | 217.96 | 3 |
| BHLHE23_C | 0.8 | <.0001 | 10.48 | 94.28 | 23 |
| MAST1 | 0.77622 | <.0001 | 7.76 | 25.77 | 160 |
| ZSCAN12 | 0.72054 | 0.0002 | 15.29 | 273.71 | 327 |
| SLC30A10 | 0.74595 | <.0001 | 8.48 | 53.74 | 279 |
| GRASP | 0.79459 | <.0001 | 5.88 | 35.98 | 105 |
| C10orf125 | 0.50757 | 0.0006 | 6.91 | 6.04 | 27 |

**Table 16E.**

| Gene Annotation | Overall AUC | p-value | %meth | FC | DMR No. |
|---|---|---|---|---|---|
| ATP6V1B1 | 0.88731 | <.0001 | 26.75 | 3.17 | 15 |
| FOXP4 | 0.62969 | 0.0032 | 47.95 | 1.39 | 96 |
| LMX1B_A | 0.86181 | <.0001 | 26.52 | 3.35 | 149 |
| BANK1 | 0.81125 | <.0001 | 28.59 | 2.25 | 17 |
| OTX1 | 0.84786 | <.0001 | 28.23 | 3.84 | 237 |
| ST8SIA4 | 0.61072 | 0.0054 | 19.51 | 1.59 | 285 |
| MAX.chr11.14926602-14927148 | 0.93745 | <.0001 | 18.72 | 33.52 | 168 |
| UBTF | 0.81517 | <.0001 | 42.18 | 3.07 | 310 |
| STX16_B | 0.66565 | <.0001 | 38.93 | 2.61 | 289 |
| KLHDC7B | 0.67241 | 0.0005 | 29.94 | 1.58 | 146 |
| PRKCB | 0.92153 | <.0001 | 19.52 | 43.21 | 256 |
| TBX1 | 0.36127 | 0.9266 | 13.81 | 1.02 | 295 |
| TRH_A | 0.94355 | <.0001 | 29.05 | 11.02 | 303 |
| MPZ | 0.93396 | <.0001 | 18.93 | 65.72 | 221 |
| GP5 | 0.79294 | <.0001 | 30.09 | 4.08 | 104 |
| DNM3_A | 0.85418 | <.0001 | 24.75 | 30.48 | 69 |
| MAX.chr17.73073682-73073814 | 0.53095 | 0.1372 | 21.71 | 1.31 | 174 |

| | | | | | |
|---|---|---|---|---|---|
| TRIM67 | 0.91391 | <.0001 | 10.80 | 41.41 | 305 |
| PLXNC1_A | 0.76983 | <.0001 | 10.49 | 16.29 | 245 |
| MAX.chr12.4273906-4274012 | 0.9017 | <.0001 | 12.09 | 55.76 | 170 |
| CALN1_A | 0.87271 | <.0001 | 11.47 | 27.59 | 36 |
| ITPRIPL1 | 0.88928 | <.0001 | 17.19 | 37.21 | 134 |
| MAX.chr12.4273906-4274012 | 0.9029 | <.0001 | 6.69 | 197.30 | 170 |
| GYPC_B | 0.87925 | <.0001 | 15.78 | 16.22 | 109 |
| MAX.chr5.42994866-42994936 | 0.8932 | <.0001 | 11.11 | 16.19 | 198 |
| OSR2_A | 0.80667 | <.0001 | 18.56 | 44.40 | 234 |
| SCRT2 | 0.841 | <.0001 | 7.84 | 59.82 | 273 |
| MAX.chr5.145725410-145725459 | 0.91303 | <.0001 | 10.12 | 52.12 | 193 |
| MAX.chr11.68622869-68622968 | 0.87947 | <.0001 | 20.07 | 23.21 | 169 |
| MAX.chr8.124173030-124173395 | 0.85636 | <.0001 | 21.94 | 3.08 | 208 |
| CXCL12 | 0.60615 | <.0001 | 41.39 | 6.66 | 63 |
| MAX.chr20.1784209-1784461 | 0.85113 | <.0001 | 12.03 | 36.91 | 185 |
| LOC100132891 | 0.89124 | <.0001 | 19.91 | 64.60 | 153 |
| BHLHE23_D | 0.82149 | <.0001 | 5.60 | 86.71 | 24 |
| ALOX5 | 0.79948 | <.0001 | 15.96 | 14.97 | 9 |
| MAX.chr19.46379903-46380197 | 0.84416 | <.0001 | 12.85 | 31.77 | 179 |
| ODC1 | 0.76024 | <.0001 | 7.77 | 16.38 | 231 |
| CHST2_B | 0.84154 | <.0001 | 12.15 | 226.06 | 57 |
| MAX.chr5.77268672-77268725 | 0.90519 | <.0001 | 12.13 | 46.85 | 199 |
| C17orf64 | 0.87293 | <.0001 | 28.03 | 31.07 | 29 |
| EMX1_A | 0.88056 | <.0001 | 11.01 | 83.60 | 74 |
| CHST2_A | 0.77114 | <.0001 | 8.00 | 97.42 | 56 |
| DSCR6 | 0.86595 | <.0001 | 7.14 | 73.34 | 72 |
| ITPRIPL1 | 0.88165 | <.0001 | 15.26 | 36.79 | 134 |
| IGF2BP3_B | 0.81822 | <.0001 | 27.51 | 69.74 | 124 |
| CDH4_E | 0.78073 | <.0001 | 6.67 | 12.81 | 53 |

| | | | | | |
|---|---|---|---|---|---|
| NACAD | 0.75207 | <.0001 | 4.29 | 38.67 | 223 |
| DLX4 | 0.86399 | <.0001 | 22.31 | 9.58 | 66 |
| ABLIM1 | 0.83054 | <.0001 | 5.25 | 265.54 | 3 |
| BHLHE23_C | 0.79174 | <.0001 | 7.40 | 66.61 | 23 |
| MAST1 | 0.73627 | <.0001 | 9.73 | 32.31 | 160 |
| ZSCAN12 | 0.75774 | <.0001 | 7.79 | 139.40 | 327 |
| SLC30A10 | 0.78182 | <.0001 | 8.75 | 55.48 | 279 |
| GRASP | 0.77114 | <.0001 | 7.10 | 43.44 | 105 |
| C10orf125 | 0.72646 | <.0001 | 11.36 | 9.94 | 27 |

**Table 17.** Table 17 highlights the top 10 MDMs for discriminating DCIS HGD from DCIS LGD.

| Gene Annotation | AUC | p-value | DMR No. |
|---|---|---|---|
| DSCR6 | 0.9127 | <.0001 | 72 |
| SCRT2 | 0.86905 | 0.0314 | 273 |
| MPZ | 0.85714 | 0.0275 | 221 |
| MAX.chr8.124173030-124173395 | 0.84127 | 0.0122 | 208 |
| OSR2_A | 0.84127 | 0.0067 | 234 |
| MAX.chr11.68622869-68622968 | 0.82143 | 0.0067 | 169 |
| ITPRIPL1 | 0.81746 | 0.0851 | 134 |
| MAX.chr5.145725410-145725459 | 0.81349 | 0.0037 | 193 |
| BHLHE23_C | 0.80952 | 0.004 | 23 |
| ITPRIPL1 | 0.80556 | 0.0658 | 134 |

**Example III.**

This example describes identification of breast tissue markers and plasma markers for detecting breast cancer.

Candidate methylation markers for the detection of breast cancer were identified by RRBS of breast cancer and normal breast tissue samples. Originally 58 markers were identified and target enrichment long-probe quantitative amplified signal assays were designed and ordered (see, e.g., WO2017/075061 and U.S. Patent Application Serial No. 15,841,006 for general techniques) (Table 18 shows the methylated regions distinguishing breast cancer tissue from normal breast tissue) (Tables 19 and 20 show the primer and probe sequences for the markers shown in Table 18). After design screening and redesign, 56 markers (see, Table 21) were chosen and assays made, triplexed and tested on tissue. Assays

were equally split between FAM and HEX reporting and triplexed with the reference assay, B3GALT6 which reports to Quasar670.

**Table 18. Methylated regions distinguishing breast cancer tissue from normal breast tissue**

| DMR No. | Gene Annotation | Region on Chromosome (starting base-ending base) |
|---|---|---|
| 329 | ABLIM1_B | chr10:116391634-116391781 |
| 330 | AJAP1_C | chr1:4715533-4715652 |
| 331 | ALOX5_B | chr10:45914740-45914889 |
| 332 | ASCL2_B | chr11:2292232-2292371 |
| 333 | BANK1_B | chr4:102711861-102712082 |
| 334 | BHLHE23_E | chr20:61638334-61638574 |
| 335 | C10orf125_B | chr10:135171404-135171514 |
| 336 | C17orf64_B | chr17:58499085-58499196 |
| 337 | CALN1_1520 | chr7:71801485-71801604 |
| 37 | CALN1_B | Chr7:71801741-71801800 |
| 339 | CD1D_1058 | chr1:158150861-158151139 |
| 340 | CDH4_7890 | chr20:59827763-59828158 |
| 341 | CHST2_8128 | chr3:142838015-142838501 |
| 342 | CHST2_8384 | chr3:142838015-142838501 |
| 343 | CHST2_9316 | chr3:142839218-142839575 |
| 344 | CHST2_9470 | chr3:142839218-142839575 |
| 345 | CLIC6_B | chr21:36042020-36042140 |
| 346 | CXCL12_B | chr10:44881200-44881315 |
| 347 | DLX4_B | chr17:48042552-48042616 |
| 348 | DNM3_D | chr1:171810425-171810575 |
| 74 | EMX1_A | chr2:73151498-73151578 |
| 349 | ESPN_B | chr1:6507924-6508087 |
| 350 | FAM59B_7764 | chr2:26407703-26407976 |
| 351 | FOXP4_B | chr6:41528816-41528912 |
| 104 | GP5 | chr3:194118738-194118924 |
| 352 | HOXA1_C | chr7:27135593-27135895 |
| 353 | IGF2BP3_C | chr7: 23513861-23514064 |
| 354 | IPTRIPL1_1138 | chr2:96990958-96991338 |
| 355 | IPTRIPL1_1200 | chr2:96990958-96991338 |

