

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 June 2009 (18.06.2009)

PCT

(10) International Publication Number
WO 2009/076478 A2

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

(21) International Application Number:
PCT/US2008/086294

(22) International Filing Date:
10 December 2008 (10.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/007,025 10 December 2007 (10.12.2007) US

(71) Applicant (*for all designated States except US*): **THE JOHNS HOPKINS UNIVERSITY** [US/US]; Johns Hopkins Technology Transfer, 100 North Charles Street, 5th Floor, Baltimore, MD 21201 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **CALIFANO, Joseph, A.** [US/US]; 228 Hopkins Lane, Owings Mills, MD 21117 (US). **SMITH, Ian, M.** [US/US]; 1911 Gough Street, Baltimore, MD 21231 (US).

(74) Agent: **HAILE, Lisa, A.**; DLA Piper LLP (US), Suite 1100, 4365 Executive Drive, San Diego, CA 92121-2133 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: HYPOMETHYLATED GENES IN CANCER

(57) Abstract: The present invention provides methods and kits for identifying a cell that exhibits or is predisposed to exhibiting unregulated growth by detecting hypomethylation of a gene or a regulatory region in at least one gene in the cell. Also provided are methods for diagnosis or prognosis of a proliferative disorder in a subject. Also provided are methods of ameliorating a cell proliferative disorder in a subject by administering to the subject an agent that methylates a hypomethylated gene or regulatory region thereof. In some aspects, the gene or regulatory region thereof is TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRTN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISSIR, KIAA1937 protein, SOX30, DEAD, or KBGP.



WO 2009/076478 A2

HYPOMETHYLATED GENES IN CANCER

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0001] The present invention relates generally to methods for detecting the presence of or risk of developing cancer and more specifically to methods for detecting the presence of hypomethylation in various genes or regulatory regions thereof.

BACKGROUND INFORMATION

[0002] Epigenetic alterations in promoter methylation and histone acetylation have been associated with cancer-specific expression differences in human malignancies, including for example, head and neck cancer and non-small cell lung carcinoma (NSCLC). Methylation has been primarily considered as a mechanism of tumor suppressor gene (TSG) inactivation, and comprehensive whole-genome profiling approaches to promoter hypermethylation have identified multiple novel putative TSGs silenced by promoter hypermethylation.

[0003] Indirect evidence supports a role for hypomethylation in tumor development. Global genomic hypomethylation has been reported in almost all solid tumors. Mice with functional disruption of DNA methyltransferase 1 (*DNMT1*) function demonstrate significant genomic hypomethylation in all tissues and develop aggressive T-cell lymphomas with chromosomal instability. In solid human tumors, meta-analysis shows an overall correlation between global hypomethylation and advanced tumor stage.

[0004] To date, only sporadic examples of promoter hypomethylation associated with unmasked expression of putative oncogenes have been reported, including: *R-Ras* in gastric cancer, *c-Neu* in transgenic mouse models, the *Hox11* proto-oncogene in leukemia, *BCL-2* gene hypomethylation and high-level expression in B-cell chronic lymphocytic lymphomas, demethylation in MMTV/N-rasN transgenic mice, and rare activation of two RAS family members in colon cancer and small cell lung cancer. These observations demonstrate that proto-oncogenes with tissue-specific or developmentally restricted expression—*i.e.*, during early growth, differentiation, or gametogenesis—may be inappropriately re-expressed in cancers via epigenetic alteration, including demethylation.

[0005] Cancer/testis antigens (CTAs) have been shown to be overexpressed in various tumor types, with little or no expression in normal human tissue; however, the mechanism of this differential expression is not well understood. It is also known that CTAs, especially those encoded by the X chromosome (CT-X antigens), are expressed in association with promoter demethylation or whole genomic hypomethylation. To date, a comprehensive, genome-wide approach to identify coordinately expressed CTAs and other differentially expressed genes activated by promoter demethylation in NSCLC has not been conducted.

SUMMARY OF THE INVENTION

[0006] The present invention is based on the discovery that some genes have promoters that are demethylated and transcriptionally upregulated in cancer. This discovery is useful for cancer screening, risk-assessment, prognosis, and identification of subjects responsive to a therapeutic regimen. Accordingly, there are provided methods for detecting a cellular proliferative disorder in a subject. The subject may have or be at risk of having a cellular proliferative disorder. The method of the invention is useful for diagnostic as well as prognostic analyses.

[0007] In one embodiment of the invention, there are provided methods for identifying a cell that exhibits or is predisposed to exhibiting unregulated growth. The method includes detecting hypomethylation of a gene or a regulatory region in at least one gene in the cell, wherein the at least one gene is hypomethylated as compared to a corresponding normal cell not exhibiting unregulated growth, thereby identifying the cell as exhibiting or predisposed to exhibiting unregulated growth. In some aspects, at least two genes or regulatory regions are hypomethylated and the at least two genes are coordinately expressed in the cell undergoing unregulated cell growth. In particular embodiments, the regulatory region of the at least one gene comprises a BORIS binding site. In certain embodiments, the regulatory region of the at least one gene includes a promoter of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL (cyteine-rich secretory protein LCCL domain), KRT86 (keratin, hair, basic, 6 monilethrix), KIPV467, KRT81 (keratin, hair, basic, 1), CSPG5 (chondroitin sulfate proteoglycan 5), PP1R14A (protein phosphatase 1, regulatory inhibitor 14), KISS1R (G protein coupled receptor 54), KIAA1937 protein, SOX30 (SRY sex

determining region Y-box 30), DEAD (Asp-Glu-Ala-Asp box polypeptide), and KBGP (Kell Blood group precursor McLeod phenomenon).

[0008] In another embodiment, there are provided methods for diagnosing a disorder in a subject having or at risk of developing a cell proliferative disorder. The method includes contacting a nucleic acid-containing sample from cells of the subject with an agent that provides a determination of the methylation state of at least one regulatory region of a gene, wherein the at least one regulatory region is hypomethylated in a cell undergoing unregulated cell growth as compared to a corresponding normal cell; and identifying hypomethylation of the regulatory region in the nucleic acid-containing sample, as compared to the same region of the at least one regulatory region in a subject not having the proliferative disorder, wherein hypomethylation is indicative of a subject having or at risk of developing the proliferative disorder. In some aspects, at least two genes or regulatory regions are hypomethylated and the at least two genes are coordinately expressed in the cell undergoing unregulated cell growth. In particular embodiments, the regulatory region of the at least one gene comprises a BORIS binding site. In certain embodiments, the regulatory region of the at least one gene includes a promoter of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP.

[0009] In yet another embodiment of the invention, there are provided methods of determining the prognosis of a subject having a cell proliferative disorder. The method includes determining the methylation state of at least one regulatory region of a gene in a nucleic acid sample from the subject, wherein hypomethylation as compared to a corresponding normal cell in the subject or a subject not having the disorder, is indicative of a poor prognosis. In some aspects, at least two genes or regulatory regions are hypomethylated and the at least two genes are coordinately expressed in the cell undergoing unregulated cell growth. In particular embodiments, the regulatory region of the at least one gene comprises a BORIS binding site. In certain embodiments, the regulatory region of the at least one gene includes a promoter of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1,

MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP.

[0010] In still another embodiment of the present invention, there are provided methods of ameliorating a cell proliferative disorder in a subject in need thereof. The method includes administering to the subject an agent that methylates at least one regulatory region in a gene that is demethylated as compared to a subject not having the disorder, thereby reducing expression of the at least one gene and ameliorating the cell proliferative disorder. In some aspects, at least two genes or regulatory regions are hypomethylated and the at least two genes are coordinately expressed in the cell undergoing unregulated cell growth. In particular embodiments, the regulatory region of the at least one gene comprises a BORIS binding site. In certain embodiments, the regulatory region of the at least one gene includes a promoter of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP.

[0011] In a further embodiment of the present invention, there are provided methods of identifying a gene activated by hypomethylation. The method includes comparing an expression analysis of a cell treated with an agent that reduces methylation to an expression analysis of a control cell not treated with the agent, wherein an increase in expression of a gene is indicative of a gene activated by demethylation. Certain embodiments may further include an expression analysis of a tissue sample and a tumor sample from the same tissue of origin as the normal cell, wherein an increase in expression of a gene in a tumor sample as compared to a normal sample is correlated to the genes activated by demethylation in the treated cell.

[0012] In another embodiment of the present invention, there are provided methods for determining whether a subject is responsive to a particular therapeutic regimen. The method includes determining the methylation status of one or more genes or regulatory regions thereof, selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5,

NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP, wherein hypomethylation of the gene or regulatory region thereof as compared with a normal subject is indicative of a subject who is responsive to the therapeutic regimen. In certain embodiments, therapeutic regimen is administration of a chemotherapeutic agent. In other embodiments, the therapeutic regimen is administration of a vaccine directed to a protein encoded by the hypomethylated gene.

[0013] In another embodiment, the invention provides a kit useful for the detection of a methylated CpG-containing nucleic acid in determining the methylation status of one or more genes or regulatory regions thereof, selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP. The kit includes carrier element containing one or more containers having a first container containing a reagent which modifies unmethylated cytosine and a second container containing primers for amplification of the one or more genes or regulatory regions thereof, wherein the primers distinguish between modified methylated and nonmethylated nucleic acid.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Figure 1A provides a schematic of the integrative epigenetic screening strategy employed in Example 1. Figure 1B shows a representative COPA graph of *MAGEA3* demonstrating the statistical approach to finding candidate overexpressed oncogenes. Figure 1C shows a plot of the upregulation of expression of the target genes after treatment with 5-aza/TSA in cell lines as measured by QRT-PCR.

[0015] Figure 2 depicts the promoter methylation status in primary tissues. Figure 2A shows the bisulfite sequencing results in 10 tumors and 10 normals for: *TKTL1* (4/10, $p < 0.05$), *H19* (6/10, $p < 0.05$), *MAGEA2* (5/10, $p < 0.05$), *MAGEA3* (5/10, $p < 0.05$), *MAGEA4* (5/10, $p < 0.05$), *MAGEA11* (5/10, $p < 0.05$), *GPR17* (3/10, $p < 0.10$), *GRIN1* (6/10, $p < 0.05$), *C19ORF28* (5/10, $p < 0.05$). Figures 2B-J show the QRT-PCR expression with the

bisulfite sequencing of the respective promoter below (white is unmethylated, grey is methylated).

[0016] Figure 3 shows depicts the results of the transient transfection of target genes in minimally transformed oral keratinocytes. Figure 3A-D show plots of the transient transfection of an *H19* construct, a *MAGEA2* construct, a *TKTL1* construct, and a *MAGEA4* construct into OKF6-Tert-1R cells, for A-D, respectively. Figure 3E shows plots of the QUMSP percentage of *C19ORF28*, *GRIN1*, *H19*, *MAGEA11*, *MAGEA2*, *MAGEA3*, *GPR17*, and *TKTL1* conducted in a separate cohort of head and neck cancer patients using 25 tumors and 11 upper aerodigestive mucosal samples to assay promoter demethylation. Figure 3F shows the QUMSP results for an independent cohort of 14 lung normals and 13 lung tumor patients. Significant differences in QUMSP were found in *H19*, *MAGEA11*, *MAGEA2*, and *MAGEA3*.

[0017] Figure 4 shows plots demonstrating TKTL1 forced overexpression via transient transfection in background low expressing JHU-011 cells (Figure 4A); induces increased anchorage dependent colony formation and TKTL1 shRNA in high-expressing FaDu cell line (Figure 4B); induces growth inhibition. Anchorage independent growth of UM-22B cells is significantly inhibited by TKTL1 shRNA (Figure 4C); with decrease in colony size (Figure 4D).

[0018] Figure 5 shows plots depicting overexpression and demethylation in human cancers (non small cell lung cancer, lymphoma, melanoma, pancreatic cancer, and urothelial cancer). Figure 5A shows the expression of *H19* in these cancers. Figure 5B shows *MAGEA2* expression, Figure 5C shows *TKTL1* expression, and Figure 5D shows *MAGEA4*.

[0019] Figure 6 shows gene expression and demethylation correlation. Figure 6A shows the gene expression correlation p-value matrix for the coexpression for each gene pair across all tumors. This comparison shows the correlation of each gene pair in 49 head and neck tumors. Figure 6B shows the gene pair expression p-value correlation matrix for 80 NSCLC. Figure 6C shows a phylogram of the promoters of interest based on ClustalW analysis after multiple sequence alignment. The region of significant homology is shown after sequence alignment. Figure 6D shows the promoter hypomethylation (QUMSP) correlation p-value

matrix for HNSCC (25 tumors). Figure 6E shows the promoter hypomethylation (QUMSP) correlation p-value matrix for NSCLC (13 tumors).

[0020] Figure 7A shows a correlation of *BORIS* expression with expression of target genes in HNSCC (QRT-PR) heat map (Pearson correlation). Figure 7B shows plots of the growth of cells following the transient transfection of *BORIS* construct into NIH-3T3 and OKF6-Tert1R cell lines. Figure 7C shows a plot of the anchorage independent growth assayed after transfection with empty vector (EV), CTCF, and *BORIS* at various concentrations of doxycycline, with a representative colony shown. Figure 7D shows a plot of QUMSP of nine targets of interest after transfection with empty vector (untreated) and *BORIS* construct (treated) in presence of 0.0625 ug/ml of doxycycline. Figure 7E shows a plot of the fold increase in expression measured by quantitative RT-PCR of nine targets of interest after *BORIS* transfection.

[0021] Figure 8 shows a plot of the fold upregulation of mRNA expression in treated minimally-transformed cell lines measured by Affymetrix U133 Plus 2.0.

[0022] Figure 9 shows *BORIS* correlates with gene expression in all cancers (using the expO cohort of 1041 human cancers of various tumor sites and histologies). Shown are microarray median-normalized expression of the targets compared to *BORIS* expression in 1041 human cancers

[0023] Figure 10A shows the integrative epigenetic screening strategy employed in Example 2. Figure 10B shows a plot of the mRNA expression in treated normal lung cell lines, NHBE and SAEC. Figure 10C shows a representative COPA graph of *MAGEA12* demonstrating the statistical approach used to find candidate overexpressed CTAs and related genes.

[0024] Figure 11 shows the bisulfite sequencing results with associated p values in 28 NSCLC tumor samples and 11 normal lung tissues for the indicated genes. Shaded boxes represent methylated promoters, "ND" = methylation status not determined by bisulfite sequencing.

[0025] Figure 12A shows a plot of QUMSP conducted in a cohort of 28 NSCLC and 11 normal lung tissues. Figure 12B shows the promoter hypomethylation (QUMSP) correlation p-value matrix for NSCLC (n=28; Spearman's correlation permutation test).

[0026] Figure 13A shows a heat map of transcript expression as measured by the Affymetrix Human Genome U133 Plus 2.0 mRNA expression platform for 40 normal lung samples from non-cancer patients and 111 NSCLC primary tissue samples. Figure 13B shows the Pearson's correlation coefficient p-value matrix for gene expression which tests for the coexpression of each gene pair across all tumors. This comparison shows the expression correlation of each gene pair in 111 NSCLC. Values to the upper right have been corrected with the Benjamin Hochberg multiple test correction to decrease the false discovery rate; uncorrected values are displayed in the lower left. Shaded cells represent significant p-values.

[0027] Figure 14 shows a plot of the relative fold upregulation of expression of each gene after treatment with 5-aza/TSA is shown in NHBE and SAEC cell lines as measured by quantitative RT-PCR.

[0028] Figure 15 depicts the target gene expression is upregulated in NSCLC vs. normal lung tissues. Figures 15A-J show the quantitative RT-PCR in a cohort of 28 NSCLC and 5 normal lung tissues for each gene.

[0029] Figure 16A-D shows scatter plots showing Log2 QRT-PCR values plotted against Log2 QUMSP for 28 NSCLC and 5 normal lung tissues for *MAGEA12* (p=0.024), *MAGEA4* (p<0.004), *SBSN* (p=0.004) and *NY-ESO-1* (p<0.004).

[0030] Figure 17 shows a table of containing a ranking of the 290 significant genes found after combing the three rank ordered lists ("-" = Not determined).

DETAILED DESCRIPTION OF THE INVENTION

[0031] The present invention is based on the discovery that several genes have promoters that are demethylated and transcriptionally upregulated in cancer. Accordingly, in a first embodiment of the invention, there are provided methods for identifying a cell that exhibits or is predisposed to exhibiting unregulated growth. The method includes detecting hypomethylation of a gene or a regulatory region in at least one gene in the cell, wherein the at least one gene is hypomethylated as compared to a corresponding normal cell not

exhibiting unregulated growth, thereby identifying the cell as exhibiting or predisposed to exhibiting unregulated growth.

[0032] The genes or regulatory regions thereof whose methylation status is detected in the methods provided herein can be any gene or regulatory region thereof identified as hypomethylated in a cell exhibiting unregulated growth as compared to a corresponding normal cell, not undergoing unregulated cell growth. In certain embodiments, at least two genes or regulatory regions are hypomethylated and the at least two genes are coordinately expressed in the cell undergoing unregulated cell growth. In other aspects, at least three, or at least four, or at least five, or more genes or regulatory regions are hypomethylated.

[0033] In certain embodiments, the gene or regulatory region is one or more of the genes identified herein (the “target genes”). In particular embodiments, the gene or regulatory region thereof is selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP. In some embodiments, the gene or regulatory region thereof is one or more of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, and ZNF711. In certain embodiments, the gene or regulatory region thereof is one or more of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, and C19ORF28. In other embodiments the gene or regulatory region thereof is one or more of MAGEA3, MAGEA12, MAGEA4, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, and CT45-2.

[0034] As provided herein, hypomethylation may occur in the gene or regulatory region thereof. In some embodiments, the hypomethylation occurs within the regulatory region of the genes identified herein, in particular embodiments, the hypomethylation is in the promoter sequence of the regulatory region. More particularly, the hypomethylation may be in a CpG dinucleotide motif of the promoter. In some embodiments, the regulatory region of the at least one gene comprises a BORIS binding site.

[0035] In particular embodiments, the methylation status of the regulatory regions of TKTL1, GRIN1, and GPR17 is determined. In other embodiments, the methylation status of the regulatory regions of MAGEA2, MAGEA3, MAGEA4, MAGEA11 and H19 is determined. In other embodiments, the methylation status of the regulatory regions of MAGEA3, MAGEA12, MAGEA4, MAGEA1, MAGEA5, and NY-ESO-1 is determined.

[0036] In another embodiment, there are provided methods for diagnosing a disorder in a subject having or at risk of developing a cell proliferative disorder. The method includes contacting a nucleic acid-containing sample from cells of the subject with an agent that provides a determination of the methylation state of at least one regulatory region of a gene, wherein the at least one regulatory region is hypomethylated in a cell undergoing unregulated cell growth as compared to a corresponding normal cell; and identifying hypomethylation of the regulatory region in the nucleic acid-containing sample, as compared to the same region of the at least one regulatory region in a subject not having the proliferative disorder, wherein hypomethylation is indicative of a subject having or at risk of developing the proliferative disorder.

[0037] The term “cell proliferative disorder” as used herein refers to malignant as well as non-malignant cell populations which often differ from the surrounding tissue both morphologically and genotypically. In some embodiments, the cell proliferative disorder is a cancer. In particular embodiments the cancer may be a carcinoma or a sarcoma. A cancer can include, but is not limited to, head cancer, neck cancer, head and neck cancer, lung cancer, breast cancer, prostate cancer, colorectal cancer, esophageal cancer, stomach cancer, leukemia/lymphoma, uterine cancer, skin cancer, endocrine cancer, urinary cancer, pancreatic cancer, gastrointestinal cancer, ovarian cancer, cervical cancer, and adenomas. In one aspect, the cancer is head and neck cancer. In another aspect, the cancer is lung cancer.

[0038] The nucleic acid-containing sample for use in the invention methods may be virtually any biological sample that contains nucleic acids from the subject. The biological sample can be a tissue sample which contains 1 to 10,000,000, 1000 to 10,000,000, or 1,000,000 to 10,000,000 somatic cells. However, it is possible to obtain samples that contain smaller numbers of cells, even a single cell in embodiments that utilize an amplification protocol such as PCR. The sample need not contain any intact cells, so long as it contains sufficient material (e.g., protein or genetic material, such as RNA or DNA) to assess

methylation status or gene expression levels. In some embodiments the nucleic acid-containing sample is obtained from cells are from a sample selected from the group consisting of a tissue sample, a frozen tissue sample, a biopsy specimen, a surgical specimen, a cytological specimen, whole blood, bone marrow, cerebral spinal fluid, peritoneal fluid, pleural fluid, lymph fluid, serum, mucus, plasma, urine, chyle, stool, ejaculate, sputum, nipple aspirate and saliva.