| | | |
|---|---|---|
| 356 | KCNK9_B | chr8:140715096-140715177 |
| 357 | KCNK17_C | chr6:39281887-39281994 |
| 358 | KLHDC7B_B | chr22:50987209-50987311 |
| 359 | LAYN_B | chr11:111412023-111412090 |
| 360 | LIME1_B | chr20:62369173-62369342 |
| 361 | LMX1B_D | chr9:129388170-129388223 |
| 362 | LOC100132891_B | chr8:72755986-72756299 |
| 375 | MAST1_B | chr19:12978496-12978642 |
| 338 | MAX.chr12.427.br | chr12:4273906-4274012 |
| 174 | MAX.chr17.73073682-73073814 | chr17:73073682-73073814 |
| 363 | MAX.chr20.4422 | chr20:1784207-1784471 |
| 364 | MPZ_5742 | chr1:161275554-161276006 |
| 365 | MPZ_5554 | chr1:161275554-161276006 |
| 366 | MSX2P1_B | chr17:56234426-56234520 |
| 367 | ODC1_B | chr2:10589075-10589225 |
| 234 | OSR2_A | chr8:99952233-99952366 |
| 368 | OTX1_B | chr2:63281460-63281599 |
| 246 | PLXNC1_B | chr12:94544333-94544426 |
| 369 | PRKCB_7570 | chr16:23847569-23847705 |
| 370 | SCRT2_C | chr20:644563-644631 |
| 279 | SLC30A10 | chr1:220101458-220101634 |
| 371 | SPHK2_B | chr19:49127571-49127685 |
| 372 | ST8SIA4_B | chr5:100240049-100240286 |
| 373 | STX16_C | chr20:57225077-57225237 |
| 374 | TBX1_B | chr22:19754226-19754419 |
| 303 | TRH_A | chr3:129693484-129693575 |
| 328 | TRIM67_B | chr1:231297039-231297163 |

Table 19.

| Gene Annotation | DMR No. | Forward Primer 5'-3' | SEQ ID NO: | Reverse Primer 5'-3' | SEQ ID NO: |
|---|---|---|---|---|---|
| ABLIM1_B | 329 | TGGTAATCGGGTTTTTCGACG | 255 | CCGCGAATCTATCTACCGAAAC | 256 |
| AJAP1_C | 330 | GTGTTAGGTTGGGCGGAAG | 257 | GTTACCCGCTTACGAAAAACGA | 258 |
| ALOX5_B | 331 | TTCGTTTTTTGTCGGGAGTTATTC | 259 | TCCAAAAATTAAATTAAAAACGCTACGC | 260 |
| ASCL2_B | 332 | ATAATACGGTTGTTCGGGAGG | 261 | GTAAATATAAACTACGCGACGCGTA | 262 |

| | | | | | |
|---|---|---|---|---|---|
| BANK1_B | 333 | GAGAGTTTAGGTAGCGTTCGG | 263 | CCTAACGCTACTAACAACATTATAACGA | 264 |
| BHLHE23_E | 334 | CGCGGTTTTGGAGCGTTAG | 265 | CCGAAACGACCGAAAACGAC | 266 |
| C10orf125_B | 335 | CGGTTCGTTGCGTTTATCGA | 267 | CCCCCGAACTACTCTACGCG | 268 |
| C17orf64_B | 336 | GATTATATTCGGATTTTGTTTATCGCGT | 269 | GACTCTTCCTACCCGCGA | 270 |
| CALN1_1520 | 337 | GCGGTTTTTAGTTCGCGGG | 271 | AACAAATAATTAACAAACAACGCCTCC | 272 |
| CALN1_B | 37 | TCGTTCGGCGTATTTATTTCGTAT | 273 | CGCGAAAAACTTCCTCCGA | 274 |
| CD1D_1058 | 339 | GGATTGGTGAGATTCGGGAC | 275 | CCCGAAACCAAAAAACAACGA | 276 |
| CDH4_7890 | 340 | CGGGGAGTTTCGTTTGTATCG | 277 | CGAATAACGACTACGAACTTTAAACG | 278 |
| CHST2_8128 | 341 | CGTAGTTATAGATTTATTAGAGAGGGCG | 279 | CTAAAACGATAAAAAAACGCGAAACG | 280 |
| CHST2_8384 | 342 | TGGTAGTTTTCGGTATCGACGAG | 281 | TAACTCTACGCGCAAAACGC | 282 |
| CHST2_9316 | 343 | GGGATTTTTAGCGGAAGCGA | 283 | CGACGAACTATCCGACTATCACT | 284 |
| CHST2_9470 | 344 | CGGAGGAATCGGGTAGAATCG | 285 | ACTCTCCCATAACAACGACTCC | 286 |
| CLIC6_B | 345 | CGCGTAGGGCGAGTTTC | 287 | GCCTCCTCCTACCTCTCG | 288 |
| CXCL12_B | 346 | TCGGCGGTTTTTAGTAAAAGCG | 289 | AAATCTCCCGTCCCACTCC | 290 |
| DLX4_B | 347 | GGTATATTCGCGTAGGTGCG | 291 | AACCGAATACCGAAATCTATAACCC | 292 |
| DNM3_D | 348 | GTAGTTGGGTTGTAGTGCGTG | 293 | CCCGAACTTCCCATCGAAC | 294 |
| EMX1_A | 74 | TTCGTACGGTTTTTTCGTTTTCG | 295 | CCACCACGTAATAATTCTTCTCGAAA | 296 |
| ESPN_B | 349 | CGGTTTGATATTATTCGGGGTTCG | 297 | AATTAACGCCCCCTATAACATCC | 298 |
| FAM59B_7764 | 350 | CGCGATAGCGTTTTTTATTGTCGCG | 299 | CGCACGACCGTAAAATACTCG | 300 |
| FOXP4_B | 351 | CGGTTCGTAGATTGTTTTAGAGCG | 301 | CAAATACCGTCGAAAAAAAACTAAATCAAAAC | 302 |
| GP5 | 104 | CGTTGTAGGACGGTTATGTCG | 303 | CATCCTACTCTTCGAAATAAACCGC | 304 |
| HOXA1_C | 352 | AGTCGTTTTTTTAGGTAGTTTAGGCG | 305 | CGACCTTTACAATCGCCGC | 306 |
| IGF2BP3_C | 353 | AGATTGGCGCGTAAAAGCG | 307 | ACCGACCCCGAAAAACG | 308 |
| IPTRIPL1_1138 | 354 | CGTTTTCGGAGTCGCGTG | 309 | AACCATACTTATCCGAACGTCTAAAC | 310 |
| IPTRIPL1_1200 | 355 | GAGTAGGGTTATTTTCGCGGG | 311 | CTACTTTTTTCCCGACAAAATAAAAACGT | 312 |
| KCNK9_B | 356 | TTTTCGCGTATTTCGTGGTTC | 313 | AACGCCGCCGTATTCG | 314 |
| KCNK17_C | 357 | TCGCGTTGGAAGTTGCG | 315 | CGTATTCTAAACGCTAAAAAACCGC | 316 |
| KLHDC7B_B | 358 | CGGCGGTAGTTTTGCGG | 317 | CTACTAAACAAAAACCAACACGTCC | 318 |
| LAYN_B | 359 | GGTAGGTTTGTTAGTTGGTTTTCG | 319 | CGCTATCTCTACGACCGCCT | 320 |

| | | | | | |
|---|---|---|---|---|---|
| LIME1_B | 360 | CGGAGGTAGCGGGCGAG | 321 | CACTCACCGCTTCCGCC | 322 |
| LMX1B_D | 361 | GGCGTTCGTTTCGGCG | 323 | CGCTTCTCCGACGCCC | 324 |
| LOC100132891_B | 362 | GCGGTTGAGTTTTTGGTCGG | 325 | CCCCGTATAACTAAAAACGACGAC | 326 |
| MAST1_B | 375 | CGTTTTTTTTATGTAGTAAGCGATTTTTCGC | 327 | AAACGACGACGAACGCC | 328 |
| MAX.chr12.427.br | 338 | GCGTTTTGGTTTTTTCGTTTCGAG | 329 | GAACGACGAAACTAAAACCGC | 330 |
| MAX.chr17.73073682-73073814 | 174 | CGTTTTTTGGTAGTTTTTTTCGAGTCG | 331 | GCTTAAACGTAACCGAAACGCC | 332 |
| MAX.chr20.4422 | 363 | GGTTGCGCGTCGTTTTTTC | 333 | CCCGACGCGTTTAAATCGT | 334 |
| MPZ_5742 | 364 | GGATGGGAATAGTTAAGTTTTAGTCGTT | 335 | TCCAACATTACATACAACACTAACGTC | 336 |
| MPZ_5554 | 365 | GGTTAGGGGTGGAGTTCGTTA | 337 | ACTCCGAACTCTACTCATCCTTTC | 338 |
| MSX2P1_B | 366 | TAGGTTGGAGATTTTGACGCG | 339 | CGAAACCTAAAAACGCCGAAAC | 340 |
| ODC1_B | 367 | GGTTGGTAGTCGTTTTTACGTTTTC | 341 | CAAAACCCATCTAATTACAAAATACCTCGA | 342 |
| OSR2_A | 234 | TGGAGTTATCGGAAGGCGA | 343 | CGAACTCCCGAAACGACG | 344 |
| OTX1_B | 368 | GGAAATGGTTTAGAGTTTTGGATTTCG | 345 | TTCTAAAAAATACTTTCGATACCGACA | 346 |
| PLXNC1_B | 246 | GTGGTTTGAAGAGTTGTTAGTTCGTTTAG | 347 | GCCAAAAATTCGATTCCAACGCA | 348 |
| PRKCB_7570 | 369 | AAGGTGGGTTGTTTGAAGAAGC | 349 | ACCCTCCGACAAAAAAACGTAC | 350 |
| SCRT2_C | 370 | GCGAGAAGGTTTTGTCGTAGA | 351 | ACCTACTCACGCACAACCT | 352 |
| SLC30A10 | 279 | CGCGGTGAGGAAGATCG | 353 | ACGCCACCTACGACTACG | 354 |
| SPHK2_B | 371 | GTACGGTTATTGGTTGAGCGG | 355 | CCGAATCCTCCTCCAAACG | 356 |
| ST8SIA4_B | 372 | GGAATTTAATTGGAGAGAAATTTTGGCG | 357 | CCAAAATTTCCCTCATCTATATACGCC | 358 |
| STX16_C | 373 | GTTGCGGGTCGGGTTGC | 359 | GCAAAAACAAAAAACGCGTAAAAACC | 360 |
| TBX1_B | 374 | GTCGTCGTTGTCGTAGTTGTC | 361 | CGTAAAAACCGAACGACGCG | 362 |
| TRH_A | 303 | TTTTCGTTGATTTTATTCGAGTCGTC | 363 | GAACCCTCTTCAAATAAACCGC | 364 |
| TRIM67_B | 328 | GATTAAATAGTCGGGGTCGCG | 365 | ATTCTCCAACGCCAACCAC | 366 |

Table 20.

| Gene Annotation | DMR No. | Probe Sequence | SEQ ID NO: |
|---|---|---|---|
| ABLIM1_B | 329 | CGCGCCGAGG CGCGCTTCCACTCC/3C6/ | 367 |
| AJAP1_C | 330 | AGGCCACGGACG GCGGCGTTTTTTTTTATGTTG/3C6/ | 368 |
| ALOX5_B | 331 | AGGCCACGGACG CAACCGAACTAAAAAAAAAAACTAACG/3C6/ | 369 |

| | | | |
|---|---|---|---|
| ASCL2_B | 332 | CGCGCCGAGG GCGCGTAAGATTTTCGG/3C6/ | 370 |
| BANK1_B | 333 | CGCGCCGAGG GCGGGTAGTAGTGCG/3C6/ | 371 |
| BHLHE23_E | 334 | CGCGCCGAGG CGACCGAAAAATCGAAAAACA/3C6/ | 372 |
| C10orf125_B | 335 | CGCGCCGAGG GCTAACGCGAATAAAACACG/3C6/ | 373 |
| C17orf64_B | 336 | CGCGCCGAGG TTTTCGTTTTCGGTTTCGG/3C6/ | 374 |
| CALN1_1520 | 337 | CGCGCCGAGG CCGTACCTATTAACTCCG/3C6/ | 375 |
| CALN1_B | 37 | AGGCCACGGACG TCGTTTTTTTTTTGCGGGT/3C6/ | 376 |
| CD1D_1058 | 339 | AGGCCACGGACG CGTATTGGCGCGATTTAG/3C6/ | 377 |
| CDH4_7890 | 340 | AGGCCACGGACG GTTCGAAAAAAACTCGACGAA /3C6/ | 378 |
| CHST2_8128 | 341 | AGGCCACGGACG GCCGTTCTCTAACTTCCG/3C6/ | 379 |
| CHST2_8384 | 342 | AGGCCACGGACG CCGAAATACGAACGCGA/3C6/ | 380 |
| CHST2_9316 | 343 | AGGCCACGGACG TCGTTCCTCGATTTCGC/3C6/ | 381 |
| CHST2_9470 | 344 | AGGCCACGGACG CGAATAAAACCTACGAAAAAAAACG /3C6/ | 382 |
| CLIC6_B | 345 | AGGCCACGGACG GAAAACCGCAAAATCCTCG/3C6/ | 383 |
| CXCL12_B | 346 | AGGCCACGGACG CGCGAAATAAACCTATAATTAACTCA/3C6/ | 384 |
| DLX4_B | 347 | CGCGCCGAGG CCGAACCAACACTCAAAAC/3C6/ | 385 |
| DNM3_D | 348 | CGCGCCGAGG GCGCGTTTGGTTTGGT/3C6/ | 386 |
| EMX1_A | 74 | AGGCCACGGACG AACGCGCTCCAACC/3C6/ | 387 |
| ESPN_B | 349 | CGCGCCGAGG CGCGACGACTAAAAAAATTCA/3C6/ | 388 |
| FAM59B_7764 | 350 | AGGCCACGGACG GTCGAAATCGAAACGCTC/3C6/ | 389 |
| FOXP4_B | 351 | CGCGCCGAGG CCGCGACTACCTCTTC/3C6/ | 390 |
| GP5 | 104 | AGGCCACGGACG CGACGTCCTACAAAACCA/3C6/ | 391 |
| HOXA1_C | 352 | CGCGCCGAGG GGCGGTAGTTGTTGC/3C6/ | 392 |
| IGF2BP3_C | 353 | CGCGCCGAGG GCGAAAACCCCGCC/3C6/ | 393 |
| IPTRIPL1_1138 | 354 | CGCGCCGAGG CGTCTAACTAAACGCGATAAAC/3C6/ | 394 |
| IPTRIPL1_1200 | 355 | CGCGCCGAGG GCGGTTTTAGCGATGAATC/3C6/ | 395 |
| KCNK9_B | 356 | CGCGCCGAGG CGATTCGAGGGCGT/3C6/ | 396 |
| KCNK17_C | 357 | AGGCCACGGACG CGCGACGCAAAACTC/3C6/ | 397 |
| KLHDC7B_B | 358 | AGGCCACGGACG GCGGCGGTTGGATT/3C6/ | 398 |
| LAYN_B | 359 | AGGCCACGGACG TCCCGAAACGAACGATAAA/3C6/ | 399 |
| LIME1_B | 360 | CGCGCCGAGG CGCCGTCGCACTAC/3C6/ | 400 |
| LMX1B_D | 361 | AGGCCACGGACG CGCGACTCCCCACT/3C6/ | 401 |

| | | | |
|---|---|---|---|
| LOC100132891_B | 362 | AGGCCACGGACG CGCAAATAATAACGCGAACG/3C6/ | 402 |
| MAST1_B | 375 | AGGCCACGGACG CGTTCGAGGTTAGTTTTTTGG/3C6/ | 403 |
| MAX.chr12.427.br | 338 | AGGCCACGGACG CGTACGTAACCCGCG/3C6/ | 404 |
| MAX.chr17.73073682-73073814 | 174 | CGCGCCGAGG CGCTACTAACCATAACCGC/3C6/ | 405 |
| MAX.chr20.4422 | 363 | CGCGCCGAGG CGTTTTCGTTTGATTCGGTT/3C6/ | 406 |
| MPZ_5742 | 364 | CGCGCCGAGG TCGGTGATTGATGTGTGCG/3C6/ | 407 |
| MPZ_5554 | 365 | CGCGCCGAGG CGTAACTCCATCTCGATAACC/3C6/ | 408 |
| MSX2P1_B | 366 | CGCGCCGAGG CGACCGCGAAAAAACG/3C6/ | 409 |
| ODC1_B | 367 | AGGCCACGGACG CGCGTTGGAAGTTTCG/3C6/ | 410 |
| OSR2_A | 234 | CGCGCCGAGG GCGCGAACACAAAACG/3C6/ | 411 |
| OTX1_B | 368 | CGCGCCGAGG ACCGAAAACGCCCTAAA/3C6/ | 412 |
| PLXNC1_B | 246 | CGCGCCGAGG GCGTGGAGAAATGTTAGTTTG/3C6/ | 413 |
| PRKCB_7570 | 369 | AGGCCACGGACG CGGGCGGTGAATTTGT/3C6/ | 414 |
| SCRT2_C | 370 | AGGCCACGGACG ACGTCGTATTTGTGGCG/3C6/ | 415 |
| SLC30A10 | 279 | AGGCCACGGACG GCGTTGTTTAGCGCG/3C6/ | 416 |
| SPHK2_B | 371 | AGGCCACGGACG GATCCCGCAAATCAACAC/3C6/ | 417 |
| ST8SIA4_B | 372 | CGCGCCGAGG CGATCCCCAACTCCC/3C6/ | 418 |
| STX16_C | 373 | CGCGCCGAGG CGCTTCTAAAACCTCGATCC/3C6/ | 419 |
| TBX1_B | 374 | CGCGCCGAGG CGCGGTCGTTAATATGTATTC/3C6/ | 420 |
| TRH_A | 303 | AGGCCACGGACG CGTTTGGCGTAGATATAAGC/3C6/ | 421 |
| TRIM67_B | 328 | AGGCCACGGACG CGAACTACGAAAACAACCTC/3C6/ | 422 |

Table 21

| Marker | DMR | Marker | DMR |
|---|---|---|---|
| AJAP1_C | 330 | CHST2_9316 | 343 |
| C10orf125_B | 335 | ASCL2_B | 332 |
| CALN1_B | 37 | ESPN_B | 349 |
| BHLHE23_E | 334 | DLX4_B | 347 |
| CD1D_1058 | 339 | KCNK17_C | 357 |
| HOXA1_C | 352 | EMX1_A | 74 |
| LOC100132891_B | 362 | MPZ_5742 | 364 |

| | | | |
|---|---|---|---|
| MSX2P1_B | 366 | LAYN_B | 359 |
| PRKCB_7570 | 369 | KCNK9_B | 356 |
| ITPRIPL1_1200 | 355 | ABLIM1_B | 329 |
| SPHK2_B | 371 | MAX.chr12.427.br | 338 |
| C17orf64_B | 336 | SCRT2_C | 370 |
| TRIM67_B | 328 | IGF2BP3_C | 353 |
| MAX.chr20.4422 | 363 | MAST1_B | 375 |
| DNM3_D | 348 | MAX.chr17.73073682-73073814 | 174 |
| ODC1_B | 367 | OTX1_B | 368 |
| OSR2_A | 234 | ST8SIA4_B | 372 |
| SLC30A10 | 279 | CXCL12_B | 346 |
| TRH_A | 303 | LIME1_B | 360 |
| ALOX5_B | 331 | TBX1_B | 374 |
| PLXNC1_B | 246 | STX16_C | 373 |
| CDH4_7890 | 340 | FOXP4_B | 351 |
| CLIC6_B | 345 | CALN1_1520 | 337 |
| LMX1B_D | 361 | ITPRIPL1_1138 | 354 |
| FAM59B_7764 | 350 | CHST2_8128 | 341 |
| GP5 | 104 | CHST2_8384 | 342 |
| BANK1_B | 333 | CHST2_9470 | 344 |
| KLHDC7B_B | 358 | MPZ_5554 | 365 |

A collection of 38 normal breast cancer samples including 6 BRCA carriers and 113 breast cancer tissue samples including Luminal A & B, HER2+, BRCA1+, BRCA2+, triple negative and DCIS varieties was tested for presence of the 56 methylation markers. The 56 markers displayed a range of sensitivities from ~15% to 92% at 95% specificity. Table 22 shows the markers demonstrating sensitivity at or above 25% at 95% specificity. A 5 marker panel (SPHK2, c17orf64_B, DLX4_B, MPZ_5742, ITPRIPL1_1138) showed 96% sensitivity at 100% specificity. The resulting ROC curve had an AUC of 0.995.