[0039] A biological or tissue sample can be drawn from any tissue that is susceptible to cancer. For example, the tissue may be obtained by surgery, biopsy, swab, stool, or other collection method. The biological sample for methods of the present invention can be, for example, a sample from colorectal tissue, or in certain embodiments, can be a blood sample, or a fraction of a blood sample such as a peripheral blood lymphocyte (PBL) fraction. Methods for isolating PBLs from whole blood are well known in the art. An example of such a method is provided in the Example section herein. In addition, it is possible to use a blood sample and enrich the small amount of circulating cells from a tissue of interest, e.g., lung, colon, breast, etc. using a method known in the art.

[0040] In the present invention, the subject is typically a human, but also can be any mammal, including, but not limited to, a dog, cat, rabbit, cow, rat, horse, pig, or monkey.

[0041] Numerous methods for analyzing methylation status of a gene or regulatory region are known in the art and can be used in the methods of the present invention to identify hypomethylation. As illustrated in the Examples herein, analysis of methylation can be performed by bisulfite genomic sequencing.

[0042] Bisulfite ions, for example, sodium bisulfite, convert non-methylated cytosine residues to bisulfite modified cytosine residues. The bisulfite ion treated gene sequence can be exposed to alkaline conditions, which convert bisulfite modified cytosine residues to uracil residues. Sodium bisulfite reacts readily with the 5,6-double bond of cytosine (but poorly with methylated cytosine) to form a sulfonated cytosine reaction intermediate that is susceptible to deamination, giving rise to a sulfonated uracil. The sulfonate group can be removed by exposure to alkaline conditions, resulting in the formation of uracil. The DNA can be amplified, for example, by PCR, and sequenced to determine whether CpG sites are methylated in the DNA of the sample. Uracil is recognized as a thymine by Taq polymerase

and, upon PCR, the resultant product contains cytosine only at the position where 5-methylcytosine was present in the starting template DNA. One can compare the amount or distribution of uracil residues in the bisulfite ion treated gene sequence of the test cell with a similarly treated corresponding non-methylated gene sequence. A decrease in the amount or distribution of uracil residues in the gene from the test cell indicates methylation of cytosine residues in CpG dinucleotides in the gene of the test cell. The amount or distribution of uracil residues also can be detected by contacting the bisulfite ion treated target gene sequence, following exposure to alkaline conditions, with an oligonucleotide that selectively hybridizes to a nucleotide sequence of the target gene that either contains uracil residues or that lacks uracil residues, but not both, and detecting selective hybridization (or the absence thereof) of the oligonucleotide.

[0043] In another embodiment, the gene is contacted with hydrazine, which modifies cytosine residues, but not methylated cytosine residues, then the hydrazine treated gene sequence is contacted with a reagent such as piperidine, which cleaves the nucleic acid molecule at hydrazine modified cytosine residues, thereby generating a product comprising fragments. By separating the fragments according to molecular weight, using, for example, an electrophoretic, chromatographic, or mass spectrographic method, and comparing the separation pattern with that of a similarly treated corresponding non-methylated gene sequence, gaps are apparent at positions in the test gene contained methylated cytosine residues. As such, the presence of gaps is indicative of methylation of a cytosine residue in the CpG dinucleotide in the target gene of the test cell.

[0044] Modified products can be detected directly, or after a further reaction which creates products which are easily distinguishable. Means which detect altered size and/or charge can be used to detect modified products, including but not limited to electrophoresis, chromatography, and mass spectrometry. Examples of such chemical reagents for selective modification include hydrazine and bisulfite ions. Hydrazine-modified DNA can be treated with piperidine to cleave it. Bisulfite ion-treated DNA can be treated with alkali. Other means which are reliant on specific sequences can be used, including but not limited to hybridization, amplification, sequencing, and ligase chain reaction. Combinations of such techniques can be used as is desired.

[0045] In another example, methylation status may be assessed using real-time methylation specific PCR. For example, the methylation level of the promoter region of one or more of the target genes can be determined by determining the amplification level of the promoter region of the target gene based on amplification-mediated displacement of one or more probes whose binding sites are located within the amplicon. In general, real-time quantitative methylation specific PCR is based on the continuous monitoring of a progressive fluorogenic PCR by an optical system. Such PCR systems are well-known in the art and usually use two amplification primers and an additional amplicon-specific, fluorogenic hybridization probe that specifically binds to a site within the amplicon. The probe can include one or more fluorescence label moieties. For example, the probe can be labeled with two fluorescent dyes: 1) a 6-carboxy-fluorescein (FAM), located at the 5'-end, which serves as reporter, and 2) a 6-carboxy-tetramethyl-rhodamine (TAMRA), located at the 3'-end, which serves as a quencher. When amplification occurs, the 5'-3' exonuclease activity of the Taq DNA polymerase cleaves the reporter from the probe during the extension phase, thus releasing it from the quencher. The resulting increase in fluorescence emission of the reporter dye is monitored during the PCR process and represents the number of DNA fragments generated.

[0046] In other embodiments, hypomethylation can be identified through nucleic acid sequencing after bisulfite treatment to determine whether a uracil or a cytosine is present at specific location within a gene or regulatory region. If uracil is present after bisulfite treatment, then the nucleotide was unmethylated. Hypomethylation is present when there is a measurable decrease in methylation.

[0047] In an alternative embodiment, the method for analyzing methylation of the target gene can include amplification using a primer pair specific for methylated residues within a the target gene. In these embodiments, selective hybridization or binding of at least one of the primers is dependent on the methylation state of the target DNA sequence (Herman et al., *Proc. Natl. Acad. Sci. USA*, 93:9821 (1996)). For example, the amplification reaction can be preceded by bisulfite treatment, and the primers can selectively hybridize to target sequences in a manner that is dependent on bisulfite treatment. For example, one primer can selectively bind to a target sequence only when one or more base of the target sequence is altered by bisulfite treatment, thereby being specific for a methylated target sequence.

[0048] Other methods are known in the art for determining methylation status of a target gene, including, but not limited to, array-based methylation analysis (*see* Gitan *et al.*, *Genome Res* 12:158-64, 2002) and Southern blot analysis.

[0049] Methods using an amplification reaction can utilize a real-time detection amplification procedure. For example, the method can utilize molecular beacon technology (Tyagi S., *et al.*, *Nature Biotechnology*, 14: 303 (1996)) or Taqman™ technology (Holland, P. M., *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:7276 (1991)).

[0050] In addition, methyl light (Trinh BN, Long TI, Laird PW. DNA methylation analysis by MethylLight technology, *Methods*, 25(4):456-62 (2001), incorporated herein in its entirety by reference), Methyl Heavy (Epigenomics, Berlin, Germany), or SNuPE (single nucleotide primer extension) (See e.g., Watson D., *et al.*, *Genet Res.* 75(3):269-74 (2000)) can be used in the methods of the present invention related to identifying altered methylation of the genes or regulatory regions provided herein. Additionally, methyl light, methyl heavy, and array-based methylation analysis can be performed, by using bisulfite treated DNA that is then PCR-amplified, against microarrays of oligonucleotide target sequences with the various forms corresponding to unmethylated and methylated DNA.

[0051] The degree of methylation in the DNA associated with the gene or genes or regulatory regions thereof, may be measured by fluorescent *in situ* hybridization (FISH) by means of probes which identify and differentiate between genomic DNAs, which exhibit different degrees of DNA methylation. FISH is described in the Human chromosomes: principles and techniques (Editors, Ram S. Verma, Arvind Babu Verma, Ram S.) 2nd ed., New York: McGraw-Hill, 1995, and de Capoa A., Di Leandro M., Grappelli C., Menendez F., Poggesi I., Giancotti P., Marotta, M. R., Spano A., Rocchi M., Archidiacono N., Niveleau A. Computer-assisted analysis of methylation status of individual interphase nuclei in human cultured cells. *Cytometry*. 31:85-92, 1998 which is incorporated herein by reference. In this case, the biological sample will typically be any which contains sufficient whole cells or nuclei to perform short term culture. Usually, the sample will be a tissue sample that contains 10 to 10,000, or, for example, 100 to 10,000, whole somatic cells.

[0052] In other embodiments, methylation-sensitive restriction endonucleases can be used to detect methylated CpG dinucleotide motifs. Such endonucleases may either preferentially

cleave methylated recognition sites relative to non-methylated recognition sites or preferentially cleave non-methylated relative to methylated recognition sites. Examples of the former are Acc III, Ban I, BstN I, Msp I, and Xma I. Examples of the latter are Acc II, Ava I, BssH II, BstU I, Hpa II, and Not I. Alternatively, chemical reagents can be used which selectively modify either the methylated or non-methylated form of CpG dinucleotide motifs.

[0053] In some embodiments, hypomethylation of the target gene is detected by detecting increased expression of the that gene. Expression of a gene can be assessed using any means known in the art. Typically expression is assessed and compared in test samples and control samples which may be normal, non-malignant cells. The test samples may contain cancer cells or pre-cancer cells or nucleic acids from them. Methods employing hybridization to nucleic acid probes can be employed for measuring specific mRNAs. Such methods include using nucleic acid probe arrays (microarray technology), *in situ* hybridization, and using Northern blots. Messenger RNA can also be assessed using amplification techniques, such as RT-PCR. Advances in genomic technologies now permit the simultaneous analysis of thousands of genes, although many are based on the same concept of specific probe-target hybridization. Sequencing-based methods are an alternative; these methods started with the use of expressed sequence tags (ESTs), and now include methods based on short tags, such as serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS). Differential display techniques provide yet another means of analyzing gene expression; this family of techniques is based on random amplification of cDNA fragments generated by restriction digestion, and bands that differ between two tissues identify cDNAs of interest. Moreover, specific proteins can be assessed using any convenient method including immunoassays and immuno-cytochemistry but are not limited to that. Most such methods will employ antibodies which are specific for the particular protein or protein fragments. The sequences of the mRNA (cDNA) and proteins of the target genes of the present invention are known in the art and publicly available.

[0054] As used herein, the term "selective hybridization" or "selectively hybridize" refers to hybridization under moderately stringent or highly stringent physiological conditions, which can distinguish related nucleotide sequences from unrelated nucleotide sequences.

[0055] As known in the art, in nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids

being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (for example, relative GC:AT content), and nucleic acid type, *i.e.*, whether the oligonucleotide or the target nucleic acid sequence is DNA or RNA, can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter. Methods for selecting appropriate stringency conditions can be determined empirically or estimated using various formulas, and are well known in the art (see, for example, Sambrook et al., *supra*, 1989).

[0056] An example of progressively higher stringency conditions is as follows: 2X SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2X SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2X SSC/0.1% SDS at about 42°C (moderate stringency conditions); and 0.1X SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, for example, high stringency conditions, or each of the conditions can be used, for example, for 10 to 15 minutes each, in the order listed above, repeating any or all of the steps listed.

[0057] The term "nucleic acid molecule" is used broadly herein to mean a sequence of deoxyribonucleotides or ribonucleotides that are linked together by a phosphodiester bond. As such, the term "nucleic acid molecule" is meant to include DNA and RNA, which can be single stranded or double stranded, as well as DNA/RNA hybrids. Furthermore, the term "nucleic acid molecule" as used herein includes naturally occurring nucleic acid molecules, which can be isolated from a cell, for example, a particular gene of interest, as well as synthetic molecules, which can be prepared, for example, by methods of chemical synthesis or by enzymatic methods such as by the polymerase chain reaction (PCR), and, in various embodiments, can contain nucleotide analogs or a backbone bond other than a phosphodiester bond.

[0058] The terms "polynucleotide" and "oligonucleotide" also are used herein to refer to nucleic acid molecules. Although no specific distinction from each other or from "nucleic acid molecule" is intended by the use of these terms, the term "polynucleotide" is used generally in reference to a nucleic acid molecule that encodes a polypeptide, or a peptide portion thereof, whereas the term "oligonucleotide" is used generally in reference to a nucleotide sequence useful as a probe, a PCR primer, an antisense molecule, or the like. Of

course, it will be recognized that an "oligonucleotide" also can encode a peptide. As such, the different terms are used primarily for convenience of discussion.

[0059] A polynucleotide or oligonucleotide comprising naturally occurring nucleotides and phosphodiester bonds can be chemically synthesized or can be produced using recombinant DNA methods, using an appropriate polynucleotide as a template. In comparison, a polynucleotide comprising nucleotide analogs or covalent bonds other than phosphodiester bonds generally will be chemically synthesized, although an enzyme such as T7 polymerase can incorporate certain types of nucleotide analogs into a polynucleotide and, therefore, can be used to produce such a polynucleotide recombinantly from an appropriate template.

[0060] In yet another embodiment of the invention, there are provided methods of determining the prognosis of a subject having a cell proliferative disorder. The method includes determining the methylation state of at least one regulatory region of a gene in a nucleic acid sample from the subject, wherein hypomethylation as compared to a corresponding normal cell in the subject or a subject not having the disorder, is indicative of a poor prognosis.

[0061] In still another embodiment of the present invention, there are provided methods of ameliorating a cell proliferative disorder in a subject in need thereof. The method includes administering to the subject an agent that methylates at least one regulatory region in a gene that is demethylated as compared to a subject not having the disorder, thereby reducing expression of the at least one gene and ameliorating the cell proliferative disorder.

[0062] Methylating agents are known in the art and include, for example, alkylating agents such as nitrosureas, triazenes, and imidazotetrzines. In particular embodiments, the methylating agent is delivered locally to a tumor site or systemically by targeted drug delivery.

[0063] Agents that methylate the demethylated gene can be contacted with cells in vitro or in vivo for the purpose of restoring normal gene expression to the cell. Efficacy of the treatment can be assessed by detecting increased expression or methylation of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9,

MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP.

[0064] In a further embodiment of the present invention, there are provided methods of identifying a gene activated by hypomethylation. The method includes comparing an expression analysis of a cell treated with an agent that reduces methylation to an expression analysis of a control cell not treated with the agent, wherein an increase in expression of a gene is indicative of a gene activated by demethylation. In one aspect, the cell is from a minimally transformed cell line. In some embodiments, the method may further include an expression analysis of a tissue sample and a tumor sample from the same tissue of origin as the treated cell, wherein an increase in expression of a gene in a tumor sample as compared to a normal sample is correlated to the genes activated by demethylation in the treated cell. The method may also include sequence analysis to identify CpG dinucleotide motifs in the regulatory region, or particularly the promoter of identified genes. Determination of the methylation status of the identified genes in tumor and corresponding normal tissue samples may also be included.

[0065] The demethylating agent can be a methyltransferase inhibitor such as 5-aza-2'-deoxycytidine (DAC). In one aspect, the histone deacetylase inhibitor trichostatin (TSA) is used to treat cells for further determination of the methylation status. In yet another aspect, a combination of 5 aza 2' deoxycytidine (5Aza-dC) and trichostatin (TSA) is utilized.

[0066] The data provided herein suggest that multiple solid tumor types undergo activation of candidate proto-oncogenes with associated demethylation in a coordinated fashion in individual tumors. Transformation-associated effects of BORIS expressed ectopically in BORIS-negative cell lines as well as growth effects with individual target genes that have been shown to be epigenetically activated and expressed by BORIS are demonstrated herein. However, this does not rule out the contribution of as yet unidentified genes to BORIS related effects or a cooperative effect between identified target genes. Cancer testes antigens including four of the genes identified herein, *MAGE A2*, *A3*, *A4*, *A11*, are part of the melanoma antigen family A (MAGE-A) family of genes initially discovered as targets for immunotherapy due to their near exclusive tumor-specific expression, but the MAGE-A family plays a functional role in cancer development. MAGEA2 binds to p53-responsive promoters and leads to assembly of a p53/MAGEA2/HDAC3 protein complex,

resulting in transcriptional silencing of genes ordinarily activated by p53 because of histone deacetylation. Additionally, different MAGE-A family members can repress downstream targets of p53, and studies have also linked MAGE-A family overexpression to chemoresistance, and MAGE family members have been shown to increase cell growth and inactivate TSG activity. Recently, MAGEA has been shown to repress p53-dependent apoptosis, and has been associated with resistance to taxanes and alkylating agents in gastric cancer.

[0067] As provided in Example 1, expression of the MAGE-A family and expression of *H19* appeared to be significantly related in primary tumors, supported by data indicating that these targets are controlled by common methylation-specific transcription factors. *H19* forms half of the best-studied example of imprinted-gene regulation, the *IGF2/H19* locus. *IGF2* (insulin-like growth factor 2) is expressed uniquely from the parental allele achieved by monoallelic methylation of the imprinting control region (ICR) at 11p15.5. Aberrant hypomethylation at this locus is one cause of Silver-Russell syndrome—a disease of asymmetry or hemihypertrophy associated with increased risk of malignancies including craniopharyngioma, testicular seminoma, hepatocellular carcinoma, and Wilms tumor. Additionally, several cases of familial Beckwith-Wiedemann syndrome (BWS), with and without Wilms' tumors, have been shown to be caused by microdeletions of the methylation-specific CTCF binding sites in the *H19* ICR, a rare familial cancer syndrome linked to epigenetics.

[0068] Other identified proto-oncogenes provided herein have been implicated recently in tumorigenesis. *TKTL1* protein expression is correlated to worse patient outcome in patients with invasive colon and urothelial tumors, and investigators hypothesize that enhanced *TKTL1* expression in tumors increases oxygen-independent glucose usage (Krockenberger et al., Int J Gynecol Cancer 17:101-6, 2007). In addition, over-expression of *TKTL1* has since been validated as a potential biomarker and treatment target in breast cancer (Foldi et al, Oncol Rep 17:841-5, 2007). *GPR17* and *GRIN1* have not been implicated in carcinogenesis to date. Although growth promoting effects of *C19ORF28* have not been demonstrated, this does not exclude the possibility that overexpression of this and any of our other targets may contribute to a malignant phenotype via other mechanisms—e.g., motility, invasion,

angiogenesis, or apoptosis resistance—or that it may cooperate with other identified targets to produce phenotypic effects.

[0069] The epigenetic reactivation of *TKTL1*, *H19*, *MAGEA2*, *MAGEA3*, *MAGEA4*, *MAGEA11*, *GPR17*, *GRIN1*, and *C19ORF28*, genes located at diverse chromosomal loci, appears to occur simultaneously in individual primary tumors from multiple tumor types. This concurrent genome-wide, promoter-specific hypomethylation that results in derepression of many potential oncogenes raises the possibility of a demethylator phenotype analogous to the CpG island methylator phenotype (CIMP) initially noted in colon cancer. Many proto-oncogenes are members of the cancer testes antigen family which are ordinarily repressed via epigenetic mechanisms during development. While not wishing to be bound by any particular theory, an attractive hypothesis is that this phenomenon represents the coordinated, but pathologic reversal of developmental epigenetic regulatory patterns in cancer cells. The validity of whole-genome integrative approach to screening for epigenetically-activated genes associated with malignancy provided herein is confirmed by the appearance of *H19* and the MAGE-A family members which have been reported to be controlled by epigenetic activation and show silencing in normal cells. Two separate groups among the nine genes identified in Example 1 showed statistically significant correlations for patterns of expression: 1) MAGEA family members with *H19* and 2) *TKTL1*, *GPR17*, and *GRIN*. These groups were also identified according to promoter homology, implicating the participation of promoter-specific binding activity in the coordinated expression of each of these groups and suggesting the existence of additional common transcriptional activators that recognize the specific demethylated promoter sequences of these genes. The strict correlation of BORIS expression with aberrant expression of multiple growth-promoting proto-oncogenes in a variety of solid tumors reinforces the postulated role for BORIS as a key participant in aberrant demethylation and transcriptional activation of putative oncogenes. This concept is supported by cell line experiments demonstrating that BORIS expression by itself is sufficient to simultaneously demethylate and activate the transcription of these genes. It is of great interest to define the factors with which BORIS cooperates to induce these epigenetic and expression changes. Recently, a role for BORIS in histone demethylation and chromatin remodeling has been demonstrated. Moreover, regardless of mechanism, the data provide strong evidence for consideration of BORIS as a dominant controlling factor for facilitating epigenetic alterations associated with coordinated demethylation and reactivation of target

genes that are of high value as potential therapeutic and diagnostic targets for NSCLC, HNSCC, and other tumors.