**Table 22.**

| Marker | DMR No. | Sensitivity |
|---|---|---|
| AJAP1_C | 330 | 66.30% |
| C10orf125_B | 335 | 58.40% |
| CALN1_B | 37 | 69.70% |
| BHLHE23_E | 334 | 43.80% |

| | | |
|---|---|---|
| CD1D_1058 | 339 | 68.50% |
| HOXA1_C | 352 | 62.90% |
| LOC100132891_B | 362 | 79.80% |
| MSX2P1_B | 366 | 79.80% |
| PRKCB_7570 | 369 | 86.50% |
| ITPRIPL1_1200 | 355 | 79.80% |
| SPHK2_B | 371 | 65.20% |
| C17orf64_B | 336 | 77.50% |
| TRIM67_B | 328 | 79.80% |
| MAX.chr20.4422 | 363 | 71.90% |
| CHST2_9316 | 343 | 73.00% |
| ASCL2_B | 332 | 53.90% |
| ESPN_B | 349 | 67.40% |
| DLX4_B | 347 | 83.10% |
| KCNK17_C | 357 | 55.10% |
| EMX1_A | 74 | 77.50% |
| MPZ_5742 | 364 | 91.00% |
| LAYN_B | 359 | 57.30% |
| KCNK9_B | 356 | 62.90% |
| ABLIM1_B | 329 | 44.90% |
| MAX.chr12.427.br | 338 | 79.80% |
| SCRT2_C | 370 | 78.70% |
| IGF2BP3_C | 353 | 70.80% |
| MAST1_B | 375 | 77.50% |
| DNM3_D | 348 | 74.20% |
| ODC1_B | 367 | 65.20% |
| OSR2_A | 234 | 70.80% |
| SLC30A10 | 279 | 60.70% |
| TRH_A | 303 | 85.40% |
| ALOX5_B | 331 | 59.60% |
| PLXNC1_B | 246 | 61.80% |
| CDH4_7890 | 340 | 71.90% |
| CLIC6_B | 345 | 48.30% |
| LMX1B_D | 361 | 56.20% |

| | | |
|---|---|---|
| FAM59B_7764 | 350 | 66.30% |
| GP5 | 104 | 61.80% |
| BANK1_B | 333 | 43.80% |
| OTX1_B | 368 | 70.80% |
| ST8SIA4_B | 372 | 40.40% |
| CXCL12_B | 346 | 56.20% |
| LIME1_B | 360 | 47.20% |
| STX16_C | 373 | 52.80% |
| FOXP4_B | 351 | 36.00% |
| CALN1_1520 | 337 | 66.30% |
| ITPRIPL1_1138 | 354 | 83.10% |
| CHST2_8128 | 341 | 62.90% |
| CHST2_8384 | 342 | 60.70% |
| CHST2_9470 | 344 | 66.30% |
| MPZ_5554 | 365 | 92.10% |

Based on the results of the tissue testing, a set of 28 markers were selected to test on a set of plasma samples collected from breast cancer patients and normal controls. The 28 markers were split into two pools of 14 due to the high number of markers to be tested. The markers in the two pools are shown in Tables 23 and 24 below.

| Table 23: Pool 7 Breast Cancer Plasma Markers | |
|---|---|
| AJAP1 | C10orf125 |
| CALN1_B | BHLHE23 |
| LOC100132891 | MSX2P1 |
| SPHK2 | C17orf64 |
| MAST1 | DNM3 |
| MAX.chr.12.427.br | OTX1 |
| SCRT2 | ALOX5 |

| Table 24: Pool 8 Breast Cancer Plasma Markers | |
|---|---|
| FAM59B | ITPRIPL1_B |
| ODC1_B | OSR2_A |
| CD1D_B | DLX4_2591 |
| PRKCB_7570 | MAX.chr20.4422 |
| TRIM67 | MPZ |

| | |
|---|---|
| TRH_A | CXCL12_B |
| EMX1_br | CHST2_B |

The testing of Pool 7 markers was done on a collection of EDTA plasma samples comprised of 85 breast cancer samples (33 stage I, 33 stage II, 18 stage III, and 1 stage IV) and 100 healthy normal controls. The testing of Pool 8 markers was done on a similar collection of EDTA plasma samples comprised of 85 breast cancer samples (34 stage I, 32 stage II, 18 stage III and 1 stage IV) and 100 healthy normal controls. Based on the results of the Pool 7 and Pool 8 testing, a collection of 14 assays were selected for further testing (shown in Table 25).

| Table 25: Pool 9 Breast Cancer Plasma Markers | |
|---|---|
| SPHK2 | C17orf64 |
| FAM59B | ITPRIPL1_B |
| ODC1_B | OSR2_A |
| TRIM67 | MPZ |
| TRH_A | CXCL12_B |
| CD1D_B | C10orf125 |
| CALN1_B | CHST2_B |

The testing of Pool 9 markers was done on a collection of LBgard (Biomatrica, San Diego, CA) plasma samples comprised of 42 breast cancer samples (1 stage I, 16 stage II, 14 stage III, and 11 stage IV) and 84 healthy normal controls. Table 26 shows the identified methylated region for the Pool 9 markers. Table 27 shows the exhibited sensitivity and 90% specificity for the Pool 9 markers. Tables 28 and 29 show the primer information, and probe information for the Pool 9 markers. A collection of 4 markers (FAM59B, ITPRIPL1, TRH_A, and C17orf64_B) exhibited a sensitivity of 74% at 90% specificity. The resulting ROC curve exhibited an AUC of 0.884.

Table 26.

| **DMR No.** | **Gene Annotation** | **Region on Chromosome (starting base-ending base)** |
|---|---|---|
| 47 | CD1D | chr1:158150864-158151129 |

| | | |
|---|---|---|
| 134 | ITPRIPL1 | chr2:96990968-96991328 |
| 90 | FAM59B | chr2:26407713-26407972 |
| 27 | C10orf125 | chr10:135171410-135171504 |
| 305 | TRIM67 | chr1:231297047-231297159 |
| 284 | SPHK2 | chr19:49127580-49127683 |
| 37 | CALN1_B | chr7:71801741-71801800 |
| 57 | CHST2_B | chr3:142839223-142839568 |
| 221 | MPZ | chr1:161275561-161275996 |
| 346 | CXCL12_B | chr10:44881200-44881315 |
| 367 | ODC1_B | chr2:10589075-10589225 |
| 234 | OSR2_A | chr8:99952233-99952366 |
| 303 | TRH_A | chr3:129693484-129693575 |
| 336 | C17orf64_B | chr17:58499085-58499196 |

Table 27.

| **Marker Name** | **AUC** | **Sens @ 90% sp** | **DMR No.** |
|---|---|---|---|
| FAM59B | 0.814 | 50.0% | 90 |
| ITPRIPL1 | 0.804 | 61.9% | 134 |
| ODC1_B | 0.809 | 59.5% | 367 |
| OSR2_A | 0.749 | 42.9% | 234 |
| TRIM67 | 0.669 | 30.9% | 305 |
| MPZ | 0.698 | 47.6% | 221 |
| TRH_A | 0.83 | 50.0% | 303 |
| CXCL12_B | 0.71 | 28.6% | 346 |
| SPHK2 | 0.585 | 31.0% | 284 |
| C17orf64_B | 0.763 | 59.5% | 336 |
| CD1D | 0.613 | 33.3% | 47 |
| C10orf125 | 0.775 | 45.2% | 27 |
| CALN1_B | 0.622 | 26.2% | 37 |
| CHST2_B | 0.687 | 38.1% | 57 |

Table 28.

| **Gene Annotation** | **DMR No.** | **Forward Primer 5'-3'** | **SEQ ID NO:** | **Reverse Primer 5'-3'** | **SEQ ID NO:** |
|---|---|---|---|---|---|
| CD1D | 47 | GGATTGGTGA GATTCGGGAC | 423 | CCCGAAACCAAA AAACAACGA | 424 |

| | | | | | |
|---|---|---|---|---|---|
| ITPRIPL1 | 134 | GAGTAGGGTTATTTTCGCGGG | 425 | CTACTTTTTTCCCGACAAAATAAAAACGT | 426 |
| FAM59B | 90 | CGCGATAGCGTTTTTTATTGTCGCG | 427 | CGCACGACCGTAAAATACTCG | 428 |
| C10orf125 | 27 | CGGTTCGTTGCGTTTATCGA | 429 | CCCCCGAACTACTCTACGCG | 430 |
| TRIM67 | 305 | GATTAAATAGTCGGGGTCGCG | 431 | ATTCTCCAACGCCAACCAC | 432 |
| SPHK2 | 284 | GTACGGTTATTGGTTGAGCGG | 433 | CCGAATCCTCCTCCAAACG | 434 |
| CALN1_B | 37 | TCGTTCGGCGTATTTATTTCGTAT | 273 | CGCGAAAAACTTCCTCCGA | 274 |
| CHST2_B | 57 | GGGATTTTTAGCGGAAGCGA | 437 | CGACGAACTATCCGACTATCACT | 438 |
| MPZ | 221 | GGTTAGGGGTGGAGTTCGTTA | 439 | ACTCCGAACTCTACTCATCCTTTC | 440 |
| CXCL12_B | 346 | TCGGCGGTTTTTAGTAAAAGCG | 441 | AAATCTCCCGTCCCACTCC | 442 |
| ODC1_B | 367 | GGTTGGTAGTCGTTTTTACGTTTTC | 443 | CAAAACCCATCTAATTACAAAATACCTCGA | 444 |
| OSR2_A | 234 | TGGAGTTATCGGAAGGCGA | 445 | CGAACTCCCGAAACGACG | 446 |
| TRH_A | 303 | TTTTCGTTGATTTTATTCGAGTCGTC | 447 | GAACCCTCTTCAAATAAACCGC | 448 |
| C17orf64_B | 336 | GATTATATTCGGATTTTGTTTATCGCGT | 449 | GACTCTTCCTACCCGCGA | 450 |

Table 29.