[0070] This simultaneous reactivation of multiple targets provides a significant challenge to the understanding of the collective, and perhaps cooperative, effects of this phenomenon in cell transformation. In particular, single targets may depend on concurrent activation of, and interaction with, other family members for oncogenic effect. Other investigators have found some evidence of coordination of cancer testes antigen family expression and the possibility of direct interactions. In addition, only the top 26/106 possible targets identified after integrative analysis in a single solid tumor type were selected for further analysis. Future studies of the remaining genes, as well as use of normal cell lines and tumors derived from other tissues in an integrative approach, would be expected to allow for discovery of additional, novel, epigenetically-controlled genes that may also act collaboratively to induce malignant transformation.

[0071] Due to lack of primary tumor data on a larger array platform a nonintegrative approach was also used, which resulted in ultimate validation of 4.4% of the targets (2/45) compared to the integrative results that produced a 27% hit rate (7/26), reflecting a higher ability to validate targets in primary tumor when these data are included in initial discovery strategies. Additional analysis of other targets that are significantly differentially regulated may also yield additional epigenetically derepressed targets. Finally, these data have therapeutic implications for demethylation therapy and targeting of therapy. The active investigation of pharmacologic demethylating agents as therapy for malignancy based on reversal of silencing of tumor suppressor genes may have unintended effects. It is possible that in certain tissues this may result in reactivation of developmentally repressed proto-oncogene targets, with the unintended effect of promoting late, second primary tumors. However, modulation of a pathway that involves the coordinated derepression of a series of growth-promoting protooncogene candidates and a key transcriptional effector, BORIS, may provide a significant opportunity for directed therapeutic intervention that simultaneously targets multiple oncogenic pathways.

[0072] In the study provided in Example 2, an integrative epigenetic screening approach was used to specifically identify coordinately expressed genes in human NSCLC whose transcription is driven by promoter demethylation. From the over 47,000 transcripts

incorporated in the Affymetrix Human Genome U133 Plus 2.0 expression platform, 10 genes were identified that showed both differential overexpression and promoter region hypomethylation in NSCLC. Surprisingly, 6 of the 10 genes discovered via this approach were known CT-X antigens, *MAGEA3*, *MAGEA12*, *MAGEA4*, *MAGEA1*, *MAGEA5* and *NY-ESO-1*. Four additional CTAs, *MAGEA9*, *MAGEA6*, *MAGEB2* and *CT45-2*, were within the top 20 on the rank list provided herein; however, these genes did not meet the screening selection criteria due to failure to show complete methylation of promoter regions in a separate cohort of normal lung tissue by bisulfite sequencing (Figure 17). It is possible that, by using less stringent selection criteria, these and other genes would be identified as differentially expressed, albeit with incomplete promoter methylation patterns in normal lung tissue. In this study, only the top 55 of the 290 possible targets identified after integrative analysis in a single solid tumor type were selected for further analysis. It is expected that further investigation of the remaining genes, as well as the use of normal cell lines and tumors derived from additional tissue types in an integrative approach, will allow for discovery of additional, novel, epigenetically-controlled genes that may also show coordinated expression in tumors and serve as possible targets for screening and immunogenic therapy.

[0073] Although some of the CTAs identified using the integrative technique provided have previously been shown to be expressed to some degree in NSCLC; the demonstration of a high degree of coordinated expression in a large sample set related to epigenetic unmasking is a new finding. In a previous study of 19 lung carcinoma cell lines expressing various MAGEA family members, there was nearly complete concordance between the RT-PCR and IHC results. Thus the use of quantitative RT-PCR is a valid method for detecting CT antigen expression, especially when dealing with primary tissue where it is usually not possible to isolate sufficient quantities of protein for analysis. In addition, this same study showed that 44% of the 187 NSCLC samples tested on tissue microarrays stained positive, to some degree, for MAGEA family expression, supporting the fact that CTAs are expressed at the protein level in NSCLC.

[0074] Four target genes showed a significant positive correlation between mRNA expression and promoter hypomethylation, *MAGEA12*, *MAGEA4*, *SBSN* and *NY-ESO-1* (Figure 16). *TKTL1*, *MAGEA5* and *MAGEA3* also showed a positive correlation between

demethylation and expression, but missed significance. The lack of correlation between expression and hypomethylation in some of the target genes is expected given the fact that multiple other mechanisms such as point mutations, insertions, deletions and loss of heterozygosity could be involved in gene expression regulation in NSCLC. Alternatively, a larger sample size may facilitate the ability to define a closer association between promoter methylation status and expression in these genes.

[0075] In addition to the 6 mentioned CT-X antigens, elucidated 4 additional target genes were elucidated that were showed to be coordinately expressed with the known CTAs and demethylated in tumors. Interestingly, three of these 4 genes are encoded by the X chromosome, *TKTL-1*, *ZNF-711* and *G6PD*. *TKTL1* has been correlated with worse outcomes in patients with invasive colon and urothelial tumors, oxygen-independent glucose usage and validated as a potential biomarker in breast cancer. *SBSN*, *ZNF-711* and *G6PD* have not previously been associated with tumor specific expression or carcinogenesis.

[0076] CTAs are attractive targets for tumor immunotherapy because of their restricted expression patterns in normal human tissue. Currently, demethylating agents and HDAC inhibitors are being studied as adjuvant treatment options for NSCLC and other human malignancies, and combinations of these drugs continue to undergo bench-top and clinical investigation. These epigenetic therapies are being utilized based on data that suggests that methylation of tumor suppressor genes plays a fundamental role in tumor formation, progression, and recurrence after resection. Promoter region methylation of certain genes in resected NSCLC specimens was recently shown to be associated with recurrence of the tumor and poorer patient outcomes. An additional study has previously shown that *NY-ESO-1* and *MAGEA3* are upregulated in a proportion of patients treated with 5-aza-2'-deoxycytidine in cancers involving the lung, esophagus, or pleura. With the data herein suggesting that multiple CTAs are coordinately expressed in NSCLC and demethylation coordinately upregulates multiple known CTAs and associated genes from the target list, combining the use of demethylating agents with immunotherapy targeted against these genes that might be derepressed after treatment with 5-AZA and other demethylating agents may be useful. Targeting multiple CTAs that are coordinately expressed would help to improve the efficacy seen with monovalent immunologic agents.

[0077] The function of these genes expressed uniquely in NSCLC has not been well explored. There are data that indicate that MAGEA family members have growth promoting effects, and CTA members have been associated with biologic pathways that support a malignant phenotype. Additional analyses of the genes that are aberrantly expressed via promoter demethylation in NSCLC would be expected to demonstrate functional effects that contribute to carcinogenesis.

[0078] Using an integrative analysis combining pharmacologic demethylation and previously published primary tissue array data, a common epigenetic mechanism for the coordinated expression of CTAs and additional targets that may serve as targets for immunotherapy have been defined. Accordingly, in a further embodiment of the present invention, the integrative epigenetic analysis provided herein may be used to identify antigens expressed in cancer cells, in particular groups of antigens that are coordinately expressed in cancer. Such antigens may be used as targets for anti-cancer immunotherapy. Further, antigens identified as having a common epigenetic mechanism for coordinated expression may be targeted in combination in immunotherapy, including antigens containing CpG islands.

[0079] In another embodiment of the present invention, there are provided methods for determining whether a subject is responsive to a particular therapeutic regimen. The method includes determining the methylation status of one or more genes or regulatory regions thereof, selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, and ZNF711, wherein hypomethylation of the gene or regulatory region thereof as compared with a normal subject is indicative of a subject who is responsive to the therapeutic regimen.

[0080] In certain embodiments, therapeutic regimen is administration of a chemotherapeutic agent. While not wanting to be limiting, chemotherapeutic agents include antimetabolites, such as methotrexate, DNA cross-linking agents, such as cisplatin/carboplatin; alkylating agents, such as canbusil; topoisomerase I inhibitors such as dactinomycin; microtubule inhibitors such as taxol (paclitaxol), and the like. Other chemotherapeutic agents include, for example, a vinca alkaloid, mitomycin-type antibiotic, bleomycin-type antibiotic, antifolate, colchicine, demecoline, etoposide, taxane,

anthracycline antibiotic, doxorubicin, daunorubicin, carminomycin, epirubicin, idarubicin, mithoxanthrone, 4-dimethoxy-daunomycin, 11-deoxydaunorubicin, 13-deoxydaunorubicin, adriamycin-14-benzoate, adriamycin-14-octanoate, adriamycin-14-naphthaleneacetate, amsacrine, carmustine, cyclophosphamide, cytarabine, etoposide, lovastatin, melphalan, topotecan, oxalaplatin, chlorambucil, methotrexate, lomustine, thioguanine, asparaginase, vinblastine, vindesine, tamoxifen, or mechlorethamine. While not wanting to be limiting, therapeutic antibodies include antibodies directed against the HER2 protein, such as trastuzumab; antibodies directed against growth factors or growth factor receptors, such as bevacizumab, which targets vascular endothelial growth factor, and OSI-774, which targets epidermal growth factor; antibodies targeting integrin receptors, such as Vitaxin (also known as MEDI-522), and the like. Classes of anticancer agents suitable for use in compositions and methods of the present invention include, but are not limited to: 1) alkaloids, including, microtubule inhibitors (e.g., Vincristine, Vinblastine, and Vindesine, etc.), microtubule stabilizers (e.g., Paclitaxel [Taxol], and Docetaxel, Taxotere, etc.), and chromatin function inhibitors, including, topoisomerase inhibitors, such as, epipodophyllotoxins (e.g., Etoposide [VP-16], and Teniposide [VM-26], etc.), and agents that target topoisomerase I (e.g., Camptothecin and Isirinotecan [CPT-11], etc.); 2) covalent DNA-binding agents [alkylating agents], including, nitrogen mustards (e.g., Mechlorethamine, Chlorambucil, Cyclophosphamide, Ifosfamide, and Busulfan [Myleran], etc.), nitrosoureas (e.g., Carmustine, Lomustine, and Semustine, etc.), and other alkylating agents (e.g., Dacarbazine, Hydroxymethylmelamine, Thiotepea, and Mitocycin, etc.); 3) noncovalent DNA-binding agents [antitumor antibiotics], including, nucleic acid inhibitors (e.g., Dactinomycin [Actinomycin D], etc.), anthracyclines (e.g., Daunorubicin [Daunomycin, and Cerubidine], Doxorubicin [Adriamycin], and Idarubicin [Idamycin], etc.), anthracenediones (e.g., anthracycline analogues, such as, [Mitoxantrone], etc.), bleomycins (Blenoxane), etc., and plicamycin (Mithramycin), etc.; 4) antimetabolites, including, antifolates (e.g., Methotrexate, Folex, and Mexate, etc.), purine antimetabolites (e.g., 6-Mercaptopurine [6-MP, Purinethol], 6-Thioguanine [6-TG], Azathioprine, Acyclovir, Ganciclovir, Chlorodeoxyadenosine, 2-Chlorodeoxyadenosine [CdA], and 2'-Deoxycoformycin [Pentostatin], etc.), pyrimidine antagonists (e.g., fluoropyrimidines [e.g., 5-fluorouracil (Adrucil), 5-fluorodeoxyuridine (FdUrd) (Floxuridine)] etc.), and cytosine arabinosides (e.g., Cytosar [ara-C] and Fludarabine, etc.); 5) enzymes, including, L-asparaginase; 6) hormones, including, glucocorticoids, such as, antiestrogens (e.g., Tamoxifen, etc.), nonsteroidal antiandrogens

(e.g., Flutamide, etc.), and aromatase inhibitors (e.g., anastrozole [Arimidex], etc.); 7) platinum compounds (e.g., Cisplatin and Carboplatin, etc.); 8) monoclonal antibodies conjugated with anticancer drugs, toxins, and/or radionuclides, etc.; 9) biological response modifiers (e.g., interferons [e.g., IFN- α , etc.] and interleukins [e.g., IL-2, etc.], etc.); 10) adoptive immunotherapy; 11) hematopoietic growth factors; 12) agents that induce tumor cell differentiation (e.g., all-trans-retinoic acid, etc.); 13) gene therapy techniques; 14) antisense therapy techniques; 15) tumor vaccines; 16) therapies directed against tumor metastases (e.g., Batimistat, etc.); and 17) inhibitors of angiogenesis.

[0081] In other embodiments, the therapeutic regimen is administration of a vaccine directed to a protein encoded by the hypomethylated gene. Vaccines may be directed to one or more of the target genes identified herein. For example, NY-ESO-1 and MAGEA3 are currently undergoing clinical trials in various human malignancies, including NSCLC (*see*, Hirschowitz, E. A., *et al.*, *J Thorac Oncol* 1: 93-104 (2006); Karanikas, V. *et al.*, *Cancer Biol Ther* 7 (2007); Raez, L. E., *et al.*, *Expert Opinion on Emerging Drugs* 11: 445-459 (2006); and Old, L. J., *Cancer Immun* 8: Suppl 1, 1 (2008)). In certain embodiments, vaccines are directed to multiple targets found to be coordinately overexpressed. Vaccines directed to the hypomethylated gene can be made by methods well-known in the art (*see e.g.*, Davis *et al.*, *Proc Natl Acad Sci U S A.* 101(29): 10697–10702, 2004).

[0082] The materials for use in the methods of the invention are ideally suited for the preparation of a kit. Such a kit may comprise a carrier device containing one or more containers such as vials, tubes, and the like, each of the containers comprising one of the separate elements to be used in the method. The kit may contain reagents, as described above for differentially modifying methylated and non-methylated cytosine residues. One of the containers may include a probe which is or can be detectably labeled. Such probe may be a nucleic acid sequence specific for a promoter region associated with a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP. For example, oligonucleotide probes of the invention can be included in a kit and used for detecting the presence of hypomethylated nucleic acid sequences in a sample containing a nucleic acid

sequence of the genes TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and/or KBGP. The kit may also include a container comprising a reporter, such as an enzymatic, fluorescent, or radionucleotide label to identify the detectably labeled oligonucleotide probe.

[0083] In certain embodiments, the kit utilizes nucleic acid amplification in detecting the target nucleic acid. In such embodiments, the kit will typically contain both a forward and a reverse primer for each target gene. Such oligonucleotide primers are based upon identification of the flanking regions contiguous with the target nucleotide sequence. Accordingly, the kit may contain primers useful to amplify and screen a promoter region of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP. If there is a sufficient region of complementarity, e.g., 12, 15, 18, or 20 nucleotides, then the primer may also contain additional nucleotide residues that do not interfere with hybridization but may be useful for other manipulations. For example, such other residues may be sites for restriction endonuclease cleavage, for ligand binding or for factor binding or linkers or repeats. The oligonucleotide primers may or may not be such that they are specific for modified methylated residues. The kit may optionally contain oligonucleotide probes. The probes may be specific for sequences containing modified methylated residues or for sequences containing non-methylated residues. The kit may optionally contain reagents for modifying methylated cytosine residues. The kit may also contain components for performing amplification, such as a DNA polymerase and deoxyribonucleotides. Means of detection may also be provided in the kit, including detectable labels on primers or probes. Kits may also contain reagents for detecting gene expression for one of the markers of the present invention. Such reagents may include probes, primers, or antibodies, for example. In the case of enzymes or ligands, substrates or binding partners may be used to assess the presence of the marker. In particular embodiments, the kit may include one or more primers or primer pairs selected from the sequences set forth in SEQ ID NOs: 1- 58.

[0084] The following examples are intended to illustrate but not limit the invention.

EXAMPLE 1

ACTIVATION OF GENES VIA PROMOTER DEMETHYLATION IN HEAD AND NECK CANCER AND LUNG CANCER

[0085] In this example, an integrative method was used to analyze expression in primary head and neck squamous cell carcinoma (HNSCC) and pharmacologically demethylated cell lines to identify aberrantly demethylated and expressed candidate proto-oncogenes and cancer testes antigens in HNSCC. HNSCC is useful as a solid tumor model system, due to the established role of epigenetic changes in its pathogenesis, as well as the availability of normal, minimally transformed cell lines for use in gene discovery strategies. Using pharmacologic demethylation in normal, minimally-transformed oral keratinocyte cell lines combined with Cancer Outlier Profile Analysis (COPA) in primary tissues as a discovery approach, a set of candidate proto-oncogenes that undergo aberrant demethylation and increased expression in primary human tumors were identified.

[0086] Functional data suggest that expression of these genes is associated with tumor promotion. Additional analyses demonstrated promoter homology and coordinated upregulation in individual tumors for subsets of these target genes (proto-oncogenes). Coordinated promoter demethylation and simultaneous transcriptional upregulation of proto-oncogene candidates with promoter homology, and phylogenetic footprinting of these promoters demonstrated potential recognition sites for the transcription factor BORIS were noted. Aberrant BORIS expression correlated with upregulation of candidate proto-oncogenes in multiple human malignancies including primary non-small cell lung cancers and HNSCC, induced coordinated proto-oncogene specific promoter demethylation and expression in non-tumorigenic cells, and transformed NIH3T3 cells. Coordinated, epigenetic unmasking of multiple genes with growth promoting activity occurred in aerodigestive cancers, and BORIS was implicated in the coordinated promoter demethylation and reactivation of epigenetically silenced genes in human cancers.

[0087] Histopathology. All samples were analyzed by the Pathology department at Johns Hopkins Hospital. Tissues were obtained via Johns Hopkins Institutional Review Board approved protocols under JHM IRB Protocol #92-07-21-01, "Detection of Genetic

Alterations in Head and Neck Tumors.” Normal samples were microdissected and DNA prepared from the mucosa. Tumor samples were confirmed to be head and neck squamous and subsequently microdissected to separate tumor from stromal elements to yield at least 80% tumor cells. Tissue DNA was extracted as described below.

[0088] 5Aza-dC and TSA Treatment of Cells. These *in vitro* techniques employed treatment of cultured cells with 5-aza-deoxycytidine (5Aza-dC, a cytosine analog which cannot be methylated) with or without Trichostatin A (TSA, a histone deacetylase inhibitor) and subsequent expression array analysis with validation of tumor suppressor gene targets. HNSCC cell lines were treated with 5Aza-dC and/or TSA as described previously. Briefly, cells were split to low density (1×10^6 cells/T-75 flask) 24 hours before treatment. Stock solutions of 5Aza-dC (Sigma, St. Louis, MO) and TSA (Sigma) were dissolved in DMSO (Sigma) and 100% ethanol, respectively. Cells were treated with 5 μ M 5-Aza-deoxycytidine for 5 days and 300 nmol/L TSA for the last 24 hours. Baseline expression was established by mock-treated cells with the same volume of DMSO or ethanol. Two normal oral keratinocyte cell lines (OKF6-Tert1 and OKF6-Tert1Q, immortalized with hTert), were treated in duplicate by 5-aza-deoxycytidine/trichostatin A.

[0089] Oligonucleotide microarray analysis and QRT-PCR analysis. Total cellular RNA was isolated using the RNEASY RNA isolation kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Oligonucleotide microarray analysis was carried out using the GENECHIP U133plus2 Affymetrix expression microarray (Affymetrix, Santa Clara, CA). Samples were converted to labeled, fragmented, cRNA per the manufacturer's protocol for use on the expression microarray. Signal intensity and statistical significance were established for each transcript using dChip version 2005. 2-fold increase based on the 90% confidence interval of the result and expression minus baseline >50 was used as the statistical cutoff value after 5Aza-dC and/or TSA treatment to identify upregulated candidate genes.

[0090] Public datasets. The public databases used in this study were the University of California Santa Cruz (UCSC) Human Genome reference sequence and the annotation database from the May 2004 freeze (hg17). 56 HNSCC expression microarrays were obtained from public datasets from Oncomine (Oncomine.org, Ann Arbor, Michigan). 14 expression microarrays that had been previously studied from the same platform were incorporated and all microarrays were normalized for COPA analysis. Also utilized were the expO datasets

(1185 tumors on the Affymetrix U133plus2 mRNA expression platform) available online as part of the Gene Expression Omnibus (GEO/NCBI). These data are publicly available online as part of the Gene Expression Omnibus (GEO/NCBI), produced by the International Genomics Consortium. This analysis utilized expression array data for 47,000+ genes measured in 1041 human tumors of various histologies.