| **Gene Annotation** | **DMR No.** | **Probe Sequence** | **SEQ ID NO:** |
|---|---|---|---|
| CD1D | 47 | AGGCCACGGACGCGTATTGGCGCGATTTAG/3C6/ | 451 |
| ITPRIPL1 | 134 | CGCGCCGAGGGCGGTTTTAGCGATGAATC/3C6/ | 452 |

| | | | |
|---|---|---|---|
| FAM59B | 90 | AGGCCACGGACG GTCGAAATCGAAACGCTC/3C6/ | 453 |
| C10orf125 | 27 | CGCGCCGAGG GCTAACGCGAATAAAACACG/3C6/ | 454 |
| TRIM67 | 305 | AGGCCACGGACG CGAACTACGAAAACAACCTC/3C6/ | 455 |
| SPHK2 | 284 | AGGCCACGGACG GATCCCGCAAATCAACAC/3C6/ | 456 |
| CALN1_B | 37 | AGGCCACGGACG TCGTTTTTTTTTTGCGGGT/3C6/ | 376 |
| CHST2_B | 57 | CGCGCCGAGG TCGTTCCTCGATTTCGC/3C6/ | 458 |
| MPZ | 221 | CGCGCCGAGG CGTAACTCCATCTCGATAACC/3C6/ | 459 |
| CXCL12_B | 346 | CGCGCCGAGG CGCGAAATAAACCTATAATTAACTCA/3C6/ | 460 |
| ODC1_B | 367 | AGGCCACGGACG CGCGTTGGAAGTTTCG/3C6/ | 461 |
| OSR2_A | 234 | CGCGCCGAGG GCGCGAACACAAAACG/3C6/ | 462 |
| TRH_A | 303 | AGGCCACGGACG CGTTTGGCGTAGATATAAGC/3C6/ | 463 |
| C17orf64_B | 336 | CGCGCCGAGG TTTTCGTTTTCGGTTTCGG/3C6/ | 464 |

All publications and patents mentioned in the above specification are herein incorporated by reference in their entirety for all purposes. Various modifications and variations of the described compositions, methods, and uses of the technology will be apparent to those skilled in the art without departing from the scope and spirit of the technology as described. Although the technology has been described in connection with specific exemplary embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in pharmacology, biochemistry, medical science, or related fields are intended to be within the scope of the following claims.

Definitions of the specific embodiments of the invention as claimed herein follow. According to a first embodiment of the invention, there is provided a method of screening for breast cancer in a sample obtained from a subject, the method comprising:

1) assaying a methylation level of one or more chromosomal regions in a sample from the subject through:

treating DNA in the sample with a reagent that modifies DNA in a methylation-specific manner;

amplifying the treated DNA using a set of primers specific for each of the one or more chromosomal regions, wherein the one or more chromosomal regions comprises FAM59B; and

determining the methylation level of the one or more chromosomal regions by polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and/or target capture; and

2) identifying the subject as having breast cancer when the methylation state of the one or more chromosomal regions is different than a methylation state of the one or more chromosomal regions assayed in a subject that does not have breast cancer.

The term "comprise" and variants of the term such as "comprises" or "comprising" are used herein to denote the inclusion of a stated integer or stated integers but not to exclude any other integer or any other integers, unless in the context or usage an exclusive interpretation of the term is required.

Any reference to publications cited in this specification is not an admission that the disclosures constitute common general knowledge in Australia.

**CLAIMS**

1. A method of screening for breast cancer in a sample obtained from a subject, the method comprising:

1) assaying a methylation level of one or more chromosomal regions in a sample from the subject through:

treating DNA in the sample with a reagent that modifies DNA in a methylation-specific manner;

amplifying the treated DNA using a set of primers specific for each of the one or more chromosomal regions, wherein the one or more chromosomal regions comprises FAM59B; and

determining the methylation level of the one or more chromosomal regions by polymerase chain reaction, nucleic acid sequencing, mass spectrometry, methylation-specific nuclease, mass-based separation, and/or target capture; and

2) identifying the subject as having breast cancer when the methylation state of the one or more chromosomal regions is different than a methylation state of the one or more chromosomal regions assayed in a subject that does not have breast cancer.

2. The method of claim 1, wherein the one or more chromosomal regions further comprises at least one of CD1D, ITPRIPL1, C10orf125, TRIM67, SPHK2, CALN1_B, CHST2_B, MPZ, CXCL12_B, ODC1_B, OSR2_A, TRH_A, and/or C17orf64_B.

3. The method of claim 1, wherein the reagent that modifies DNA in a methylation-specific manner comprises a methylation-sensitive restriction enzyme, a methylation-dependent restriction enzyme, and/or a bisulfite reagent.

4. The method of claim 1, wherein determining the methylation level of the one or more chromosomal regions comprises using multiplex amplification, methylation-specific PCR, quantitative methylation-specific PCR, methylation-specific DNA restriction enzyme analysis, quantitative bisulfite pyrosequencing, flap endonuclease assay, PCR-flap assay, and/or bisulfite genomic sequencing PCR.

5. The method of claim 1 or claim 2, wherein the specific set of primers for each of the one or more chromosomal regions is selected from the group consisting of:

for TRH_A a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 245 and 246,

for MPZ a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 175 and 176, or SEQ ID NOS: 439 and 440,

for ITPRIPL1 a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 97 and 98, SEQ ID NOS: 99 and 100, or SEQ ID NOS: 425 and 426,

for OSR2_A a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 187 and 188,

for CHST2_B a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 39 and 40,

for C17orf64_B a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 269 and 270, or SEQ ID NOS: 449 and 450,

for CXCL12_B a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 289 and 290, or SEQ ID NOS: 441 and 442,

for ODC1_B a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 341 and 342, or SEQ ID NOS: 443 and 444,

for CD1D a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 423 and 424,

for FAM59B a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 427 and 428,

for C10orf125 a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 429 and 430,

for TRIM67 a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 431 and 432,

for SPHK2 a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 433 and 434, and

for CALN1_B a set of primers capable of binding an amplicon bound by a sequence comprising SEQ ID NOS: 273 and 274.

6. The method of claim 1, wherein the sample comprises tissue.

7. The method of claim 6, wherein the tissue is breast tissue.

8. The method of claim 1, wherein the sample is blood, serum, or plasma.

SEQUENCE LISTING

<110> MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH
EXACT SCIENCES DEVELOPMENT COMPANY, LLC

<120> DETECTING BREAST CANCER

<130> EXCTM-35440/WO-1/ORD

<150> US 62/592,828
<151> 2017-11-30

<160> 464

<170> PatentIn version 3.5

<210> 1
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 1
gagtttcggc ggcgtttttc g 21

<210> 2
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 2
cgctacgtct aacttcccgc gc 22

<210> 3
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 3
ttttcgacga gtaggattga agaaggaac 29

<210> 4
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 4
gcgaatctat ctaccgaaac gcgct 25

<210> 5
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 5
ttttgatttg taatatagag gaaagcgtcg t 31

<210> 6
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 6
gtataaacgc gtaaatacca aactaaacga a 31

<210> 7
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 7
gtttcgagaa aggagaaggg ggagc 25

<210> 8
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 8
actcccaacg aaaacttcgc aaacg 25

<210> 9
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 9
gtttttgtc gggagttatt cgt 23

<210> 10
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 10
ccaaaaatta aattaaaaac gctacgca 28

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 11
gttttaggag ggtggggcgt 20

<210> 12
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 12
aacacgacta ttcgaaaaac gcgca 25

<210> 13
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 13
ttcgtagtat cgggagtcga 20

<210> 14
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 14
gaaataataa aaacgccgca cgct 24

<210> 15
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 15
gtcgtagttt tcgcgggtgg taagc 25

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 16
cgaacgctac ctaaactctc ccgac 25

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 17
ggaatcgcga gttttgggat agtcg 25

<210> 18
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 18
aaatacaatt acaccctcta ccgcc 25

<210> 19
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 19
gaggcgttcg gtgggatttc 20

<210> 20
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 20
ccccgaccta taaacctacg acgct 25

<210> 21
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 21
gaggaggtag cgggcgtcga 20

<210> 22
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 22
cgcgtcgatc taacttacct acgaa 25

<210> 23
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 23
ttgcgtttat cgatttcgtt ttcgt 25

<210> 24
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 24
gcactactat cccccgaact actctacgc 29

<210> 25
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 25
ttattaggcg gggagtcggg tgtc 24

<210> 26
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 26
ctcgaatccc taaaaaactc gcgaa 25

<210> 27
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 27
aggaaattcg gtagcgatta tacgg 25

<210> 28
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 28
aaacccctac aacctcaccg tacacgat 28

<210> 29
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 29
cggagttaat aggtacggga ggcgt 25

<210> 30
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 30
caaacccccg aactatcgcg aa 22

<210> 31
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 31
cgggtatcgc ggttaagttg gc 22

<210> 32
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 32
tatcgtaaaa acccaacccc tcgac 25

<210> 33
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 33
gggattggtg agattcggga cgt 23

<210> 34
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 34
ctccccgaaa ccaaaaaaca acgaa 25

<210> 35
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 35
gttttaaatc gtattcgtag ttcgg 25

<210> 36
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 36
acgaacgaaa actttcctaa acgaa 25

<210> 37
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 37
gcgtttttt atcgttttag ggcgt 25

<210> 38
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 38
accgacacta ccaacctctc cgaa 24

<210> 39
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 39
tgcggggatt tttagcggaa gc 22

<210> 40
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 40
ccgacgaact atccgactat cactcgtt 28

<210> 41
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 41
gtagtaggtg gagggggcga gttc 24

<210> 42
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 42
ctctcgaaaa ccgcaaaatc ctcg 24

<210> 43
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 43
ggtaatattg cgatatttcg tagacgt 27

<210> 44
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 44
aacaatcaaa taatcgaacg cacgc 25

<210> 45
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 45
gtcgtttttc gttacgaagc ggc 23

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 46
aaaactaaat aaatctatcc tcgat 25

<210> 47
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 47
gcgtcggcgg tttttagtaa aagc 24

<210> 48
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 48
aacgaatctc attaaatctc ccgtc 25

<210> 49
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 49
gattttcggg agcggcga 18

<210> 50
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 50
cttccccgca acgaaccg 18

<210> 51
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 51
ttcgttggta tattcgcgta ggtgc 25

<210> 52
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 52
cgaataccga aatctataac cccgaa 26

<210> 53
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 53
attatgatta cgatggttga cgg 23

<210> 54
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 54
ctccataaaa acgaatttaa acgaa 25

<210> 55
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 55
tttggttata gaacgtagag gtcgt 25