[0091] Cancer outlier profile analysis (COPA). COPA was applied to a cohort of 68 tissues (49 tumors, 19 normals), with each gene expression data set containing 14,500 probe sets. Briefly, gene expression values were median centered, setting each gene's median expression value to zero. The median absolute deviation (MAD) was calculated and scaled to 1 by dividing each gene expression value by its MAD. Of note, median and MAD were used for transformation as opposed to mean and standard deviation so that outlier expression values did not unduly influence the distribution estimates, and were thus preserved post-normalization. Finally, the 75th, 90th, and 95th percentiles of the transformed expression values were calculated for each gene and then genes were rank-ordered by their percentile scores, providing a prioritized list of outlier profiles. For the purposes of a rank-list, the 90th percentile was chosen based on sample-size analysis (49 tumors, 19 normals). For details of the method refer to Tomlins et. al. (Science 310:644-648, 2005).

[0092] Integrative epigenetics. Target genes from the Affymetrix U133A mRNA expression microarray platform were ranked by COPA upregulation at the 90th percentile (from 49 tumors and 19 normal tissues). The U133A microarray platform (Affymetrix, Santa Clara California) has approximately 14,500 probe sets. A second rank list was produced by ranking genes in descending order of the degree of upfold regulation upon 5-aza/TSA treatment. These two sources of information (gene set demonstrating upregulation with 5-aza) and COPA score were combined by using a rank product. These two rankings were combined to rank all targets and permutation of the data was used to establish significance with a threshold of ≤ 0.005 . This resulted in 106 genes deemed significant. The top 26 of these targets were comprehensively evaluated. Presence of CpG islands in these genes was determined by MethPrimer. In order to not exclude genes outside the U133A platform, also considered were all other genes in the U133plus2 platform on the sole basis of 5-aza/TSA upfold regulation. For all genes which did not have tissue mRNA expression array information amenable to COPA analysis, only statistically significant reexpression after 5-aza

treatment was considered. 45 genes were studied that had an experimental versus baseline expression (E/B) > 2.0 , based on the 90% confidence interval and $E-B > 50$. All genes were then studied for the presence of CpG islands in promoters or the first intron. Initially, an *in silico* approach was used to confirm the presence of a CpG island using the UCSC genome browser which relies on GC content of $> 50\%$, > 200 bp, > 0.6 observed to expected CG's.

[0093] DNA extraction. Samples were centrifuged and digested in a solution of detergent (sodium dodecylsulfate) and proteinase K, for removal of proteins bound to the DNA. Samples were first purified and desalted with phenol/chloroform extraction. Digested sample was subjected twice to ethanol precipitation, and subsequently resuspended in 500 μ L of LoTE (EDTA 2.5 mmol/L and Tris-HCl 10 mmol/L) and stored at -80°C .

[0094] Bisulfite treatment. DNA from salivary rinses was subjected to bisulfite treatment, as described previously. In short, 2 μ g of genomic DNA was denatured in 0.2 M of NaOH for 30 minutes at 50°C . This denatured DNA was then diluted into 500 μ L of a solution of 10 mmol/L hydroquinone and 3 M sodium bisulfite. This was incubated for 3 hours at 70°C . After, the DNA sample was purified with a SEPHAROSE column (WIZARD DNA Clean-Up System; Promega, Madison, WI). Eluted DNA was treated with 0.3 M of NaOH for 10 minutes at room temperature, and precipitated with ethanol. This bisulfite-modified DNA was subsequently resuspended in 120 μ L of LoTE (EDTA 2.5 mmol/L and Tris-HCl 10 mmol/L) and stored at -80°C .

[0095] Bisulfite sequencing. Bisulfite sequence analysis was performed to check the methylation status in primary tumors and normal tissues, as well as in cell lines. Bisulfite-treated DNA was amplified using primers designed using the MethPrimer program (Li and Dahiya, *Bioinformatics* 18(11):1427-31, 2002) to span areas of CpG islands in the promoter or first intron. Primer sequences were designed to not have CG dinucleotides (see Table 1 below for primer sequences). The PCR products were gel-purified using the QIAQUICK gel extraction kit (Qiagen), according to the manufacturer's instructions. Each amplified DNA sample was applied with nested primers to the Applied Biosystems 3700 DNA analyzer using BD terminator dye (Applied Biosystems, Foster City, CA).

[0096] Quantitative unmethylation-specific PCR (QUMSP). To selectively amplify demethylated promoter regions in genes of interest, probe and primers were designed using

data from bisulfite sequencing of primary tumors which are complimentary only to bisulfite-converted sequences known to be demethylated in tumor (see Table 1). Primer combinations were validated using *in vitro* methylated and demethylated controls.

Table 1. QUMSP Primers

Target Gene	Forward Primers	SEQ ID NO:
MageA2B-QU	GTTGTGAATTTAGGGAAGTTATGG	1
MageA3-QU	ATTATTTTATTTTGTTGATTTTGTGT	3
MageA4-QU	TAAAAATTTGTTTTTTTATTGA	5
Magea11-QU	AGTAGATATTATTGGTAGAGGATGG	7
GPR17-QU	TTATTTTCGTGAGTAAGATTTTTCGG	9
GRIN1-QU	TTTTTTAAGTTAGGGAGGGGTATGT	11
H19-QU	TTTAGAGTTTGTGGTTAAGGTGG	13
C19ORF28-QU	TGTATATTGTTTGTGGGTATGA	15
TKTL1-QU	TGTGTAGAGAAAGAAGATTTTGTATTTGT	17
	Reverse Primers	
MageA2B-QU	CATCAAACCATTAACCTCAAAACAAA	2
MageA3-QU	CTCAAACTCACATACCTAACCAC	4
MageA4-QU	ATAATACAAAATTATTTAAAATCATC	6
Magea11-QU	CTTCCCTAAATTTACAACAAAAACG	8
GPR17-QU	GACTATCATAACAACAACGACGC	10
GRIN1-QU	AAACACCAAAAATCACTAAACACC	12
H19-QU	AACAAAAAATCATATAAAAAAACATA	14
C19ORF28-QU	CAAAATCTCAAAAAAACTACCCAAC	16
TKTL1-QU	CCCAATACATTTATATAATATCAAT	18

[0097] qRT-PCR. Total RNA was measured and adjusted to the same amount for each cell line, and then cDNA synthesis was performed using oligo-dt with the SUPERScript first-strand DNA synthesis kit (Invitrogen). The final cDNA products were used as the templates for subsequent PCR with primers designed specifically for each candidate gene (see Table 2). GAPDH was examined to ensure accurate relative quantitation in qRT-PCR. qRT-PCR heat maps were generated by median-normalization by gene, logged and heat maps generated using Excel.

Table 2. qRT-PCR Primers

Target Gene	Forward Primer	SEQ ID NO:	Reverse Primer	SEQ ID NO:
magea2	ATCTGCCTGTGGGTCTTCAT	19	GAAGAGGAAGAAGCGGTCTG	20
magea3	AAGATCTGCCAGTGGGTCTC	21	CCCCAGGGTGACTTCAACTA	22
magea4	AGGAGAAGATCTGCCTGTGG	23	CAGCCTCCTGCTCCTCAGTA	24
boris	cacattcgtacccacactgg	25	tcccctgatccacactctc	26
grin1	gtcctccaaagacacgagca	27	ccgtctgtctgtctgtctgc	28
tktl1	gatcactaccgcaagggtg	29	cacagggtaaagaggccaaa	30
h19	gagctctcaggaggaggat	31	agacctggcctcgtctcc	32
c19orf28	acccttgcccctcagagc	33	ggccacagcaggaggcta	34
gpr17	aactctcaggctctgactcca	35	gccagggattgccaactaa	36
magea11	gaggagaacaagtctgtgg	37	agcagcaggcaactcctcta	38

[0098] Transfection of human expression vectors. A full-length ORF cDNA of *MAGEA2B*, *MAGEA4*, *H19*, *TKTL1* were obtained for transient transfections. Cell lines were plated at 2×10^5 /well using 6-well plates and transfected with either empty vector or gene of interest using the FUGENE 6 transfection reagent (Roche, Basel, Switzerland) according to the manufacturer's protocol. Calcein fluorescence was measured by the Spectramax M2e 96-well fluorescence plate reader Molecular Devices (Sunnyvale, California). Live cells are distinguished by the presence of ubiquitous intracellular esterase activity, determined by the enzymatic conversion of the virtually nonfluorescent cell-permeable calcein AM to the intensely fluorescent calcein. The polyanionic calcein dye is well retained within live cells, producing an intense uniform green fluorescence (excitation/emission ~495 nm/515 nm).

[0099] Anchorage-independent growth assay. Soft agar assays were conducted after transfection of cells with mammalian expression vectors. Cells were counted and approximately 5000 were added into each 6-well plate. The bottom layer was composed of 0.5% agar, DMEM + 10% FBS, plus additives, while the cells were suspended in a top layer of 0.35% agar, DMEM + 10% FBS, plus additives. BORIS Inducible promoter constructs were incubated in the presence of low doxycycline (0.01 mg/mL). Soft agar assays were incubated at 37 °C for 2 weeks.

[0100] Statistical analysis. The QUMSP data was analyzed using a Wilcoxon-Mann-Whitney rank test. The p-values were corrected using the Benjamini-Hochberg procedure, and significance was defined as $p_{\text{corr}} < 0.05$. We looked for similarities in the methylation

patterns between genes by performing an analysis of correlations between QUMSP readings on the genes across all samples. 1000 permutations of the samples were used to establish significance, with $\alpha = 0.05$. For the expression data, the normalized data was log-transformed and correlation analysis was performed across all samples between each of the genes in the study. Significance was determined by assuming a normal distribution in the log-transformed expression levels and applying Student's t-distribution with an alpha of 0.05. All analyses were performed using Matlab. Comparisons of promoter homology were done with European Bioinformatics Institutes' ClustalW sequence alignment and phylogram software and the PromoterWise application.

[0101] Integrative Discovery of Epigenetically Unmasked Genes in HNSCC. We hypothesized that normal cell lines contain methylated genes that are typically repressed in normal tissues, but that these genes can be re-expressed by pharmacologic manipulation. A subset of these genes would include candidate proto-oncogenes activated by demethylation in human cancers that could be further selected on the basis of primary tumor expression array analysis using integrative methods. Methods of epigenetic screening using 5-aza/TSA treatment that have been found to be successful in defining candidate tumor suppressor genes were adapted for the present study. Two TERT-transformed normal oral keratinocyte cell lines were treated with 5 μ M 5-aza deoxycytidine for four days and Trichostatin A for one day prior to harvesting total RNA for expression array analysis using dChip.

[0102] Concurrently, a comparative epigenetic approach utilizing Cancer Outlier Profiling Analysis (COPA) was performed using 49 primary HNSCC and 19 normal mucosal tissues assayed for mRNA expression on the Affymetrix U133A mRNA expression microarray platform (16,383 probe sets) compiled from prior work and public sources of expression (oncomine.org). COPA is particularly useful to determine differences in expression for particular genes in subsets of primary tumor samples, with improved performance compared to statistical tools that rely on median or average expression difference between two datasets (Tomlins et al., Science 310:644-8, 2005). COPA was calculated at the 90th percentile for the final rankings of all 16,383 features of the arrays, as this resulted in the most pronounced differences in expression with this sample size. Statistical significance of the expression differences in the COPA diagrams was measured by Mann-Whitney U test (Figure 1B).

[0103] Gene ranks were determined in two ways: 1) COPA ranking at the 90th percentile of upregulation in primary tumor tissue versus normal tissue expression, and 2) fold upregulation after pharmacologic demethylation after dChip normalization in cell lines. An integrative rank product was calculated (Figure 1A). Using a significance threshold ($\alpha = 0.005$) and subsequent random permutation of the rank-lists, 106 genes were identified that were significantly differentially upregulated based on epigenetic screening and tissue microarray expression (Table 3). The top scoring 26 genes were selected for further analyses. Seventeen of 26 genes contained promoter-associated CpG islands as identified utilizing the MethPrimer software and were selected for further studies. In a separate parallel analysis to account for possible activated proto-oncogenes not included in the U133A platform, 32,500 genes were analyzed in the U133plus2 platform ranked on the sole basis of 5-aza/TSA fold upregulation in the normalized cell lines that were not included in primary tumor expression array analysis. 45 target genes were identified with > 2-fold upregulation at 90% confidence interval and an average difference value expression over baseline greater than 50. Among these, 30 were confirmed to have CpG islands (Table 4).

Table 3.

Gene	Accession	Desc	FoldChange	AzRank	OPAScore	SDDiff	OverallRank
melanoma antigen family A, 4	AW438674	gb:AW438674	13	7	39.62488	1	
melanoma antigen family A, 6	U10691	gb:U10691	14.55	168	4	57.21295	2
melanoma antigen family A, 3	BC000340	gb:BC000340	2.94	539	3	57.5791	3
melanoma antigen family A, 2	U82671	gb:U82671	1.91	2148	1	160.0283	4
chemokine (C-C motif) ligand 1	NM_004591	gb:NM_004591	27.75	7	318	5.489771	5
dehydrogenase/reductase (SD)	AK000345	gb:AK000345	58	3	1024	3.408703	6
oncostatin M	BG437034	gb:BG437034	9.99	71	46	14.6672	7
melanoma antigen family A, 1	BC003408	gb:BC003408	3.59	308	15	25.73548	8
glutamate receptor, ionotropic	NM_007327	gb:NM_007327	16.43	15	380	5.095238	9
interleukin 8	AF043337	gb:AF043337	26	88	117	8.772762	10
interleukin 6 (interferon, beta 2)	NM_000600	gb:NM_000600	3.33	86	138	8.163837	11
transketolase-like 1	Z49258	gb:Z49258	56.88	6	2001	2.503557	12
enolase 2 (gamma, neuronal)	NM_001975	gb:NM_001975	7.6	59	216	6.399704	13
fatty acid binding protein 4, ad	NM_001442	gb:NM_001442	5.4	107	128	8.483477	14
immunoglobulin heavy constant	U80139	gb:U80139	4.6	162	88	10.4507	15
ornithine decarboxylase 1	NM_002539	gb:NM_002539	6.8	155	96	9.695386	16
dehydrogenase/reductase (SD)	NM_005794	gb:NM_005794	20.44	11	1363	3.037964	17
chemokine (C-X-C motif) ligand	AF002985	gb:AF002985	2.73	671	23	21.87494	18
melanoma antigen family A, 1	BC004479	gb:BC004479	2.7	690	24	19.81325	19
lymphocyte-specific protein ty	NM_005356	gb:NM_005356	5.38	116	147	7.868405	20
cytokeratin type II	NM_004693	gb:NM_004693	4.55	166	111	8.994016	21
G protein-coupled receptor 17	NM_005291	gb:NM_005291	2.8	622	37	15.99263	22
cytoplasmic FMR1 interacting	AL161999	gb:AL161999	5.89	97	248	5.967384	23
matrix metalloproteinase 13 (c	NM_002427	gb:NM_002427	4.03	226	134	8.264698	24
nuclear receptor subfamily 4, c	D49728	gb:D49728	6.38	85	374	5.137292	25
chemokine (C-X-C motif) ligand	AF030514	gb:AF030514	2.26	1191	28	18.11945	26
dipeptidylpeptidase 4 (CD26, p	M80536	gb:M80536	2.46	914	40	15.33985	27
melanoma antigen family B, 2	NM_002364	gb:NM_002364	3.92	247	158	7.682364	28
early growth response 4	NM_001965	gb:NM_001965	2.26	1190	33	16.85125	29
keratin, hair, basic, 1	NM_002281	gb:NM_002281	25.69	8	5537	0.431888	30
baculoviral IAP repeat-containi	U37546	gb:U37546	6.66	77	596	4.27014	31
gb:BC006164.1 /DB_XREF=gi	BC006164	gb:BC006164	3.88	256	181	7.056745	32
apolipoprotein C-I	NM_001645	gb:NM_001645	3.9	251	193	6.75713	33
dickkopf homolog 1 (Xenopus	NM_012242	gb:NM_012242	3.93	543	95	9.819803	34
armadillo repeat containing, X-	NM_014782	gb:NM_014782	3.43	353	160	7.640024	35
bone marrow stromal cell anti	NM_004335	gb:NM_004335	6.67	76	767	3.858544	36
fatty acid binding protein 6, ile	U19869	gb:U19869	2.19	1333	54	13.53217	37
calbindin 1, 28kDa	NM_004929	gb:NM_004929	1.91	2136	34	16.70063	38
DnaJ (Hsp40) homolog, subfam	AV729634	gb:AV729634	4.63	160	511	4.530018	39
lysosomal associated multisp	NM_005762	gb:NM_005762	3.1	63	1432	2.978136	40
lysosomal associated multisp	A1589086	gb:A1589086	5.46	110	873	3.672105	41
early growth response 2 (Krox)	NM_000399	gb:NM_000399	4.27	194	497	4.604386	42
transglutaminase 2 (C polypep	AL031651	gb:AL031651	9.9	96	1026	3.407147	43
homeo box HB9	A1738662	gb:A1738662	11.84	27	3683	1.386694	44
interferon, alpha-inducible prot	NM_005532	gb:NM_005532	6.13	89	1174	3.210361	45
nuclear receptor subfamily 4, c	S77154	gb:S77154	2.81	612	180	7.059269	46
myosin, light polypeptide kina	NM_005965	gb:NM_005965	6.99	699	166	7.437435	47
nucleoporin 210kDa	A1867102	gb:A1867102	4.27	195	622	4.160394	48
interleukin 8	NM_000584	gb:NM_000584	4.22	209	592	4.279749	49
Full-length cDNA clone CS0D1	N30878	gb:N30878	4.3	951	131	8.327486	50
insulin-like 3 (Leydig cell)	A1991694	gb:A1991694	3.96	243	538	4.462851	51
chitinase 3-like 1 (cartilage gly	M80927	gb:M80927	3	1126	119	8.710771	52
pentraxin-related gene, rapidly	NM_002852	gb:NM_002852	2.12	1478	98	9.635762	53
chemokine (C-X-C motif) ligand	NM_002993	gb:NM_002993	2.03	1725	87	10.51205	54
carboxyl ester lipase (bile salt	NM_001807	gb:NM_001807	2.71	685	227	6.267059	55
oxidised low density lipoprotei	AF035776	gb:AF035776	2.2	1310	121	8.607828	56
nuclear receptor subfamily 4, c	NM_002135	gb:NM_002135	2.97	524	306	5.549491	57
ATP-binding cassette, sub-fan	U88667	gb:U88667	2.77	639	251	5.947201	58
CD74 antigen (invariant polype	K01144	gb:K01144	5.62	104	1549	2.8797	59
glycerol kinase	NM_000167	gb:NM_000167	3.38	371	448	4.783837	60
chorionic gonadotropin, beta p	NM_000737	gb:NM_000737	1.96	1956	89	10.42807	61
chemokine (C-X-C motif) ligand	NM_001511	gb:NM_001511	3.07	478	365	5.200393	62
neurofilament 3 (150kDa medi	NM_005382	gb:NM_005382	3.96	242	721	3.935926	63
matrix metalloproteinase 9 (ge	NM_004994	gb:NM_004994	3.62	299	619	4.167812	64

[0104] Validation of tumor specific promoter demethylation of target genes. CpG islands in the promoter region of the 47 selected gene targets with CpG islands were bisulfite sequenced in normal mucosal samples from patients without a cancer diagnosis to confirm epigenetic silencing in mature upper aerodigestive tract mucosa (Table 4). Only 18/47 promoter regions demonstrated complete methylation at all sequenced CpG sites in all normal tissues. These targets were subsequently bisulfite sequenced in 10 primary HNSCC to assay for the presence of hypomethylation. (Figure 2A). Of these remaining targets, 9/18 showed demethylation in tumor tissues in greater than 30% of the samples, including *TKTL1* (4/10, $p < 0.05$), *H19* (6/10, $p < 0.05$), *MAGEA2* (5/10, $p < 0.05$), *MAGEA3* (5/10, $p < 0.05$),

MAGEA4 (5/10, $p < 0.05$), *MAGEA11* (5/10, $p < 0.05$), *GPR17* (3/10, $p < 0.10$), *GRIN1* (6/10, $p < 0.05$), *C19ORF28* (5/10, $p < 0.05$), (chi-squared). To confirm transcriptional upregulation of target genes with 5-aza/TSA treatment in the cell line system, quantitative RT-PCR was performed on 5-aza/TSA-treated normal cells compared to mock-treated cells for these nine genes (Figure 1C). Each gene demonstrated significant upregulation by 5-aza/TSA treatment in at least one cell line supporting functional gene regulation by promoter hypomethylation. Using the initial cohort of 10 primary tumors, a preliminary analysis was performed to determine the relationship of promoter hypomethylation to expression. QRT-PCR expression with the bisulfite sequencing of the respective promoter below is shown in Figure 2B-J. The Mann Whitney U test was employed to compare QRT-PCR expression of the methylated and unmethylated groups. Three genes had statistically significant increased expression in the unmethylated group: *MAGEA2* ($p = 0.007$), *MAGEA3* ($p = 0.007$), *MAGEA11* ($p = 0.05$). Possible associations between expression and promoter methylation status in this small cohort were also suggested for *TKTL1* ($p = 0.06$), *MAGEA4* ($p = 0.09$), *C19ORF28* ($p = 0.09$), *GRIN1* ($p = 0.06$), yet *H19* ($p = 0.7$) but *GPR17* ($p = 0.38$) did not show this association.

Table 4.

Probe set	Accession	Name	CpG Island	Meth Nls	Meth Tumor	Fold Change
214183_s_at	X91817	transketolase-like 1	Y	Y	N	131.98
214023_x_at	AL533838	"tubulin	Y	N		15.2
227182_at	AW966474	sushi domain containing 3	Y	N		13.09
220779_at	NM_016233	"peptidyl arginine deiminase	N	N		10.88
231729_s_at	NM_004058	calcyphosine	N	N		10.82
204802_at	NM_004165	Ras-related associated with dia	Y	N		10.26
209742_s_at	AF020768	"myosin	N	N		9.61
219554_at	NM_016321	"Rhesus blood group	Y	N		9.36
201387_s_at	NM_004181	ubiquitin carboxyl-terminal este	Y	N		8.92
204803_s_at	NM_004165	Ras-related associated with dia	Y	N		8.54
224997_x_at	AL575306	"H19	Y	Y	N	7.67
1563357_at	AL049245	"Serine (or cysteine) proteinase	Y	Y	Y	7.66
214368_at	AI688812	RAS guanyl releasing protein 2	Y	N		7.03
223734_at	AF329088	ovary-specific acidic protein	Y	N		6.93
239430_at	AA195677	insulin growth factor-like family	N	N		6.75
222746_s_at	AJ276691	B-box and SPRY domain conta	Y	N		6.57
221909_at	AW299700	hypothetical protein FLJ14627	N	N		6.49
209723_at	BC002538	"serine (or cysteine) proteinase	Y	N		6.28
227711_at	BG150433	hypothetical protein FLJ32942	N	N		6.28
210130_s_at	AF096304	transmembrane 7 superfamily n	Y	N		6.19
241835_at	AI733297	LOC440570	N	N		5.98
219429_at	NM_024306	fatty acid 2-hydroxylase	Y	N		5.62
230675_at	BE671925	LOC441546	N	N		5.6
211573_x_at	M98478	"transglutaminase 2 (C polypep	Y	Y	Y	5.55
220921_at	NM_013453	"SPANX family	N	N		5.46
227468_at	AL565745	carnitine palmitoyltransferase 1	Y	N		5.34
219045_at	NM_019034	"ras homolog gene family	Y	N		5.05
202283_at	NM_002615	"serine (or cysteine) proteinase	N	N		5.05
227240_at	AV703769	neuronal guanine nucleotide ex	N	N		4.33
218918_at	NM_020379	"mannosidase	Y	N		4.29
1558216_at	BC043614	hypothetical protein LOC25484f	N	N		4.29
228132_at	AI240129	"actin binding LIM protein family	Y	N		4.16
227890_at	AL524643	similar to RIKEN cDNA A23007	Y	N		4.09
215813_s_at	S36219	prostaglandin-endoperoxide syr	Y	N		3.99
238669_at	BE613133	prostaglandin-endoperoxide syr	Y	N		3.93
222812_s_at	AF239923	"ras homolog gene family	Y	N		3.91
205127_at	NM_000962	prostaglandin-endoperoxide syr	Y	N		3.62
219529_at	NM_004669	chloride intracellular channel 3	N	N		3.48
1554539_a_at	BC018208	"ras homolog gene family	Y	Y	Y	3.45
223549_s_at	AL136880	espin	Y	N		3.4
226771_at	AB032963	"ATPase	N	N		3.32
224818_at	BE622952	sortilin 1	N	N		3.32
205508_at	NM_001037	"sodium channel	Y	N		3.16
224378_x_at	AF276658	microtubule-associated protein	Y	N		3.15
205691_at	NM_004209	synaptogyrin 3	Y	N		2.99
232164_s_at	AL137725	epiplakin 1	Y	N		2.73

[0105] Functional validation of candidate proto-oncogenes. Transient transfections were then performed to evaluate and/or confirm growth-promoting effects of these nine targets that showed transcriptional upregulation in primary tissue with concomitant tumor-specific promoter hypomethylation. Although H19 codes for a nontranslated RNA transcript, the H19 product appears to induce growth in lung and breast cancer cell lines (Barsyte-Lovejoy *et al.*, *Cancer Res* 66:5330-7, 2006) and may induce drug resistance in hepato-cellular carcinoma (Tsang and Kwok, *Oncogene* 26(33):4877-81, 2007). Figure 3A shows results obtained by transient transfection of an *H19* construct into OKF6-Tert-1R cells. At four days, there was a 41.4% ($\pm 15\%$) increase in growth over control cells transfected with empty vector. The MAGE family consists of related family members that are known to be upregulated in a variety of tumor types (Tsai *et al.* *Lung Cancer* 56(2):185-92, 2007), but have recently been

implicated in inducing transcriptional reprogramming in tumor cells (Laduron *et al.*, *Nucleic Acids Res* 32:4340-50, 2007). *MAGEA2* induced a 72.7% ($\pm$ 26%) increase in growth at day three (Figure 3B). *MAGEA4* transfection induced a 203% ($\pm$ 17%) increase in growth (Figure 3D). Functional growth differences were tested, but not found for *C19ORF28*.

[0106] In Figure 3C, *TKTL1* induced a 50.1% ($\pm$ 38%) increase in growth at day four. Enhanced expression of *TKTL1* has recently been implicated in the conversion of cells to aerobic, glycolytic metabolism as well as increased proliferation in colon cancer cells (Foldi *et al.*, *Oncol Rep* 17:841-845, 2007; Hu *et al.*, *Anticancer Drugs* 18:427-433, 2007; Krockenberger *et al.*, *Int J Gynecol Cancer* 17: 101-106, 2007; Langbein *et al.*, *Br J Cancer* 94:578-585, 2006; Staiger *et al.*, *Oncol Rep* 16: 657-661, 2006; and Zhang *et al.*, *Cancer Lett* 253(1):108-14, 2007). *TKTL1* is independently associated with poor survival in laryngeal carcinoma, colon and urothelial cancers, as well as distant metastasis in ovarian carcinoma (Volker *et al.*, *Eur Arch Otorhinolaryngol* 264z;1431-6, 2007). To further confirm *TKTL1* as a candidate proto-oncogene in HNSCC, adherent colony focus assays were performed in *TKTL1* low- expressing HNSCC cell lines JHU-011 and JHU-028, and found significant growth increase in both cell lines (Figure 4 A,B). shRNA constructs were then employed in a *TKTL1* high-expressing cell line UM-22B in anchorage independent growth assays, and a dramatic decrease in size and number of colonies (Figure 4 C,D) compared to mock transfected cells was noted.

[0107] Candidate proto-oncogenes are aberrantly expressed and promoter demethylated in multiple cancer type. To determine if candidate proto-oncogene expression was altered in a broader range of tumor types, expression data available through the expO datasets for 1041 human tumors of all histologies was analyzed. Data was first median-expression normalized by each array and subsequently by median normalization by probe set feature across the 1041 tumors from many cancer types including lung and urothelial, but not HNSCC. A subset of these tumors, non-small cell lung cancer (NSCLC), lymphoma, melanoma, pancreatic cancer, prostate cancers, and urothelial cancers, was chosen for presentation (Figure 5A-D). *H19* was significantly upregulated in NSCLC ($p = 0.008$) and in urothelial cancer ($p = 0.0013$), as calculated by Mann-Whitney U test comparing array-normalized expression in tumor type to all other tumors. Significantly increased expression of *MAGEA2* was noted in NSCLC ($p = 0.005$) but not in urothelial cancers ($p = 0.18$). *TKTL1* also showed overexpression in NSCLC

($p = 0.05$), but not urothelial cancer ($p = 0.55$), and *MAGEA4* was overexpressed in NSCLC ($p = 0.04$), but not significantly so in urothelial cancer ($p = 0.12$). In order to confirm target-specific demethylation noted in primary tumors, a rapid, quantitative assay for specifically measuring non-methylated promoters was devised, which was termed Quantitative Unmethylation-Specific PCR (QUMSP). Twenty-five HNSCC tumors and 11 upper aerodigestive mucosal samples were assayed for promoter demethylation (Figure 3E). Tumor-specific demethylation was found in *GRIN1* ($p = 0.005$), *MAGEA11* ($p = 0.001$), and *MAGEA2* ($p = 0.002$). A similar analysis was performed using a separate, independent cohort of 13 NSCLC samples with 14 lung samples from patients without neoplastic disease and confirmed promoter hypomethylation in target genes. Significant differences at < 0.05 in QUMSP were found in *H19* ($p = 0.02$), *MAGEA11* ($p = 0.03$), *MAGEA2* ($p = 0.005$), and *MAGEA3* ($p = 0.02$). See Figure 3F.

[0108] Aberrant expression of candidate proto-oncogenes occurs in a coordinated fashion in individual primary tumors. During these analyses, it was noted that transcriptional upregulation via promoter hypomethylation tended to occur synchronously in a subset of tumors. In the cohort of 49 primary HNSCC assayed via expression array analysis, a matrix of Pearson's correlation coefficients between the expression levels of each target was constructed (Figure 6A). For the nine target genes, significant clustering of increased expression was noted within the MAGEA family of genes. *H19* was not included because of its absence on the U133A platform. A separate cluster of associated overexpression was noted for *TKTL1*, *GRIN1*, and *GPR17*. From NSCLC expression data derived from the expO datasets we created similar matrices to examine correlations between individual genes. It was noted that MAGEA family expression and *H19* expression showed highly significant correlations in individual NSCLC (see Figure 6B). In contrast, there were no target-target correlations for NSCLC expression of the other cluster (*TKTL1*, *GRIN1*, and *GPR17*) that exhibited coordinated expression in HNSCC.

[0109] Expression patterns correlate with promoter homology for promoter demethylated target genes. The question as to whether promoter homology was associated with the linked expression of the two proto-oncogene clusters was then addressed. The European Bioinformatics Institutes' ClustalW tool (Figure 6C) for phylogram analysis was used after multiple sequence alignment of the respective promoters. To confirm homology

quantitatively, EMBL-EBI's PromoterWise comparison tool was used, which found significant pair-wise areas of promoter homology in *GPR17*, *GRIN1*, and *TKTL1*. As expected from earlier studies, the MAGE-A family clustered together, as the MAGE-A family members and H19 are known to have consensus-binding sites for methylation-sensitive binding factors *CTCF* and *CTCF*L/*BORIS*. In addition, this second group of *GRIN1*, *GPR17*, and *TKTL1* clustered together by sequence homology.

[0110] Finally, the question as to whether the degree of promoter hypomethylation was correlated in individual tumors was addressed. For both primary HNSCC (Figure 6D) and NSCLC (Figure 6E), multiple significant correlations between methylation status were found between targets, but methylation status did not cluster in groups defined by the MAGE-A family/H19 expression cluster or by the *TKTL1*, *GRIN1*, *GPR17* cluster. Rather, there were significant correlations between all identified candidate proto-oncogenes. Hypomethylation, therefore, appeared to occur in a related fashion in individual tumors for all target genes, but the concurrent expression of genes within the two clusters was associated with promoter homology rather than methylation status. This implied that specific transcriptional factors may be involved in the regulation of epigenetic unmasking and/or transcriptional activation based on promoter homology among these candidate proto-oncogenes.

[0111] *BORIS* expression is associated with proto-oncogene activation in primary tumors, induces promoter demethylation, candidate proto-oncogene expression, and cell transformation. The obvious presence of several MAGE genes among the identified targets prompted the study of upstream regulatory pathways of known cancer-testis antigens. *BORIS* and *CTCF* are a unique cognate pair of transcriptional factors involved in epigenetic regulation that share an identical DNA-binding domain. *BORIS* is transcriptionally silenced in most normal tissues, but expressed in normal embryonic, germ cell, and cancer tissues. Thus, it was determined if expression of *BORIS* correlated with candidate proto-oncogene expression in a separate cohort of 36 primary HNSCC. Figure 7A presents a heat map constructed from median normalized, qRT-PCR expression data of the proto-oncogenes, sorted by *BORIS* expression. In these 36 cancers, *BORIS* overexpression was significantly correlated to overexpression of 6/9 proto-oncogenes including: *MAGEA3* ($p = 0.0017$), *MAGEA4* ($p = 0.04$), *MAGEA11* ($p < 0.001$), *GPR17* ($p = 0.01$), and *C19ORF28* ($p = 0.001$). To further examine the correlation of *BORIS* expression with the target genes in solid

cancers, the expO dataset data for 1041 human tumors of a wide variety of tissue sources and histologies was analyzed. Significant positive correlation of BORIS expression with expression of each of the nine proto-oncogenes was noted: *GRIN1* ($p < 0.001$), *C19ORF28* ($p < 0.001$), *H19* ($p < 0.001$), *MAGEA11* ($p < 0.001$), *MAGEA2* ($p < 0.001$), *MAGEA3* ($p = 0.003$), *MAGE4* ($p < 0.001$), *TKTL1* ($p < 0.001$), *GPR17* ($p < 0.001$), (Suppl. Figure 2). Although BORIS transcripts are usually undetectable in normal cells, it was determined that 59% of all tumors have a BORIS level that exceeds the median expression of all genes, and 90% of tumors have a BORIS expression level $>25\%$ of median expression value for all genes, indicating that aberrant BORIS expression is a common event in human cancer.

[0112] To explore the functional and epigenetic effects of BORIS, tetracycline inducible pBIG-BORIS constructs were transiently transfected into NIH-3T3 and OKF6-Tert1R cell lines in the presence of doxycycline, resulting in increased adherent cell growth in wild type, BORIS non-expressing NIH3T3, and OKF6-Tert1R cell lines. 3T3 cells had a $77\% \pm 34\%$ growth increase at day three. OKF6 cell lines had a $161\% \pm 78\%$ growth increase at day three (Figure 7B). Importantly, these effects were seen when levels of *BORIS* expression was regulated to be similar to the levels found in primary tumors. This effect was not seen with increased concentrations of doxycycline that induced super-“physiologic” levels of *BORIS* transcripts. An analysis of transcripts showed that expression of seven of nine target genes was significantly increased in OKF6-Tert1R cell expressing BORIS (Figure 7E). To test if BORIS expressed at “physiologic” levels might contribute to transformation, NIH3T3 cells were studied for anchorage independent growth. After 12 days, significant numbers of colonies (30 ± 3) were observed in tests of BORIS-expressing cells but not in cells transfected with a control plasmid (Figure 7C).

[0113] Finally, to test the possibility that BORIS may be associated with epigenetic alterations as well as transcriptional upregulation of the identified target genes, methylation status of our candidate proto-oncogenes was quantitatively assayed after BORIS transfection and it was noted that six out of nine targets (*C19 ORF28*, *GPR17*, *GRIN1*, *MAGEA2*, *MAGEA3*, and *MAGE11*) showed a greater than 100% increase in demethylated promoter as early as 48 hours after induction of BORIS (Figure 7D).

EXAMPLE 2**INTEGRATIVE DISCOVERY OF EPIGENETICALLY DERERESSED
CANCER/TESTIS ANTIGENS IN NSCLC**

[0114] In this study, an integrative epigenetic screening approach was used to specifically identify coordinately expressed genes in human non-small cell lung cancer (NSCLC) whose transcription is driven by promoter demethylation. Our screen found 290 significant genes from the over 47,000 transcripts incorporated in the Affymetrix Human Genome U133 Plus 2.0 expression array. Of the top 55 candidates, 10 showed both differential overexpression and promoter region hypomethylation in NSCLC. Surprisingly, 6 of the 10 genes discovered by this approach were CTAs. Using a separate cohort of primary tumor and normal tissue, NSCLC promoter hypomethylation and increased expression by quantitative RT-PCR was validated for all 10 genes. Significant, coordinated coexpression of multiple target genes, as well as coordinated promoter demethylation, in a large set of individual tumors was also noted. These data suggested that epigenetic alterations are highly associated with coordinated CTA expression in NSCLC, and have significant implications for discovery of novel CTAs and CT antigen directed immunotherapy.

[0115] Histopathology. All samples were analyzed by the pathology department at Johns Hopkins Hospital. Tissues were obtained via Johns Hopkins Institutional Review Board approved protocol NA_00001911. Tumor and normal lung tissues from surgical specimens were frozen in liquid nitrogen immediately after surgical resection and stored at -80 °C until use. Normal samples were microdissected and DNA prepared from normal lung parenchyma. Tumor samples were confirmed to be NSCLC and subsequently microdissected to yield at least 80% tumor cells. Tissue DNA and RNA was extracted as described below.

[0116] 5Aza-dC and TSA treatment of cells. Normal human lung cell lines (NHBE and SAEC, Lonza, Walkersville, MD) were treated in triplicate with 5-aza-deoxycytidine (5Aza-dC, a cytosine analog which cannot be methylated) and trichostatin A (TSA, a histone deacetylase inhibitor). Briefly, cells were split to low density (2.5×10^5 cells and 6×10^5 /100mm dish for SAEC and NHBE, respectively) 24 hours before treatment. Stock solutions of 5Aza-dC (Sigma, St. Louis, MO) and TSA (Sigma) were dissolved in 50% acetic acid and 100% ethanol, respectively. Cells were treated with 5 uM 5Aza-deoxycytidine for 72 hours

and 300 nmol/L TSA for last 24 hours. Baseline expression was established by mock-treated cells with the same volume of acetic acid or ethanol in triplicate.

[0117] RNA extraction and oligonucleotide microarray analysis. Total cellular RNA was isolated using TRIZOL reagent (Life Technologies, Gaithersburg, MD) and the RNEASY RNA isolation kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Oligonucleotide microarray analysis was carried out using the GENECHIP U133 Plus 2.0 Affymetrix expression microarray (Affymetrix, Santa Clara, CA). Samples were converted to labeled, fragmented, cRNA per the Affymetrix protocol for use on the expression microarray. Signal intensity and statistical significance was established for each transcript using dChip version 2005 software to initially analyze and normalize the array data and then Significance Analysis of Microarrays (SAM). SAM output was calculated at a d-value of 1.126 yielding a false discovery rate and d-score cutoff of 5.065% and 1.885. This identified a total of 12,132 upregulated candidate genes after 5Aza-dC/TSA treatment.

[0118] Public datasets. The public databases used in this study were the University of California Santa Cruz (UCSC) Human Genome reference sequence and the annotation database from the March 2006 freeze (hg18). 40 normal lung and 111 NSCLC expression microarrays were obtained from the expO datasets (all performed on the Affymetrix U133 Plus 2.0 mRNA expression platform) available online as part of the Gene Expression Omnibus (GEO/NCBI). The microarrays from normal tissue and tumor were first normalized for COPA analysis using dChip version 2005.

[0119] Cancer outlier profile analysis (COPA). COPA was applied to a cohort of 151 tissues (111 tumors, 40 normals), with each gene expression data set containing 54,613 probe sets from the Affymetrix U133 Plus 2.0 mRNA expression platform. Briefly, gene expression values were median centered, setting each gene's median expression value to zero. The median absolute deviation (MAD) was calculated and scaled to 1 by dividing each gene expression value by its MAD. Of note, median and MAD were used for transformation as opposed to mean and standard deviation so that outlier expression values do not unduly influence the distribution estimates, and are thus preserved post-normalization. Finally, the 75th, 90th, and 95th percentiles of the transformed expression values were calculated for each gene and then genes were rank-ordered by their percentile scores, providing a prioritized list of outlier profiles. For the purposes of the rank-list, the 90th percentile for tumors was

chosen based on sample-size analysis (111 tumors, 40 normals). Normal tissue that had a 95th percentile >2 was eliminated from the rank list. A total of 35,764 transcripts met the above criteria and were ranked. For details of the method refer to Tomlins *et al.* (*Science* 310:644-8, 2005).