<210> 56
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 56
atcgaaccac caaaccaaac gc 22

<210> 57
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 57
gggaagttta gtaggtgagc gt 22

<210> 58
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 58
actaaaaacg tttccgtcga acgca 25

<210> 59
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 59
gttggtagga gtagggttgg ttcga 25

<210> 60
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 60
atcgcaatcg taacccgtaa acgc 24

<210> 61
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 61
attcgtacgg ttttttcgtt ttcgt 25

<210> 62
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 62
gaccaactac ttccgctcga cgc 23

<210> 63
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 63
cggatttagc ggtcgagacg 20

<210> 64
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 64
tttaaaacgt ttctcgcgac gcc 23

<210> 65
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 65
tcgttaggcg atgataatta gcga 24

<210> 66
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 66
taaaaaaacc ataaacccta acgac 25

<210> 67
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 67
gttggagaag acgattcgtt cggac 25

<210> 68
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 68
ccaaaacctc actcctcaac cgc 23

<210> 69
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 69
cgcgatgcgc gttttgaac 19

<210> 70
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 70
gacgcgacta acttccaacc taacgaa 27

<210> 71
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 71
ttttcgtggt tgtcgtcgtt gc 22

<210> 72
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 72
gccgcgctct acactaaaca tattcgc 27

<210> 73
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 73
cggggaagtg ggagttttta gcg 23

<210> 74
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 74
aaaaaaacta aatcaaaacc gcgac 25

<210> 75
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 75
gcgagttcgc gttgtttacg tttc 24

<210> 76
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 76
accgacgcta cctataactc cacgct 26

<210> 77
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 77
ttaggtttgt ttattaattt tacgt 25

<210> 78
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 78
tctacaaaac gccgcgac 18

<210> 79
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 79
gttaattcga gagcgcgagg cgt 23

<210> 80
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 80
gaccaaaaaa aataaaaaat cccgcgac 28

<210> 81
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 81
taaagaaata gaaagcgggc gatacgt 27

<210> 82
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 82
cgaactaaaa aaaccgccaa cccg 24

<210> 83
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 83
gggttttgcg gttaatggcg 20

<210> 84
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 84
aataacaaac gcgtcccgaa aacga 25

<210> 85
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 85
ttagtttttt ttggttttta tttgaatttc ga 32

<210> 86
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 86
aacttttcca ccgattctca attccg 26

<210> 87
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 87
atttaaattt tcggcgtttc gtcgt 25

<210> 88
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 88
acactccaaa tcgaccttta caatcgc 27

<210> 89
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 89
agtttggttc gtttagcgat tgcgt 25

<210> 90
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 90
aacgcgacta aaaccaattt ccgca 25

<210> 91
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 91
tttatttgtt tttatcgttc gtcgg 25

<210> 92
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 92
aaatatatac ccgatttccc cgtt 24

<210> 93
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 93
taatcggcgt cgagagagat atcgt 25

<210> 94
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 94
ccgtcaacca atcgaaaacg aa 22

<210> 95
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 95
tcgtttattt cgttttttt gtcga 25

<210> 96
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 96
aaccaaccta aaatctacac tcgca 25

<210> 97
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 97
gggtcgtagg ggtttatcgc 20

<210> 98
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 98
catacttatc cgaacgtcta aacgtc 26

<210> 99
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 99
ggttttagcg atgaatcgga cgt 23

<210> 100
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 100
cacgatctta aaaaaacaac gcgac 25

<210> 101
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 101
cgtatttta ggtttagttc ggcgt 25

<210> 102
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 102
acactattac ccgcgaaaaa acgat 25

<210> 103
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 103
gagtttgttt gggggttggt cgtattc 27

<210> 104
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 104
ccaaatataa cgtttaactc tttaccacga a 31

<210> 105
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 105
tttttttga ttcggatttt ttcgg 25

<210> 106
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 106
ctaataaacg ccgccgtatt cgacg 25

<210> 107
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 107
ttttcgcgtt gtttttattt atcgt 25

<210> 108
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 108
tacacaacca cccaactact ccgcg 25

<210> 109
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 109
tgttgttggg taaaggttag tacgt 25

<210> 110
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 110
cgaaaaccca actcccgaa 19

<210> 111
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 111
tttttgcggt cgtttttcgg agc 23

<210> 112
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 112
cttaccaact aacccccgcc taccg 25

<210> 113
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 113
cgttttagta gggattgggg gcga 24

<210> 114
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 114
cccgaaaacc aaaataaaat ccgca 25

<210> 115
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 115
cggaatagcg cggtcgtttt ttc 23

<210> 116
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 116
tttaaccgta acgctcgcct cgac 24

<210> 117
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 117
gtcggttgtg tttagagcgt agcgt 25

<210> 118
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 118
aaaaaaaacc ccgacgacga a 21

<210> 119
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 119
gttgcgattg tttgtatttt gcgg 24

<210> 120
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 120
ataataacaa aaaacccctc ccgac 25

<210> 121
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 121
agtttcgtta gggaagggtt gcgtc 25

<210> 122
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 122
caactaaaac tctaccgcgc tcgat 25

<210> 123
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 123
aggaagggtt tcgagtttag tgcgg 25

<210> 124
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 124
aaaaaaatca acgcgtcctc ctcgc 25

<210> 125
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 125
tttcgatttc gtttttaaat ttcgt 25

<210> 126
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 126
aaactaaacg acctaaccct acgta 25

<210> 127
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 127
aagtttacgc gcgagtttga tcgtc 25

<210> 128
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 128
cgaaacgact tctctccccg ca 22

<210> 129
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 129
tttagttcgc ggaagttagg ttcgg 25

<210> 130
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 130
gaaaacacaa taaaccccgc cgtc 24

<210> 131
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 131
gttagattgt aggagggatt agcgg 25

<210> 132
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 132
aaaaaacgac taaaaaattc acgcc 25

<210> 133
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 133
tttggagttt gggggatcga tagtc 25

<210> 134
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 134
cgacgaaact aaaaccgcgt acgta 25

<210> 135
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 135
tttggagttt gggggatcga tagtc 25

<210> 136
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 136
cgacgaaact aaaaccgcgt acgta 25

<210> 137
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 137
attatattgg gggcgttagg ttcgg 25

<210> 138
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 138
aacaaacaat tcgcacgtaa acgaa 25

<210> 139
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 139
gggcggttta cgtggatttt tatagatttt c 31

<210> 140
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 140
gcgtctcgaa ccgtacccta acgta 25

<210> 141
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 141
cgtcgttgtt gattatgatc gcgg 24

<210> 142
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 142
cgcttcctaa caaccttcct cgaa 24

<210> 143
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 143
ttaacggtat tttttgtttt ttcgt 25

<210> 144
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 144
aaaaaaaact cgtccccgcg ct 22

<210> 145
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 145
tcggttagtt cgaggtagga agttttgc 28

<210> 146
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 146
tattaaccga aaaacgaaaa ccaaatccga 30

<210> 147
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 147
agttttgttg ttttgggtag gtcgg 25

<210> 148
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 148
aaaaactaaa aacctttctc tcgac 25

<210> 149
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 149
gcgttgagag tgacggatat ttttcgtc 28

<210> 150
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 150
actacctaaa ctccgaacac gcccg 25

<210> 151
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 151
ttagcgtatc gggaattagg gggac 25

<210> 152
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 152
gaaaacgaaa aaacgacgcg ca 22

<210> 153
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 153
tcgttttta ggtggggaag aagcg 25

<210> 154
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 154
gaaccgtatt taaaaccaat ccccgc 26

<210> 155
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 155
aattggggtt cggggttcgg tac 23

<210> 156
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 156
ttacccctac ccaaaaaaat acgct 25

<210> 157
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 157
ggggttagag tttcgcgttc gc 22

<210> 158
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 158
cgcgtctccc gtcctatcta tatacgtc 28

<210> 159
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 159
taggaatttt ttaaattcgt tttacgg 27

<210> 160
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 160
cacaaaaact cgatacaatt accgtt 26

<210> 161
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 161
tattttatag tcgcgttaaa agcgt 25

<210> 162
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 162
gtcgataaaa aacctacgcg acgaa 25

<210> 163
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 163
gatttagttt ttcgggttta tagcgg 26

<210> 164
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 164
tattaaaaac gaccaaacct ccgca 25

<210> 165
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 165
tggttgtagg cgttttgttg gagttc 26

<210> 166
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 166
aaaaacgacc ctaaccaccc tcgtt 25

<210> 167
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 167
ttttgcgtag ttgggtaggg ttcgg 25

<210> 168
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 168
cccgcattcc cgaaaaaaac gat 23

<210> 169
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 169
ttagggtttt tttcgaggag ttcga 25

<210> 170
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 170
atcccccgta cgaaactaaa cgcg 24

<210> 171
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 171
gcgttcgtat tttcgggaga ggc 23

<210> 172
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 172
tctacgtaac taaacaaaac ccgaa 25

<210> 173
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 173
cgttttgtgt tttataaaaa gaaagatttt cgg 33

<210> 174
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 174
aaaaccccaa aaacgcccga t 21

<210> 175
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 175
ggggcgtata tattagttat cgagcga 27

<210> 176
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 176
aaaaaaaacc ctaaaaaccg ccgaa 25

<210> 177
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 177
ttcgtttaat gagaaggggt tagcgg 26

<210> 178
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 178
taaaacaaac taaaaacctt aacgcgacgc t 31

<210> 179
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 179
ggggagggag tttttttac 20

<210> 180
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 180
gtacgcgaac tcgccaaaca ctacg 25

<210> 181
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 181
gtagggttgg tagtcgtttt tacgt 25

<210> 182
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 182
aacccatcta attacaaaat acctcgat 28

<210> 183
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 183
ggttttatag gggaaattat tttcgt 26

<210> 184
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 184
aaaacctcgt ctttataaca tcgaa 25

<210> 185
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 185
taggatattt cgatgttata aagacga 27

<210> 186
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 186
aacaaaacta acaaccgcct ccacg 25

<210> 187
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 187
tttggagtta tcggaaggcg aaagtac 27

<210> 188
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 188
gcacgccgaa aaaataaaaa cgaa 24

<210> 189
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 189
ttttcgatat cgatatcgaa ggcgt 25

<210> 190
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 190
ataacttaaa accctaaatt ccgcc 25

<210> 191
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 191
gcgggtagta ggaagattag tagcgg 26