[0120] Integrative epigenetics. Target genes from the Affymetrix U133 Plus 2.0 mRNA expression platform by COPA upregulation were ranked at the 90th percentile (from 111 tumors and 40 normal tissues). The U133 Plus 2.0 mRNA expression platform (Affymetrix, Santa Clara California) has approximately 55,000 probe sets. A second rank list was produced by ranking genes in descending order of their d-score as computed by SAM following 5-aza/TSA treatment of normal lung cell lines (NHBE and SAEC). A third rank list was computed using 111 NSCLC and an additional expO dataset with 79 additional NSCLC primary tumor tissues also run on the Affymetrix Human Genome U133 Plus 2.0 mRNA expression platform. In these 190 primary NSCLC samples, BORIS expression patterns within each tumor was correlated with expression of all transcripts incorporated in the U133 Plus 2.0 array by calculating a correlation coefficient using Excel. All genes were then ranked based on the strength of the correlation between their expression and that of BORIS expression across all 190 samples. These three sources of information (gene set demonstrating upregulation with 5-aza, COPA score, and BORIS correlation) were combined by using a rank product ($x*y*z$). These three rankings were combined to rank all targets and permutation of the data was used to establish significance with a threshold of $\alpha = 0.005$, yielding 290 significant genes. Genomic sequences were obtained for 122 of these genes using the UCSC genome browser, and the presence of CpG islands in the promoters or first intron of these genes was determined by MethPrimer which relies on GC content of $> 50\%$, > 100 bp, > 0.6 observed to expected CG's.

[0121] DNA extraction. Samples were centrifuged and digested in a solution of detergent (sodium dodecylsulfate) and proteinase K, for removal of proteins bound to the DNA. DNA was purified by phenol-chloroform extraction and ethanol precipitation. The DNA was subsequently resuspended in 500 μ L of LoTE (EDTA 2.5 mmol/L and Tris-HCl 10 mmol/L) and stored at -80°C until use.

[0122] Bisulfite treatment and sequencing. 2 μ g of DNA from 28 NSCLCs and 11 normal lung tissues were subjected to bisulfite treatment using the EpiTect® Bisulfite Kit (Qiagen,

Valencia, CA) according to the manufacturer's instructions. This bisulfite-modified DNA was then stored at -80°C . Subsequently, bisulfite-treated DNA was amplified using primers designed by MethPrimer to span areas of CpG islands in the promoter or first intron (Li et al., Bioinformatics 18:1427-31, 2002). Primer sequences were designed to not contain CG dinucleotides (see Table 5 below for primer sequences). The PCR products were gel-purified using the QIAQUICK gel extraction kit (Qiagen, Valencia, CA), according to the manufacturer's instructions. Each amplified DNA sample was applied with nested primers to the Applied Biosystems 3700 DNA analyzer using BD terminator dye (Applied Biosystems, Foster City, CA).

[0123] Quantitative RT-PCR. Total RNA extracted as described above and the concentration for each sample was measured. 1 μg of RNA was then used for cDNA synthesis performed using oligo-dt with the SUPERScript First- Strand Synthesis kit (Invitrogen, Carlsbad, CA). The final cDNA products were used as the templates for subsequent RT-PCR with primers designed specifically for each candidate gene. 18s rRNA was examined to ensure accurate relative quantitation in quantitative RT-PCR. Each experiment was performed in triplicate using the TAQMAN 7900 (ABI) real-time PCR machine and the QUANTIFAST SYBR Green PCR Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions.

[0124] Quantitative unmethylation-specific PCR (QUMSP). To selectively amplify demethylated promoter regions in genes of interest, primers were designed using data from bisulfite sequencing of primary tumors which are complementary only to bisulfite-converted sequences known to be demethylated in tumors. Primer combinations were validated using *in vitro* methylated and demethylated controls. These experiments were performed in triplicate using the QUANTIFAST SYBR green PCR kit and the TAQMAN 7900 (ABI) real-time PCR machine with standard curves and normalization to beta-actin primers that do not contain CpG's in the sequence.

Table 5. QUMSP Primers

Primer Name	Sequence	SEQ ID NO:
quMAGEA3F	TTGTGAGATTTTGTGTTTGAGTAATG	39
quMAGEA3R	CAACCTAAAAATCTTCCCCTACAA	40
quMAGEA12F	TTTTGTTTGTGAGTAATGGTTTGATG	41
quMAGEA12R	CCCTCTATCTAAAATAAAAACCCACC	42
quMAGEA4F	ATTTATATTTTATTTAGGTAGGATTTTGT	43
quMAGEA4R	TCTCAAACCTACAAATCAAACACA	44
quMAGEA1F	TTTGGTTTTTGTGTTAGGAAATATTTG	45
quMAGEA1R	ACAAAACCTAAATCAAATTCCTTCA	46
quSBSN_F	TAAGATGTGATTGGTTTAGATGTTGA	47
quSBSN_R	CCTACAACCTACCATACCCACT	48
quTKTL1F	TAAATATTTTAGATTTTTTGATTGTTTGT	49
quTKTL1R	TAATATCAATACAACCTCCCTCCACA	50
quMAGEA5F	TGGTTTTGTTTTGTTTTTAATTGA	51
quMAGEA5R	AAATAACCCAAACTAAAAAATCCAC	52
quZNF711F	TTTTATTTTGTAGTAAGGGTAGTGTGAT	53
quZNF711R	CCCTTAAAATCCTCACATTCAA	54
quNYESO1F	GGTATTGTGGTTATTTTTTGGTTTG	55
quNYESO1R	CACACAAAAACCCCTACTTCCAAC	56
quG6PD_F	GTTTTGGAGAGAAGTTTGAGTTTGT	57
quG6PD_R	TCCAAAAAACAAAAAACATATAAACA	58

[0125] Statistical analysis. Similarities in the methylation patterns between genes were identified by performing an analysis of correlations between QUMSP readings on the genes across all samples. Spearman's correlation permutation testing was used with 1000 permutations of the samples to establish significance, with $\alpha = 0.05$. For the expression data, the normalized data was log-transformed and correlation analysis was performed across all samples between each of the genes in the study. Significance was determined by assuming a normal distribution in the log-transformed expression levels and applying Student's t-distribution with an alpha of 0.05. All analyses were performed using Matlab.

[0126] A novel integrative epigenetic approach to screen for CTAs and related epigenetically regulated genes. An integrative, high-throughput approach to screen for CTAs and other coordinately expressed genes was developed based on three key previously published factors: (1) CTAs are expressed in germline cells and many tumors, but not in normal somatic tissue, (2) CTAs have promoter CpG islands that are methylated and silenced in normal somatic tissue, but, experimentally, can be expressed by promoter demethylation

and (3) the transcription factor BORIS has been shown to induce de-repression of several CTAs in NSCLC and other tumor/tissue types (Figure 10A).

[0127] The first arm of the screening approach used herein involved the pharmacologic demethylation of 2 normal lung cell lines, Normal Human Bronchial Epithelial (NHBE) and Human Small Airway Epithelial (SAEC) cells (Lonza, Walkersville, MD), using a 5-aza/TSA treatment protocol that has previously been successful in defining candidate tumor suppressor genes by demethylating tumor cell lines. With the understanding that CTAs are silenced by methylation in normal tissue, normal cell lines were used to identify genes that are typically repressed in normal tissues, but can be re-expressed by pharmacologic manipulation. Two normal lung cell lines, NHBE and SAEC, were treated with 5 μ M 5-aza deoxycytidine for 72 hours and Trichostatin A for 24 hours prior to harvesting total RNA for expression array analysis using the Affymetrix Human Genome U133 Plus 2.0 expression platform. These results were then analyzed using dChip and Significance Analysis of Microarrays (SAM). Genes were ranked based on their SAM score(d). SAM also reported the fold change in the mean expression of the target genes in the 5-aza/TSA treatment group versus the control group (Figure 10B).

[0128] Data from 40 normal lung and 111 NSCLC expression microarrays from expO datasets (all run on the Affymetrix Human Genome U133 Plus 2.0 mRNA expression platform) publicly available online as part of the Gene Expression Omnibus (GEO/NCBI) were concurrently analyzed. For analysis of these 151 primary tissue expression array data sets, a technique known as Cancer Outlier Profile Analysis (COPA) was used. COPA is a method to search for marked overexpression of particular genes that occur only in a subset of cases, whereas traditional analytical methods based on standard statistical measures fail to find genes with this type of expression profile. COPA was a particularly useful method for the search for CTAs and genes with similar expression profiles based on previous studies showing that CTAs are heterogeneously expressed both across a wide patient population and within individual tumor specimens. Genes with a normal tissue COPA expression scaled score >2 at the 95th percentile were eliminated from the rank list. All remaining genes were then ranked based on their COPA score at the 90th percentile; statistical significance of the expression differences in the COPA diagrams were measured by Mann-Whitney U test (Figure 10C).

[0129] For the final arm of the screening approach provided herein, the previous data set with 111 NSCLC and an additional expO dataset with 79 additional NSCLC primary tumor tissues also run on the Affymetrix Human Genome U133 Plus 2.0 mRNA expression platform was used. In these 190 primary NSCLC samples, BORIS expression patterns within each tumor were correlated with expression of all transcripts incorporated in the U133 Plus 2.0 array by calculating a correlation coefficient using Excel. All genes were then ranked based on the strength of the correlation between their expression and that of BORIS expression across all 190 samples.

[0130] Three rank lists were produced by ranking genes by SAM score(d) following 5-aza/TSA treatment in normal lung cell lines, COPA score in primary tissue, and BORIS correlation in primary tissue. These 3 rank lists were combined by using a rank product ($x*y*z$). Using a significance threshold ($\alpha=0.005$) and subsequent random permutations of the rank-lists, 290 genes were identified that were significantly differentially upregulated based on epigenetic screening and tissue microarray expression patterns (Figure 17).

[0131] Initially, an *in silico* approach utilizing MethPrimer was used to confirm the presence of CpG islands in the promoter regions of our top candidates. The top 100 of the 290 significant genes as well as 22 genes selected based on biological relevance in cancer related pathways were selected to be screened via this approach, and 101 were found to contain 1 or more promoter CpG islands.

[0132] A separate cohort of 11 normal lung tissues from patients without a cancer diagnosis was then used to confirm epigenetic silencing via promoter methylation in normal lung mucosa from patients without a lung neoplasm. Bisulfite sequencing of CpG islands in the promoter regions of 55 selected gene targets with CpG islands was used to determine the methylation status. Only 17/55 promoter regions demonstrated complete methylation at all sequenced CpG sites in all or nearly all of the normal tissues (Figure 17). These targets were subsequently bisulfite sequenced in a separate cohort of 28 primary NSCLC to search for the presence of promoter hypomethylation. Of these remaining targets, 10/17 showed promoter demethylation in some fraction of tumors including: *MAGEA3* (13/28, $p=0.0067$), *MAGEA12* (19/28, $p=0.0001$), *MAGEA4* (9/27, $p=0.0378$), *MAGEA1* (21/27, $p=0.0001$), *SBSN* (13/28, $p=0.0067$), *TKTL1* (5/27, $p=0.2949$), *MAGEA5* (9/23, $p=0.0172$), *ZNF711* (21/24, $p=0.0001$),

NY-ESO-1 (14/20, $p=0.0002$), *G6PD* (17/18, $p=0.0014$), (Fisher's exact test) (Table 6 and Figure 11).

Table 6. Genes coordinately expressed and demethylated in NSCLC

Symbol	Description	Integrated Rank Position	COPA Score *	Upregulated with 5-Aza Average Fold Change (SAM Score (d))	Promoter CpG Island Present	Methylated in Normal Lung Tissue	Unmethylated in NSCLC Tumor Tissue
MAGEA3	Melanoma antigen family A, 3	4	60.6	1.8 (9.9)	Y	Y	Y
MAGEA12	Melanoma antigen family A, 12	5	24.1	4.8 (7.8)	Y	Y	Y
MAGEA4	Melanoma antigen family A, 4	6	101.0	14.9 (11.3)	Y	Y	Y
MAGEA1	Melanoma antigen family A, 1	11	12.8	6.5 (10.5)	Y	Y	Y
SBSN	Suprabasin	27	6.4	13.5 (31.3)	Y	Y	Y
TKTL1	Transketolase-like 1	35	1.8	3.5 (15.3)	Y	Y	Y
MAGEA5	Melanoma antigen family A, 5	41	16.9	1.5 (3)	Y	Y	Y
ZNF711	Zinc finger protein 6	43	10.0	3.8 (9.3)	Y	Y	Y
NY-ESO-1	Cancer/testis antigen 1B	72	117.6	5.9 (4)	Y	Y	Y
G6PD	Glucose-6-phosphate dehydrogenase	105	22.4	2.1 (10.6)	Y	Y	Y

[0133] Transcriptional upregulation of target genes after 5-aza/TSA treatment in the cell line system was confirmed using quantitative RT-PCR on the 5-aza/TSA-treated normal cells compared to mock-treated cells for these 10 genes (Figure 14). Each gene, with the exception *MAGEA12*, demonstrated significant upregulation by 5-aza/TSA treatment in at least one cell line supporting functional gene regulation by promoter hypomethylation.

[0134] CTAs and associated genes are coordinately demethylated and expression is correlated with promoter demethylation. In order to confirm the bisulfite sequencing results in the target genes and to provide a dataset of continuous variables to express the status of promoter demethylation, a rapid, quantitative assay for specifically measuring non-

methyated promoters was devised, which was termed Quantitative Unmethylation-Specific PCR (QUMSP). DNA extracted from the cohort of 28 primary NSCLC tumor samples and 11 normal lung samples from non-cancer patients was assayed (Figure 12A). Significant tumor-specific demethylation was found in *MAGEA3* ($p < 0.005$), *MAGEA12* ($p < 0.025$), *MAGEA4* ($p < 0.018$), *MAGEA1* ($p < 0.001$), *TKTL1* ($p < 0.025$) and *MAGEA5* ($p < 0.007$). Two additional targets slightly missed significance at $\alpha < 0.05$, *SBSN* ($p < 0.07$) and *NY-ESO-1* ($p < 0.09$) (2 tailed Student's t-test assuming unequal variance).

[0135] Given the tumor-specific demethylation pattern seen for these target genes, the question as to whether demethylation of the promoter regions of these genes occurred in a coordinated fashion within tumor samples was addressed. Spearman's correlation permutation testing was utilized to determine significant coordinated demethylation using the QUMSP results from the cohort of 28 NSCLC (Figure 12B). The p-value matrix of the Spearman's correlation coefficient showed that for any of the target genes, demethylation tended to coordinately occur with a minimum of 6 of the other genes. Shaded cells represent significant p-values. This offered evidence that demethylation is highly associated with coordinated regulation of these CTAs and related target genes in NSCLC, and strongly suggested an epigenetic mechanism of activation.

[0136] In order to confirm tumor specific expression of the target genes, quantitative RT-PCR was used to determine mRNA expression in the cohort of NSCLC and normal lung tissue (Figure 15A-J). Six genes had significantly increased expression in tumors *MAGEA12* ($p < 0.02$), *SBSN* ($p < 0.002$), *TKTL1* ($p < 0.02$), *ZNF711* ($p < 0.008$), *NY-ESO-1* ($p < 0.001$), *G6PD* ($p < 0.006$). Three genes slightly missed significance at the $\alpha < 0.05$ level: *MAGEA3* ($p < 0.09$), *MAGEA4* ($p < 0.06$) and *MAGEA1* ($p < 0.08$) (2 tailed Student's t-test assuming unequal variance).

[0137] The question as to whether demethylation was responsible for the derepression of the CTAs and related genes in NSCLC was next addressed. Four target genes showed a significant positive correlation between mRNA expression (quantitative RT-PCR) and promoter hypomethylation (QUMSP): *MAGEA12* ($p = 0.024$), *MAGEA4* ($p < 0.004$), *SBSN* ($p = 0.004$) and *NY-ESO-1* ($p < 0.004$) (Figure 16A-D) (Spearman's correlation permutation test). *TKTL1* ($p = 0.1$), *MAGEA5* ($p = 0.104$) and *MAGEA3* ($p = 0.2$) also showed a positive correlation between demethylation and expression, but missed significance. These data

suggested demethylation of promoter regions was partially responsible for the regulation of the majority of the target genes.

[0138] CTAs and associated target genes are coordinately expressed. Given the findings in the previous analyses of the cohort of primary tissue showing that these target genes were differentially expressed in tumors, their promoter regions were coordinately demethylated within tumors and expression was correlated with demethylation, the initial cohort of 111 tumors assayed using the Affymetrix Human Genome U133 Plus 2.0 mRNA expression platform was examined to determine if the target genes were coordinately expressed within tumor samples in this large sample set. Figure 4A shows a heat map of transcript expression as measured by the U133 Plus 2.0 array for 40 normal lung samples from non-cancer patients and 111 NSCLC primary tissue samples. This analysis not only provided confirmation that expression of the target genes is limited to a subset of tumors with little or no expression in the normal tissue, but also, that these targets appear to be coordinately expressed in a subset of these tumors.

[0139] To formally test the coordinate expression of these genes, a p-value matrix derived from the Pearson's correlation coefficients calculated between the expression levels of each target was next constructed (Figure 13B). Values to the upper right were corrected with the Benjamin Hochberg multiple test correction to decrease the false discovery rate; however, there was no change in significance after this correction (uncorrected values displayed in lower left). Shaded cells represent significant p-values. Using this pairwise comparison method, a highly significant coordinated upregulation of all 10 target genes within a subset of tumor samples was found.

[0140] Although the invention has been described with reference to the above examples, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. A method for identifying a cell that exhibits or is predisposed to exhibiting unregulated growth, comprising detecting hypomethylation of a gene or a regulatory region in at least one gene in the cell, wherein the at least one gene is hypomethylated as compared to a corresponding normal cell not exhibiting unregulated growth, thereby identifying the cell as exhibiting or predisposed to exhibiting unregulated growth.
2. The method of claim 1, wherein at least two genes or regulatory regions are hypomethylated.
3. The method of claim 1, wherein the regulatory region of the at least one gene comprises a BORIS binding site.
4. The method according to claim 1, wherein the regulatory region of the at least one gene comprises a promoter of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP.
5. The method of claim 1, wherein the cell exhibiting, or predisposed to exhibiting unregulated growth, is a cancer cell.
6. The method of claim 5, wherein the cancer cell is a carcinoma or a sarcoma.
7. The method of claim 5, wherein the cancer cell is from a cancer selected from the group consisting of head cancer, neck cancer, head and neck cancer, lung cancer, breast cancer, prostate cancer, colorectal cancer, esophageal cancer, stomach cancer, leukemia/lymphoma, uterine cancer, skin cancer, endocrine cancer, urinary cancer, pancreatic cancer, gastrointestinal cancer, ovarian cancer, cervical cancer, and adenomas.
8. The method of claim 7, wherein the cancer is head and neck cancer.
9. The method of claim 7, wherein the cancer is lung cancer.

10. The method of claim 1 wherein the hypomethylation is of a CpG dinucleotide motif in the at least one gene or regulatory region.

11. The method of claim 1 hypomethylation is of a CpG dinucleotide motif in a promoter of the regulatory region of the at least one gene.

12. The method of claim 1 wherein hypomethylation is detected by detecting increased expression of the at least one gene.

13. The method of claim 1 wherein hypomethylation is detected by detecting increased mRNA of the at least one gene.

14. The method of claim 12 wherein increased expression of the at least one gene is determined by comparison to a control sample.

15. The method of claim 13 wherein increased mRNA of the at least one gene is determined by hybridization to a nucleotide probe.

16. The method of claim 12 wherein increased expression is detected by reverse transcription-polymerase chain reaction (RT-PCR).

17. The method of claim 16 wherein the RT-PCR is performed in a real-time and quantitative manner.

18. The method of claim 12 wherein hypomethylation is detected by detecting increased protein encoded by the gene.

19. The method of claim 1 wherein hypomethylation is detected by contacting at least a portion of the gene with a methylation-sensitive restriction endonuclease, said endonuclease preferentially cleaving non-methylated recognition sites relative to methylated recognition sites, whereby cleavage of the portion of the gene indicates non-methylation of the portion of the gene provided that the gene comprises a recognition site for the methylation-sensitive restriction endonuclease.