<210> 192
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 192
ccgacttccg tacgaaaccg ta 22

<210> 193
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 193
taatagaggt ttgcgttgga atcga 25

<210> 194
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 194
aacgcaccct aaacaaaacc acgac 25

<210> 195
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 195
tgaagagttg ttagttcgtt tagcgt 26

<210> 196
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 196
gccaaaaatt cgattccaac gca 23

<210> 197
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 197
tagtggtagg tatagttggt agcgg 25

<210> 198
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 198
atcaaaactc ccctcctcga aaacg 25

<210> 199
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 199
gtttttaagc ggcggtcgt 19

<210> 200
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 200
aaaaaaaatc ccgttcgct 19

<210> 201
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 201
gcgcgcgttt attagatgaa gtcg 24

<210> 202
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 202
aaaatcaaaa accacaaatt caccgcc 27

<210> 203
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 203
cggggagagg aggggtagga tttac 25

<210> 204
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 204
caacttaaac accacttcct ccgaa 25

<210> 205
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 205
tgttttttt gttcgggcgg 20

<210> 206
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 206
aaataactaa ctcctactct cgcccgct 28

<210> 207
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 207
atagtttttt aattttcgcg tttcgtcga 29

<210> 208
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 208
aaaaacaact ccaacccaca ccgc 24

<210> 209
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 209
gcgggaggag taggttaatt ttcga 25

<210> 210
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 210
aaaaacaaaa tacgcgaaac gcacg 25

<210> 211
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 211
cgagaaggtt ttgtcgtaga cgtcgt 26

<210> 212
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 212
tacgtatcca tacccgcgct cg 22

<210> 213
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 213
tttgtttgta taataggggt tgcgg 25

<210> 214
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 214
cgcctaacta ccgaaaaata ccgaa 25

<210> 215
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 215
ggtggggtta ttttttttatg gagtcgattc 30

<210> 216
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 216
cgaaccaaac ctacgattcc cgaa 24

<210> 217
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 217
gggagaagag aatggttttt tgtcgtc 27

<210> 218
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 218
tcttatactc aaccccgacc taccgac 27

<210> 219
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 219
gttttattcg gggttttagc gttatttacg g 31

<210> 220
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 220
aaaaaaccgc gttactcaac gcgc 24

<210> 221
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 221
gtttagagcg gaggtagcgg ttgc 24

<210> 222
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 222
cgcctattct taaacctaaa cccgtc 26

<210> 223
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 223
cgtagaggat tataaagatt tgtacga 27

<210> 224
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 224
tactataact actacgataa cgacgacgac 30

<210> 225
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 225
agatttcggt ttttgtttcg attttcgt 28

<210> 226
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 226
attaatacta acttacgaaa ccgcc 25

<210> 227
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 227
attatttttg agcgtgaaaa atcgt 25

<210> 228
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 228
aaatttctct ccaattaaat tccgta 26

<210> 229
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 229
gtgggtttgt cgtcggattt cg 22

<210> 230
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 230
aaataaccgc gtcatccgat tcgtt 25

<210> 231
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 231
tggatgtttt atattaattt ttagttgtat aacg 34

<210> 232
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 232
gtacttttttc tctcacgaaa aatattcccg c 31

<210> 233
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 233
tgcgtggaat aaattttata tacgt 25

<210> 234
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 234
gctcaacaca cgaaaaaccc tcgaa 25

<210> 235
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 235
cggtgcgggg ttttaataaa ggatc 25

<210> 236
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 236
tccacgcaaa aacaaaaaac gcgta 25

<210> 237
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 237
ggcgtagtta tgatttcgtt ttttcgt 27

<210> 238
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 238
atcctttcga ccctacgtac ctcgat 26

<210> 239
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 239
tttacgatta ttgttttaga taatacgg 28

<210> 240
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 240
gaacccgacg aacttcgaa 19

<210> 241
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 241
gggaaatcgc gtagtttggg c 21

<210> 242
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 242
aaaacgacga aaaaacgaaa acgac 25

<210> 243
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 243
agttatcgcg atcggtttgg gttaac 26

<210> 244
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 244
aaacgattac ctctttcgtt cgttcgtt 28

<210> 245
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 245
cggcggttta tttgaagagg gttc 24

<210> 246
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 246
cgacaaatca aaaatctaca acgct 25

<210> 247
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 247
ttttaacgtt agttacgagt tgcgg 25

<210> 248
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 248
cgaacaaacc aaacaaccga a 21

<210> 249
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 249
gtagattagg cgggggcga 19

<210> 250
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 250
gaacaaaaac ataaactaat acaaatatct cccg 34

<210> 251
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 251
ggagggagag tttttcgcgg attc 24

<210> 252
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 252
ctaaacccct caaaccctaa ccgat 25

<210> 253
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 253
tgttttcgga tacggcgagc 20

<210> 254
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 254
acgaacgaac tatacgcgac gct 23

<210> 255
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 255
tggtaatcgg gtttttcgac g 21

<210> 256
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 256
ccgcgaatct atctaccgaa ac 22

<210> 257
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 257
gtgttaggtt gggcggaag 19

<210> 258
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 258
gttacccgct tacgaaaaac ga 22

<210> 259
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 259
ttcgtttttt gtcgggagtt attc 24

<210> 260
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 260
tccaaaaatt aaattaaaaa cgctacgc 28

<210> 261
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 261
ataatacggt tgttcgggag g 21

<210> 262
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 262
gtaaatataa actacgcgac gcgta 25

<210> 263
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 263
gagagtttag gtagcgttcg g 21

<210> 264
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 264
cctaacgcta ctaacaacat tataacga 28

<210> 265
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 265
cgcggttttg gagcgttag 19

<210> 266
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 266
ccgaaacgac cgaaaacgac 20

<210> 267
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 267
cggttcgttg cgtttatcga 20

<210> 268
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 268
cccccgaact actctacgcg 20

<210> 269
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 269
gattatattc ggattttgtt tatcgcgt 28

<210> 270
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 270
gactcttcct acccgcga 18

<210> 271
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 271
gcggttttta gttcgcggg 19

<210> 272
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 272
aacaaataat taacaaacaa cgcctcc 27

<210> 273
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 273
tcgttcggcg tatttatttc gtat 24

<210> 274
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 274
cgcgaaaaac ttcctccga 19

<210> 275
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 275
ggattggtga gattcgggac 20

<210> 276
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 276
cccgaaacca aaaaacaacg a 21

<210> 277
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 277
cggggagttt cgtttgtatc g 21

<210> 278
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 278
cgaataacga ctacgaactt taaacg 26

<210> 279
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 279
cgtagttata gatttattag agagggcg 28

<210> 280
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 280
ctaaaacgat aaaaaaacgc gaaacg 26

<210> 281
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 281
tggtagtttt cggtatcgac gag 23

<210> 282
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 282
taactctacg cgcaaaacgc 20

<210> 283
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 283
gggattttta gcggaagcga 20

<210> 284
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 284
cgacgaacta tccgactatc act 23

<210> 285
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 285
cggaggaatc gggtagaatc g 21

<210> 286
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 286
actctcccat aacaacgact cc 22

<210> 287
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 287
cgcgtagggc gagtttc 17

<210> 288
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 288
gcctcctcct acctctcg 18

<210> 289
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 289
tcggcggttt ttagtaaaag cg 22

<210> 290
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 290
aaatctcccg tcccactcc 19

<210> 291
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 291
ggtatattcg cgtaggtgcg 20

<210> 292
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 292
aaccgaatac cgaaatctat aaccc 25

<210> 293
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 293
gtagttgggt tgtagtgcgt g 21

<210> 294
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 294
cccgaacttc ccatcgaac 19

<210> 295
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 295
ttcgtacggt tttttcgttt tcg 23

<210> 296
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 296
ccaccacgta ataattcttc tcgaaa 26

<210> 297
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 297
cggtttgata ttattcgggg ttcg 24

<210> 298
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 298
aattaacgcc ccctataaca tcc 23

<210> 299
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 299
cgcgatagcg ttttttattg tcgcg 25

<210> 300
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 300
cgcacgaccg taaaatactc g 21

<210> 301
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 301
cggttcgtag attgttttag agcg 24

<210> 302
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 302
caaataccgt cgaaaaaaaa ctaaatcaaa ac 32

<210> 303
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 303
cgttgtagga cggttatgtc g 21

<210> 304
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 304
catcctactc ttcgaaataa accgc 25

<210> 305
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 305
agtcgttttt ttaggtagtt taggcg 26

<210> 306
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 306
cgacctttac aatcgccgc 19

<210> 307
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 307
agattggcgc gtaaaagcg 19

<210> 308
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 308
accgaccccg aaaaacg 17

<210> 309
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 309
cgttttcgga gtcgcgtg 18

<210> 310
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 310
aaccatactt atccgaacgt ctaaac 26

<210> 311
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 311
gagtagggtt attttcgcgg g 21

<210> 312
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 312
ctacttttt cccgacaaaa taaaaacgt 29

<210> 313
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 313
ttttcgcgta tttcgtggtt c 21

<210> 314
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 314
aacgccgccg tattcg 16

<210> 315
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 315
tcgcgttgga agttgcg 17

<210> 316
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 316
cgtattctaa acgctaaaaa accgc 25

<210> 317
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 317
cggcggtagt tttgcgg 17

<210> 318
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 318
ctactaaaca aaaaccaaca cgtcc 25

<210> 319
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 319
ggtaggtttg ttagttggtt ttcg 24

<210> 320
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 320
cgctatctct acgaccgcct 20

<210> 321
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 321
cggaggtagc gggcgag 17

<210> 322
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 322
cactcaccgc ttccgcc 17

<210> 323
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 323
ggcgttcgtt tcggcg 16

<210> 324
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 324
cgcttctccg acgccc 16

<210> 325
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 325
gcggttgagt ttttggtcgg 20

<210> 326
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 326
ccccgtataa ctaaaaacga cgac 24

<210> 327
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 327
cgtttttttt atgtagtaag cgatttttcg c 31

<210> 328
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 328
aaacgacgac gaacgcc 17

<210> 329
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 329
gcgttttggt tttttcgttt cgag 24

<210> 330
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 330
gaacgacgaa actaaaaccg c 21

<210> 331
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 331
cgtttttgg tagtttttt cgagtcg 27

<210> 332
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 332
gcttaaacgt aaccgaaacg cc 22