20. The method of claim 1 wherein hypomethylation is detected by contacting at least a portion of the gene of the cell with a chemical reagent that selectively modifies a non-methylated cytosine residue relative to a methylated cytosine residue, or selectively modifies a methylated cytosine residue relative to a non-methylated cytosine residue; and detecting a product generated by the contacting step.

21. The method of claim 20 wherein the step of detecting comprises hybridization with at least one probe that hybridizes to a sequence comprising a modified non-methylated CpG dinucleotide motif but not to a sequence comprising an unmodified methylated CpG dinucleotide.

22. The method of claim 20 wherein the step of detecting comprises amplification with at least one primer that hybridizes to a sequence comprising a modified non-methylated CpG dinucleotide motif but not to a sequence comprising an unmodified methylated CpG dinucleotide motif thereby forming amplification products.

23. The method of claim 20 wherein the step of detecting comprises amplification with at least one primer that hybridizes to a sequence comprising an unmodified methylated CpG dinucleotide motif but not to a sequence comprising a modified non-methylated CpG dinucleotide motif thereby forming amplification products.

24. The method of claim 20 wherein the product is detected by a method selected from the group consisting of electrophoresis, hybridization, amplification, primer extension, sequencing, ligase chain reaction, chromatography, mass spectrometry, and combinations thereof.

25. The method of claim 20 wherein the chemical reagent is hydrazine.

26. The method of claim 25 further comprising cleaving the hydrazine-contacted at least a portion of the gene with piperidine.

27. The method of claim 20 wherein the chemical reagent comprises bisulfite ions.

28. The method of claim 27 further comprising treating the bisulfite ion-contacted at least a portion of the gene with alkali.

29. A method for diagnosing a disorder in a subject having or at risk of developing a cell proliferative disorder comprising:

contacting a nucleic acid-containing sample from cells of the subject with an agent that provides a determination of the methylation state of at least one regulatory region of a gene, wherein the at least one regulatory region is demethylated as compared to a corresponding normal cell; and

identifying hypomethylation of the regulatory region as compared to the same region of the at least one regulatory region in a subject not having the proliferative disorder, wherein hypomethylation is indicative of a subject having or at risk of developing the proliferative disorder.

30. The method of claim 29, wherein at least two regulatory regions are hypomethylated.

31. The method of claim 29, wherein the regulatory region of the at least one gene comprises a BORIS binding site.

32. The method according to claim 29, wherein the regulatory region of the at least one gene comprises a promoter of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP.

33. The method of claim 29, wherein the cell exhibiting, or predisposed to exhibiting unregulated growth, is a cancer cell.

34. The method of claim 33, wherein the cancer cell is a carcinoma or a sarcoma.

35. The method of claim 33, wherein the cancer cell is from a cancer selected from the group consisting of head cancer, neck cancer, head and neck cancer, lung cancer, breast

cancer, prostate cancer, colorectal cancer, esophageal cancer, stomach cancer, leukemia/lymphoma, uterine cancer, skin cancer, endocrine cancer, urinary cancer, pancreatic cancer, gastrointestinal cancer, ovarian cancer, cervical cancer, and adenomas.

36. The method of claim 35, wherein the cancer is head and neck cancer.

37. The method of claim 35, wherein the cancer is lung cancer.

38. The method of claim 29 wherein the cells are from a sample selected from the group consisting of a tissue sample, a frozen tissue sample, a biopsy specimen, a surgical specimen, a cytological specimen, whole blood, bone marrow, cerebral spinal fluid, peritoneal fluid, pleural fluid, lymph fluid, serum, mucus, plasma, urine, chyle, stool, ejaculate, sputum, nipple aspirate and saliva.

39. The method of claim 29 wherein the hypomethylation is of a CpG dinucleotide motif in the at least one gene or regulatory region.

40. The method of claim 29 hypomethylation is of a CpG dinucleotide motif in a promoter of the regulatory region of the at least one gene.

41. The method of claim 29 wherein hypomethylation is detected by detecting increased expression of the at least one gene.

42. The method of claim 29 wherein hypomethylation is detected by detecting increased mRNA of the at least one gene.

43. The method of claim 41 wherein increased expression is detected by reverse transcription-polymerase chain reaction (RT-PCR).

44. The method of claim 29 wherein hypomethylation is detected by detecting increased protein encoded by the gene.

45. The method of claim 29 wherein hypomethylation is detected by contacting at least a portion of the gene with a methylation-sensitive restriction endonuclease, said endonuclease preferentially cleaving non-methylated recognition sites relative to methylated

recognition sites, whereby cleavage of the portion of the gene indicates non-methylation of the portion of the gene provided that the gene comprises a recognition site for the methylation-sensitive restriction endonuclease.

46. The method of claim 29 wherein hypomethylation is detected by contacting at least a portion of the gene of the cell with a chemical reagent that selectively modifies a non-methylated cytosine residue relative to a methylated cytosine residue, or selectively modifies a methylated cytosine residue relative to a non-methylated cytosine residue; and detecting a product generated by the contacting step.

47. The method of claim 46 wherein the step of detecting comprises hybridization with at least one probe that hybridizes to a sequence comprising a modified non-methylated CpG dinucleotide motif but not to a sequence comprising an unmodified methylated CpG dinucleotide.

48. The method of claim 46 wherein the step of detecting comprises amplification with at least one primer that hybridizes to a sequence comprising a modified non-methylated CpG dinucleotide motif but not to a sequence comprising an unmodified methylated CpG dinucleotide motif thereby forming amplification products.

49. The method of claim 46 wherein the step of detecting comprises amplification with at least one primer that hybridizes to a sequence comprising an unmodified methylated CpG dinucleotide motif but not to a sequence comprising a modified non-methylated CpG dinucleotide motif thereby forming amplification products.

50. The method of claim 46 wherein the product is detected by a method selected from the group consisting of electrophoresis, hybridization, amplification, primer extension, sequencing, ligase chain reaction, chromatography, mass spectrometry, and combinations thereof.

51. The method of claim 46 wherein the chemical reagent is hydrazine.

52. The method of claim 51 further comprising cleaving the hydrazine-contacted at least a portion of the gene with piperidine.

53. The method of claim 46 wherein the chemical reagent comprises bisulfite ions.

54. The method of claim 53 further comprising treating the bisulfite ion-contacted at least a portion of the gene with alkali.

55. A method of determining the prognosis of a subject having a cell proliferative disorder comprising:

determining the methylation state of at least one regulatory region of a gene in a nucleic acid sample from the subject, wherein hypomethylation as compared to a corresponding normal cell in the subject or a subject not having the disorder, is indicative of a poor prognosis.

56. A method of ameliorating a cell proliferative disorder in a subject in need thereof comprising:

administering to the subject an agent that methylates at least one regulatory region in a gene that is demethylated as compared to a subject not having the disorder, thereby reducing expression of the at least one gene and ameliorating the cell proliferative disorder.

57. The method of claim 56, wherein at least two genes or regulatory regions are hypomethylated.

58. The method of claim 56, wherein the regulatory region of the at least one gene comprises a BORIS binding site.

59. The method of claim 56, wherein the regulatory region comprises of the at least one gene comprises a promoter of a gene selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP.

60. The method of claim 56, wherein the cell proliferative disorder is a cancer.
61. The method of claim 60, wherein the cancer is a carcinoma or a sarcoma.
62. The method of claim 60, wherein the cancer is selected from the group consisting of head cancer, neck cancer, head and neck cancer, lung cancer, breast cancer, prostate cancer, colorectal cancer, esophageal cancer, stomach cancer, leukemia/lymphoma, uterine cancer, skin cancer, endocrine cancer, urinary cancer, pancreatic cancer, gastrointestinal cancer, ovarian cancer, cervical cancer, and adenomas.
63. The method of claim 62, wherein the cancer is head and neck cancer.
64. The method of claim 62, wherein the cancer is lung cancer.
65. A method of identifying a gene activated by hypomethylation comprising:

comparing an expression analysis of a cell treated with an agent that reduces methylation to an expression analysis of a control cell not treated with the agent, wherein an increase in expression of a gene is indicative of a gene activated by demethylation.
66. The method of claim 65, wherein the cell is a minimally transformed cell line.
67. The method of claim 65, wherein the demethylating agent is 5-aza-deoxycytidine.
68. The method of claim 65, further comprising an expression analysis of a tissue sample and a tumor sample from the same tissue of origin as the normal cell, wherein an increase in expression of a gene in a tumor sample as compared to a normal sample is correlated to the genes activated by demethylation in the treated cell.
69. The method of claim 68, further comprising sequence analysis of the identified genes to confirm the presence of CpG islands in the promoter region of the genes.
70. The method of claim 69, further comprising determining the methylation status of the promoter regions of the identified genes.
71. The method of claim 70, wherein the determining of the methylation status comprises contacting the gene with a chemical reagent that selectively modifies a non-methylated cytosine residue relative to a methylated cytosine residue, or selectively modifies

a methylated cytosine residue relative to a non-methylated cytosine residue; and detecting a product generated by the contacting step.

72. The method of claim 71 wherein the chemical reagent comprises bisulfite ions.

73. The method of claim 72 further comprising treating the bisulfite ion-contacted at least a portion of the gene with alkali.

74. A method for determining whether a subject is responsive to a particular therapeutic regimen comprising determining the methylation status of one or more genes or regulatory regions thereof, selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP wherein hypomethylation of the gene or regulatory region thereof as compared with a normal subject is indicative of a subject who is responsive to the therapeutic regimen.

75. The method of claim 74, wherein the therapeutic regimen is administration of a chemotherapeutic agent.

76. The method of claim 75, wherein the chemotherapeutic agent is selected from the group consisting of methotrexate, cisplatin/carboplatin; canbusil, dactinomycin; taxol (paclitaxol), a vinca alkaloid, a mitomycin-type antibiotic, a bleomycin-type antibiotic, antifolate, colchicine, demecoline, etoposide, taxane, anthracycline antibiotic, doxorubicin, daunorubicin, carminomycin, epirubicin, idarubicin, mithoxanthrone, 4-dimethoxy-daunomycin, 11-deoxydaunorubicin, 13-deoxydaunorubicin, adriamycin-14-benzoate, adriamycin-14-octanoate, adriamycin-14-naphthaleneacetate, amsacrine, carmustine, cyclophosphamide, cytarabine, etoposide, lovastatin, melphalan, topotecan, oxalaplatin, chlorambucil, methotrexate, lomustine, thioguanine, asparaginase, vinblastine, vindesine, tamoxifen, and mechlorethamine.

77. The method of claim 74, wherein the therapeutic regimen is administration of a vaccine directed to a protein encoded by the hypomethylated gene.

78. A kit useful for the detection of a methylated CpG-containing nucleic acid in determining the methylation status of one or more genes or regulatory regions thereof, selected from the group consisting of TKTL1, H19, MAGEA2, MAGEA3, MAGEA4, MAGEA11, GPR17, GRIN1, C19ORF28, MAGEA12, MAGEA1, MAGEA5, NY-ESO-1, MAGEA9, MAGEA6, MAGEB2, CT45-2, SBSN, G6PD, ZNF711, CrispL, KRT86, KIPV467, KRT81, CSPG5, PP1R14A, KISS1R, KIAA1937 protein, SOX30, DEAD, and KBGP comprising: a carrier element containing one or more containers comprising a first container containing a reagent which modifies unmethylated cytosine and a second container containing primers for amplification of the one or more genes or regulatory regions thereof, wherein the primers distinguish between modified methylated and nonmethylated nucleic acid.

79. The kit of claim 78, wherein the modifying reagent is bisulfite.

1 / 49

Integrative Epigenomic Screening Strategy

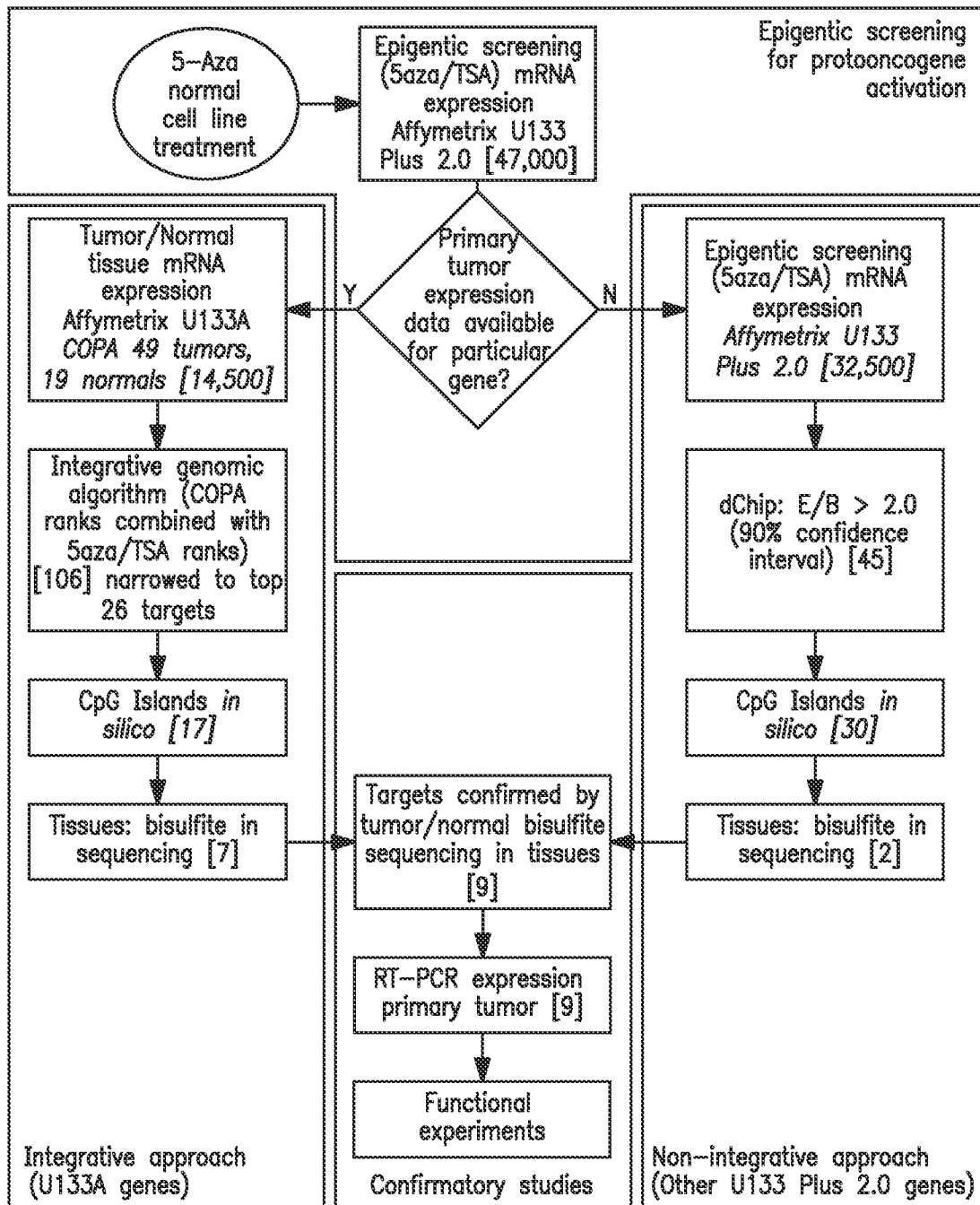


FIG. 1a

2 / 49

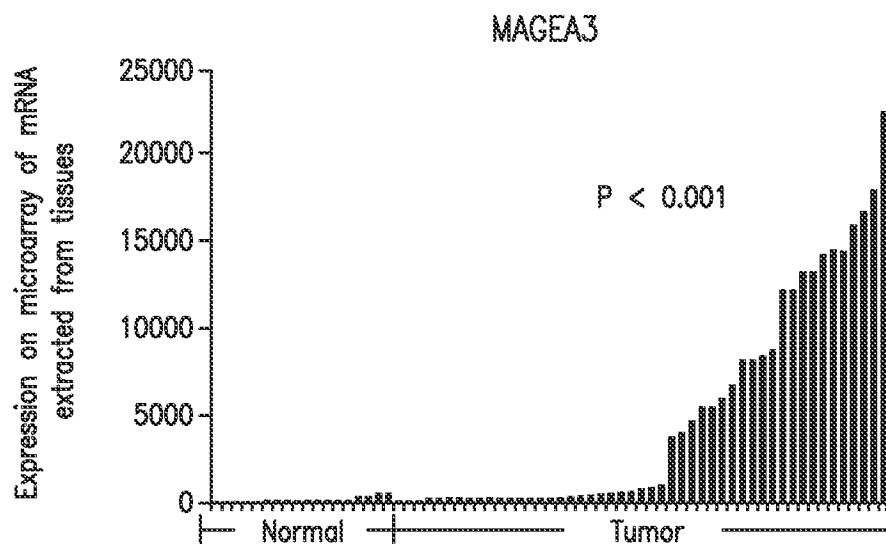


FIG. 1b

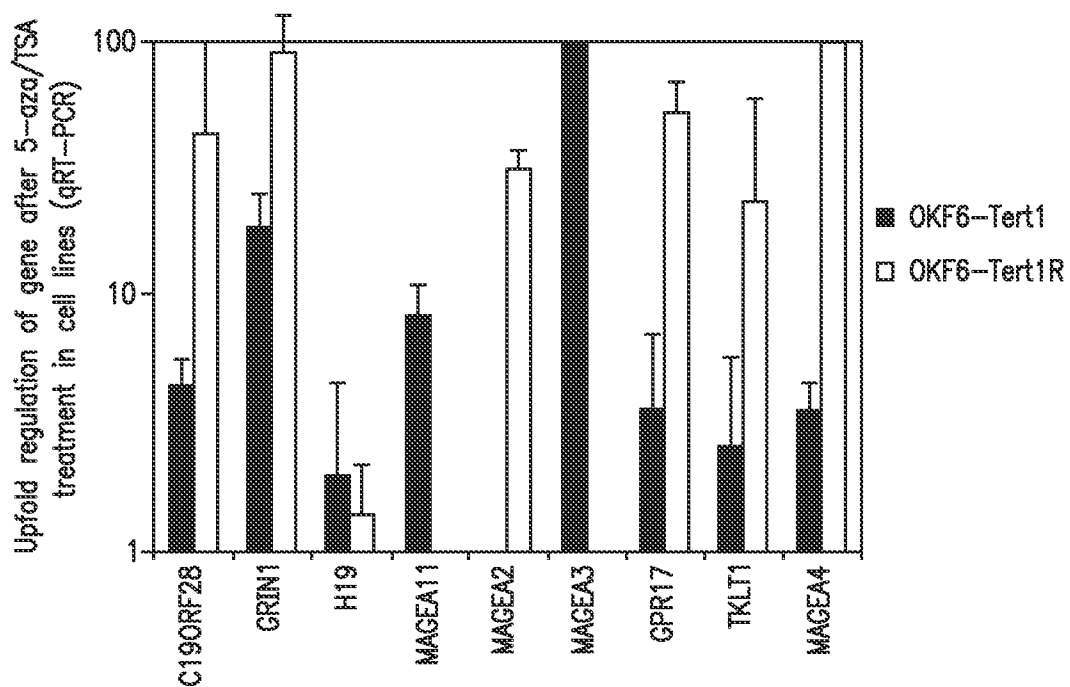
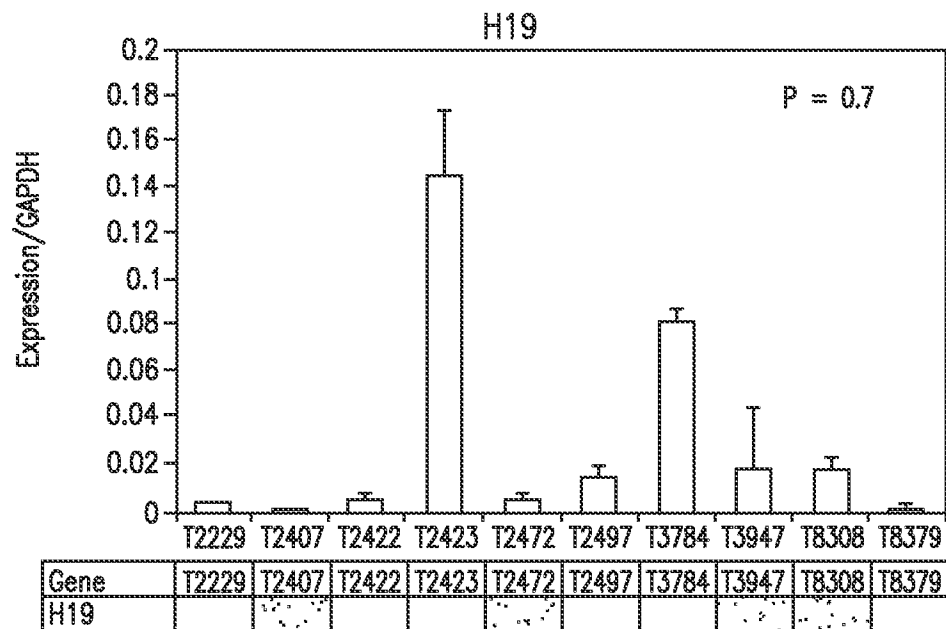
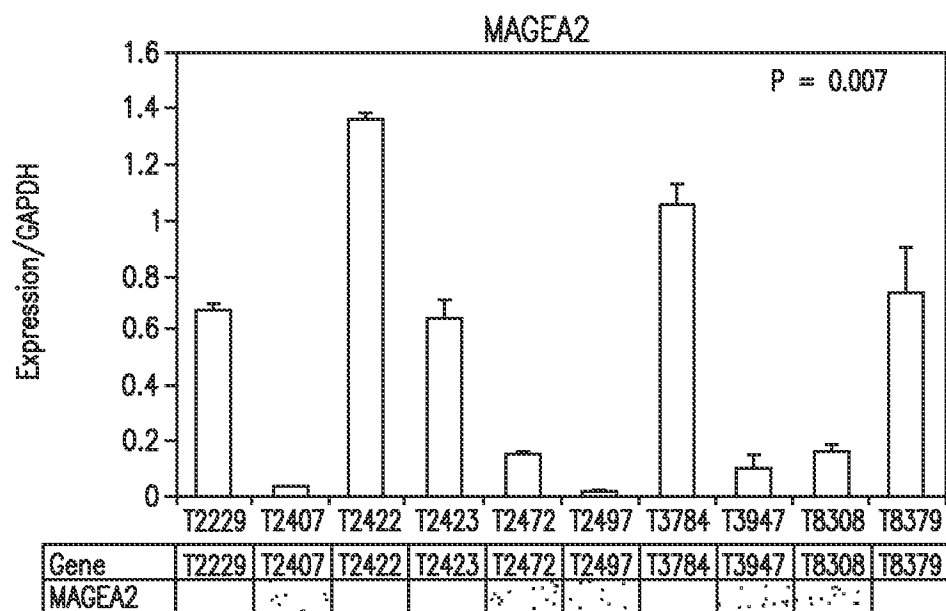