<210> 333
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 333
ggttgcgcgt cgttttttc 19

<210> 334
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 334
cccgacgcgt ttaaatcgt 19

<210> 335
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 335
ggatgggaat agttaagttt tagtcgtt 28

<210> 336
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 336
tccaacatta catacaacac taacgtc 27

<210> 337
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 337
ggttaggggt ggagttcgtt a 21

<210> 338
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 338
actccgaact ctactcatcc tttc 24

<210> 339
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 339
taggttggag attttgacgc g 21

<210> 340
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 340
cgaaacctaa aaacgccgaa ac 22

<210> 341
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 341
ggttggtagt cgtttttacg ttttc 25

<210> 342
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 342
caaaacccat ctaattacaa aatacctcga 30

<210> 343
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 343
tggagttatc ggaaggcga 19

<210> 344
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 344
cgaactcccg aaacgacg 18

<210> 345
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 345
ggaaatggtt tagagttttg gatttcg 27

<210> 346
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 346
ttctaaaaaa tactttcgat accgaca 27

<210> 347
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 347
gtggtttgaa gagttgttag ttcgtttag 29

<210> 348
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 348
gccaaaaatt cgattccaac gca 23

<210> 349
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 349
aaggtgggtt gtttgaagaa gc 22

<210> 350
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 350
accctccgac aaaaaaacgt ac 22

<210> 351
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 351
gcgagaaggt tttgtcgtag a 21

<210> 352
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 352
acctactcac gcacaacct 19

<210> 353
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 353
cgcggtgagg aagatcg 17

<210> 354
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 354
acgccaccta cgactacg 18

<210> 355
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 355
gtacggttat tggttgagcg g 21

<210> 356
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 356
ccgaatcctc ctccaaacg 19

<210> 357
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 357
ggaatttaat tggagagaaa ttttggcg 28

<210> 358
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 358
ccaaaatttc cctcatctat atacgcc 27

<210> 359
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 359
gttgcgggtc gggttgc 17

<210> 360
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 360
gcaaaaacaa aaaacgcgta aaaacc 26

<210> 361
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 361
gtcgtcgttg tcgtagttgt c 21

<210> 362
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 362
cgtaaaaacc gaacgacgcg 20

<210> 363
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 363
ttttcgttga ttttattcga gtcgtc 26

<210> 364
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 364
gaaccctctt caaataaacc gc 22

<210> 365
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 365
gattaaatag tcggggtcgc g 21

<210> 366
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 366
attctccaac gccaaccac 19

<210> 367
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 367
cgcgccgagg cgcgcttcca ctcc 24

<210> 368
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 368
aggccacgga cggcggcgtt ttttttatg ttg 33

<210> 369
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 369
aggccacgga cgcaaccgaa ctaaaaaaaa aaactaacg 39

<210> 370
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 370
cgcgccgagg gcgcgtaaga ttttcgg 27

<210> 371
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 371
cgcgccgagg gcgggtagta gtgcg 25

<210> 372
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 372
cgcgccgagg cgaccgaaaa atcgaaaaac a 31

<210> 373
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 373
cgcgccgagg gctaacgcga ataaaacacg 30

<210> 374
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 374
cgcgccgagg ttttcgtttt cggtttcgg 29

<210> 375
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 375
cgcgccgagg ccgtacctat taactccg 28

<210> 376
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 376
aggccacgga cgtcgttttt tttttgcggg t 31

<210> 377
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 377
aggccacgga cgcgtattgg cgcgatttag 30

<210> 378
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 378
aggccacgga cggttcgaaa aaaactcgac gaa 33

<210> 379
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 379
aggccacgga cggccgttct ctaacttccg 30

<210> 380
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 380
aggccacgga cgccgaaata cgaacgcga 29

<210> 381
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 381
aggccacgga cgtcgttcct cgatttcgc 29

<210> 382
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 382
aggccacgga cgcgaataaa acctacgaaa aaaaacg 37

<210> 383
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 383
aggccacgga cggaaaaccg caaaatcctc g 31

<210> 384
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 384
aggccacgga cgcgcgaaat aaacctataa ttaactca 38

<210> 385
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 385
cgcgccgagg ccgaaccaac actcaaaac 29

<210> 386
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 386
cgcgccgagg gcgcgtttgg tttggt 26

<210> 387
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 387
aggccacgga cgaacgcgct ccaacc 26

<210> 388
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 388
cgcgccgagg cgcgacgact aaaaaaattc a 31

<210> 389
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 389
aggccacgga cggtcgaaat cgaaacgctc 30

<210> 390
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 390
cgcgccgagg ccgcgactac ctcttc 26

<210> 391
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 391
aggccacgga cgcgacgtcc tacaaaacca 30

<210> 392
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 392
cgcgccgagg ggcggtagtt gttgc 25

<210> 393
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 393
cgcgccgagg gcgaaaaccc cgcc 24

<210> 394
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 394
cgcgccgagg cgtctaacta aacgcgataa ac 32

<210> 395
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 395
cgcgccgagg gcggttttag cgatgaatc 29

<210> 396
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 396
cgcgccgagg cgattcgagg gcgt 24

<210> 397
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 397
aggccacgga cgcgcgacgc aaaactc 27

<210> 398
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 398
aggccacgga cggcggcggt tggatt 26

<210> 399
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 399
aggccacgga cgtcccgaaa cgaacgataa a 31

<210> 400
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 400
cgcgccgagg cgccgtcgca ctac 24

<210> 401
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 401
aggccacgga cgcgcgactc cccact 26

<210> 402
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 402
aggccacgga cgcgcaaata ataacgcgaa cg 32

<210> 403
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 403
aggccacgga cgcgttcgag gttagttttt tgg 33

<210> 404
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 404
aggccacgga cgcgtacgta acccgcg 27

<210> 405
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 405
cgcgccgagg cgctactaac cataaccgc 29

<210> 406
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 406
cgcgccgagg cgttttcgtt tgattcggtt 30

<210> 407
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 407
cgcgccgagg tcggtgattg atgtgtgcg 29

<210> 408
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 408
cgcgccgagg cgtaactcca tctcgataac c 31

<210> 409
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 409
cgcgccgagg cgaccgcgaa aaaacg 26

<210> 410
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 410
aggccacgga cgcgcgttgg aagtttcg 28

<210> 411
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 411
cgcgccgagg gcgcgaacac aaaacg 26

<210> 412
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 412
cgcgccgagg accgaaaacg ccctaaa 27

<210> 413
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 413
cgcgccgagg gcgtggagaa atgttagttt g 31

<210> 414
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 414
aggccacgga cgcgggcggt gaatttgt 28

<210> 415
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 415
aggccacgga cgacgtcgta tttgtggcg 29

<210> 416
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 416
aggccacgga cggcgttgtt tagcgcg 27

<210> 417
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 417
aggccacgga cggatcccgc aaatcaacac 30

<210> 418
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 418
cgcgccgagg cgatccccaa ctccc 25

<210> 419
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 419
cgcgccgagg cgcttctaaa acctcgatcc 30

<210> 420
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 420
cgcgccgagg cgcggtcgtt aatatgtatt c 31

<210> 421
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 421
aggccacgga cgcgtttggc gtagatataa gc 32

<210> 422
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 422
aggccacgga cgcgaactac gaaaacaacc tc 32

<210> 423
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 423
ggattggtga gattcgggac 20

<210> 424
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 424
cccgaaacca aaaaacaacg a 21

<210> 425
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 425
gagtagggtt attttcgcgg g 21

<210> 426
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 426
ctacttttt cccgacaaaa taaaaacgt 29

<210> 427
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 427
cgcgatagcg ttttttattg tcgcg 25

<210> 428
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 428
cgcacgaccg taaaatactc g 21

<210> 429
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 429
cggttcgttg cgtttatcga 20

<210> 430
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 430
cccccgaact actctacgcg 20

<210> 431
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 431
gattaaatag tcggggtcgc g 21

<210> 432
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 432
attctccaac gccaaccac 19

<210> 433
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 433
gtacggttat tggttgagcg g 21

<210> 434
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 434
ccgaatcctc ctccaaacg 19

<210> 435
<211> 0
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 435
000

<210> 436
<211> 0
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 436
000

<210> 437
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 437
gggatttta gcggaagcga 20

<210> 438
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 438
cgacgaacta tccgactatc act 23

<210> 439
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 439
ggttaggggt ggagttcgtt a 21

<210> 440
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 440
actccgaact ctactcatcc tttc 24

<210> 441
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 441
tcggcggttt ttagtaaaag cg 22

<210> 442
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 442
aaatctcccg tcccactcc 19

<210> 443
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 443
ggttggtagt cgtttttacg ttttc 25

<210> 444
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 444
caaaacccat ctaattacaa aatacctcga 30

<210> 445
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 445
tggagttatc ggaaggcga 19

<210> 446
<211> 18

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 446
cgaactcccg aaacgacg 18

<210> 447
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 447
ttttcgttga ttttattcga gtcgtc 26

<210> 448
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 448
gaaccctctt caaataaacc gc 22

<210> 449
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 449
gattatattc ggattttgtt tatcgcgt 28

<210> 450
<211> 18
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 450
gactcttcct acccgcga 18

<210> 451
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 451
aggccacgga cgcgtattgg cgcgatttag 30

<210> 452
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 452
cgcgccgagg gcggttttag cgatgaatc 29

<210> 453
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 453
aggccacgga cggtcgaaat cgaaacgctc 30

<210> 454
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 454
cgcgccgagg gctaacgcga ataaaacacg 30

<210> 455
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 455
aggccacgga cgcgaactac gaaaacaacc tc 32

<210> 456
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 456
aggccacgga cggatcccgc aaatcaacac 30

<210> 457
<211> 0
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 457
000

<210> 458
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 458
cgcgccgagg tcgttcctcg atttcgc 27

<210> 459
<211> 31
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 459
cgcgccgagg cgtaactcca tctcgataac c 31

<210> 460
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 460
cgcgccgagg cgcgaaataa acctataatt aactca 36

<210> 461
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 461
aggccacgga cgcgcgttgg aagtttcg 28

<210> 462
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 462
cgcgccgagg gcgcgaacac aaaacg 26

<210> 463
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 463
aggccacgga cgcgtttggc gtagatataa gc 32

<210> 464
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 464
cgcgccgagg ttttcgtttt cggtttcgg 29