FIG. 1c

Gene	Tumors										Normals										p value (chi-squared)
	T2229	T2407	T2422	T2423	T2472	T2497	T3784	T3947	T8308	T8379	N8108	N8120	N8248	N8415	N8442	N8469	N8483	N8488	N8499	N8805	
C19ORF28																					p < 0.05
GRIN1																					p < 0.05
GPR17																					p < 0.1
TKTL1																					p < 0.05
MAGEA2																					p < 0.05
MAGEA3																					p < 0.05
MAGEA4																					p < 0.05
MAGEA11																					p < 0.05
H19																					p < 0.05

FIG. 2a

4 / 49

**FIG. 2b****FIG. 2c**

5 / 49

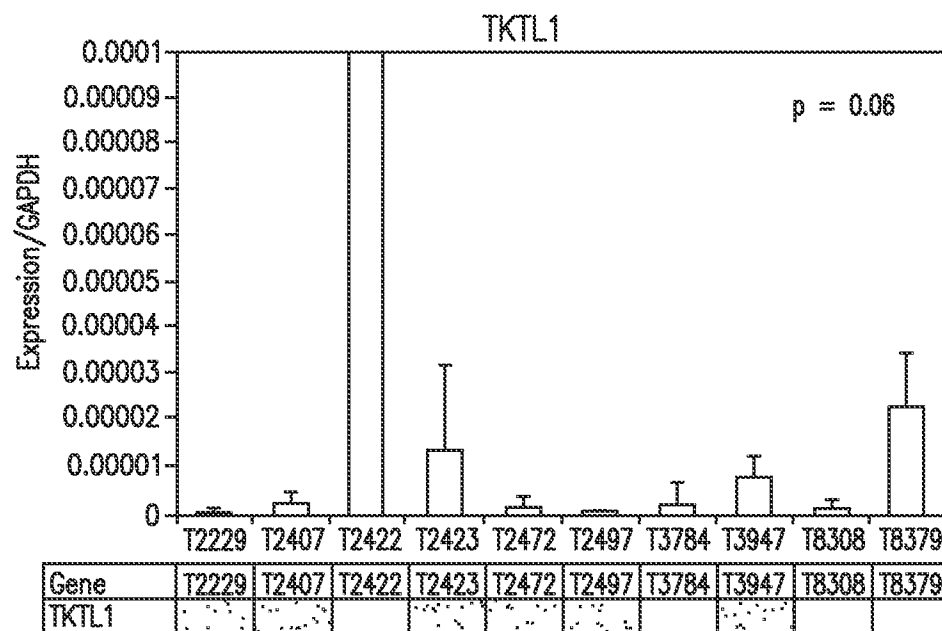


FIG. 2d

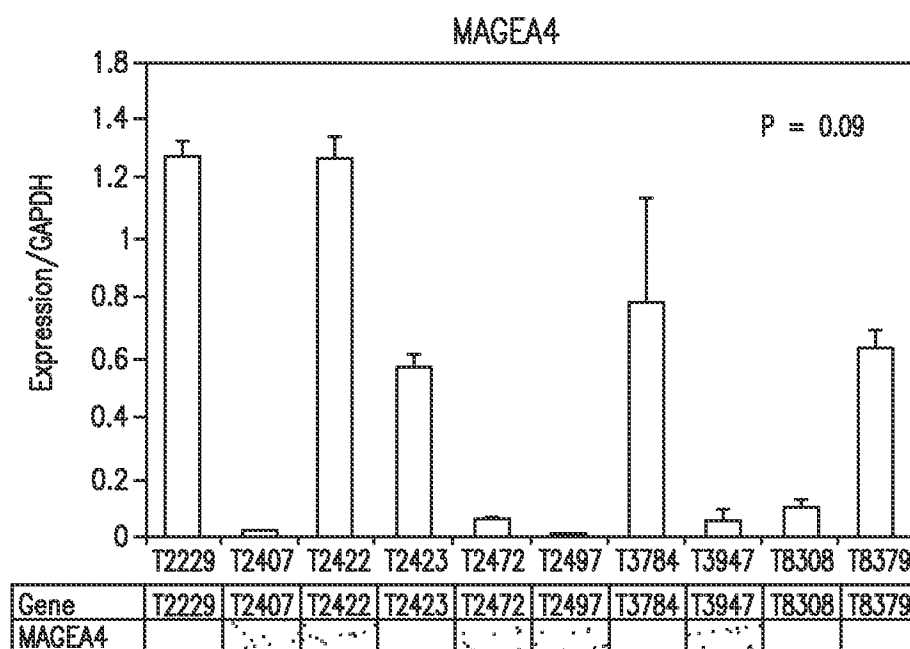
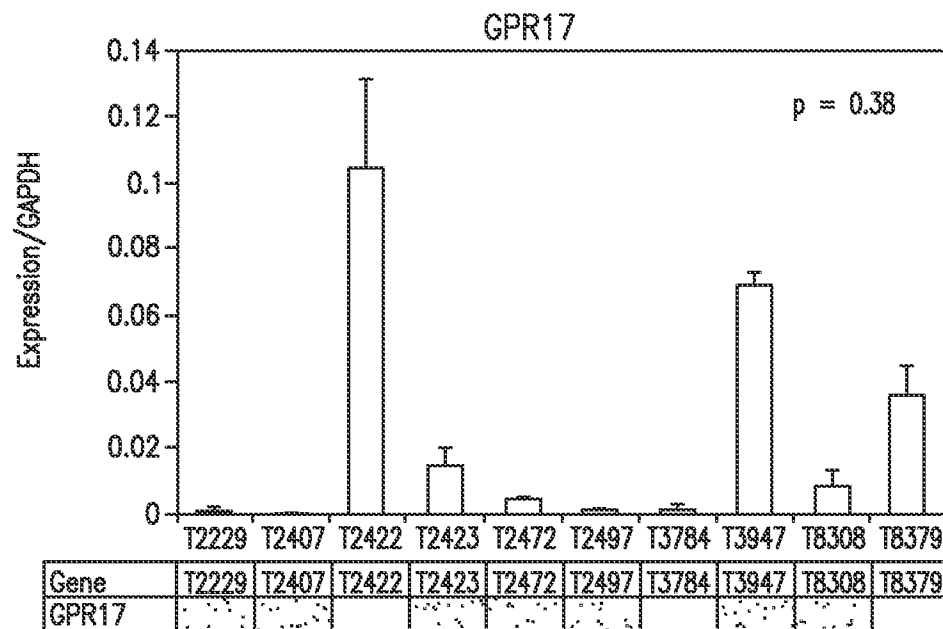
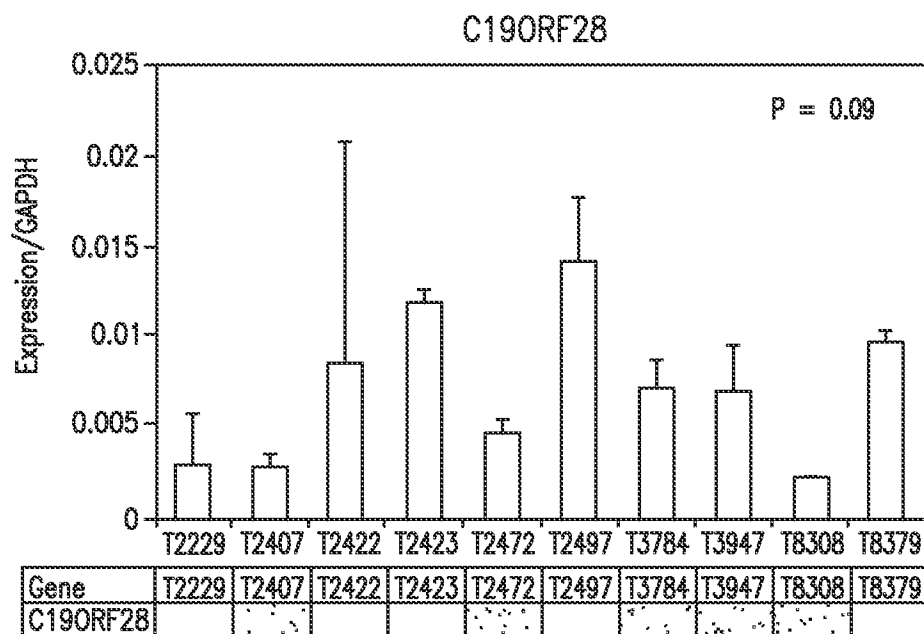
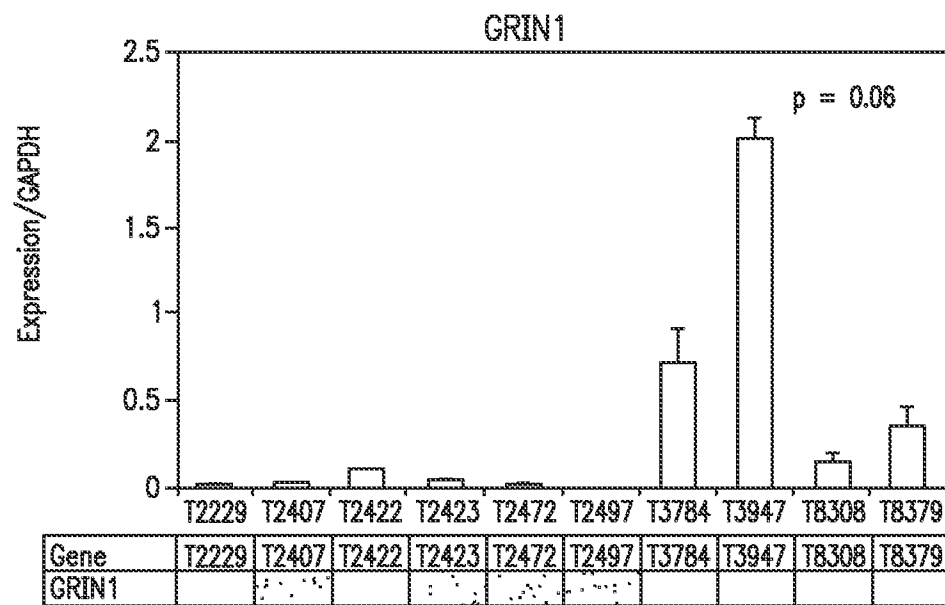
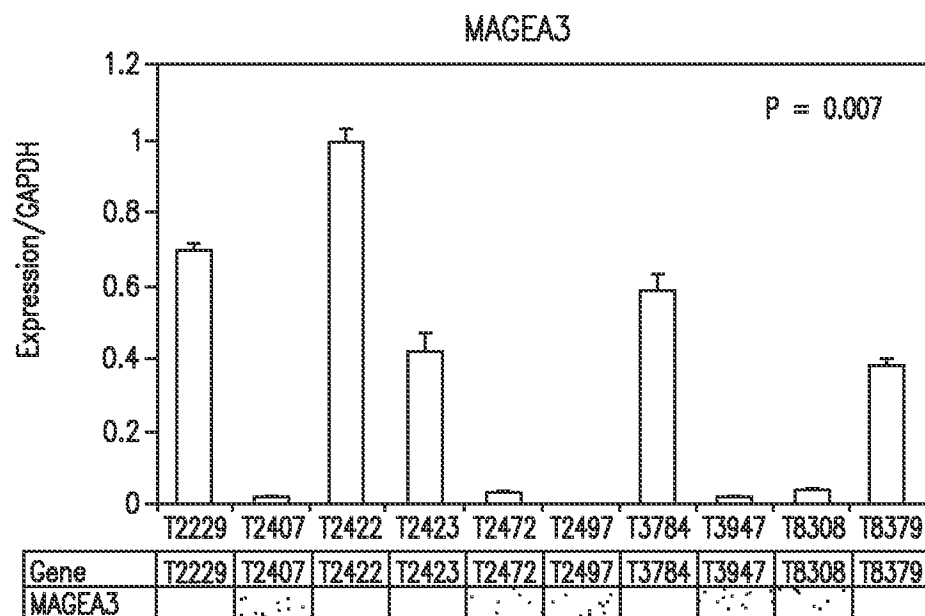


FIG. 2e

6 / 49

**FIG. 2f****FIG. 2g**

7 / 49

**FIG. 2h****FIG. 2i**

8 / 49

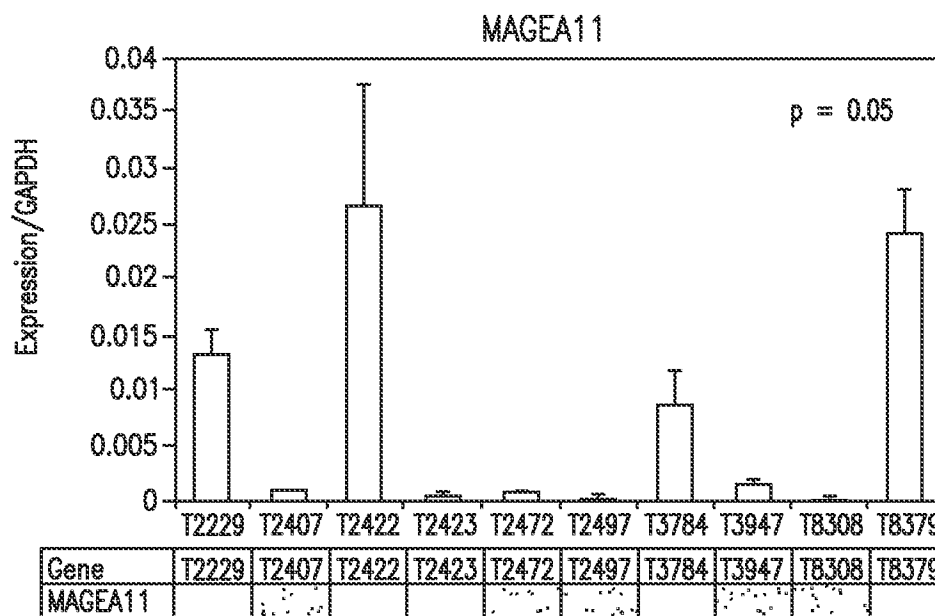


FIG. 2j

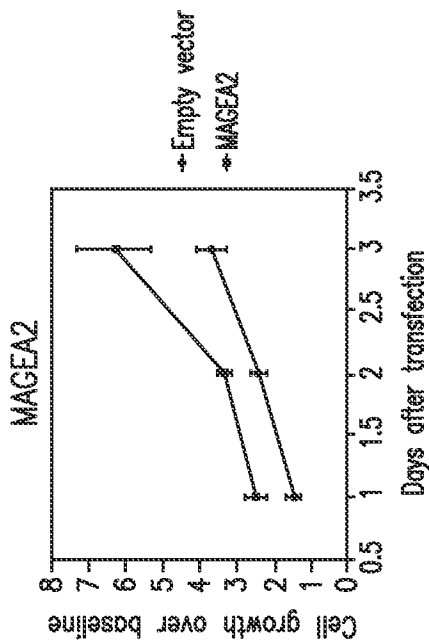


FIG. 3b

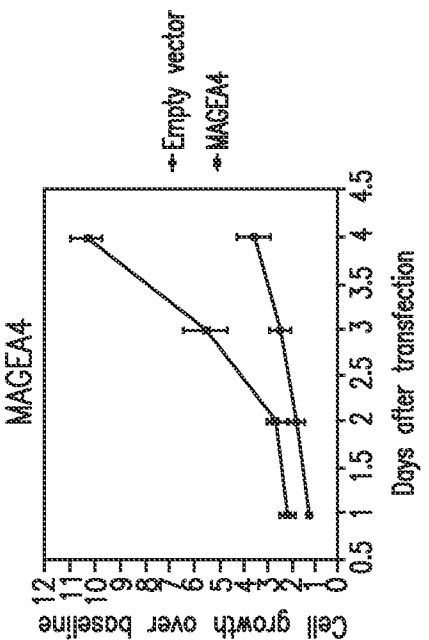


FIG. 3d

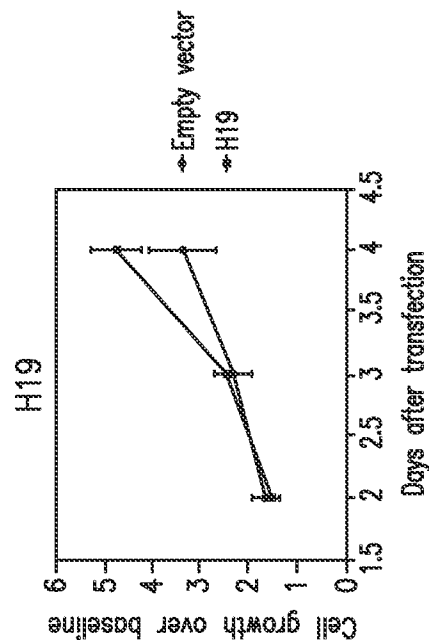


FIG. 3a

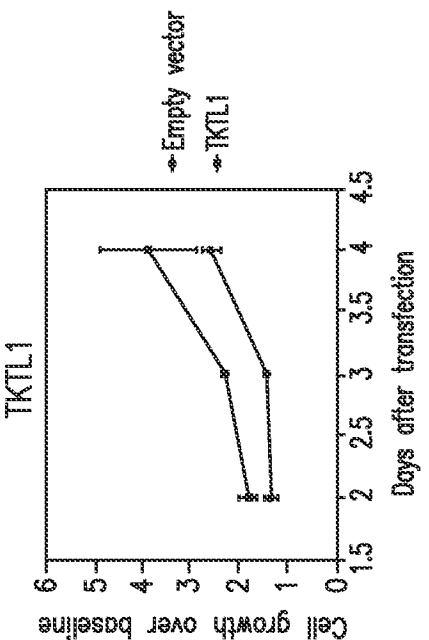


FIG. 3c

FIG. 3e

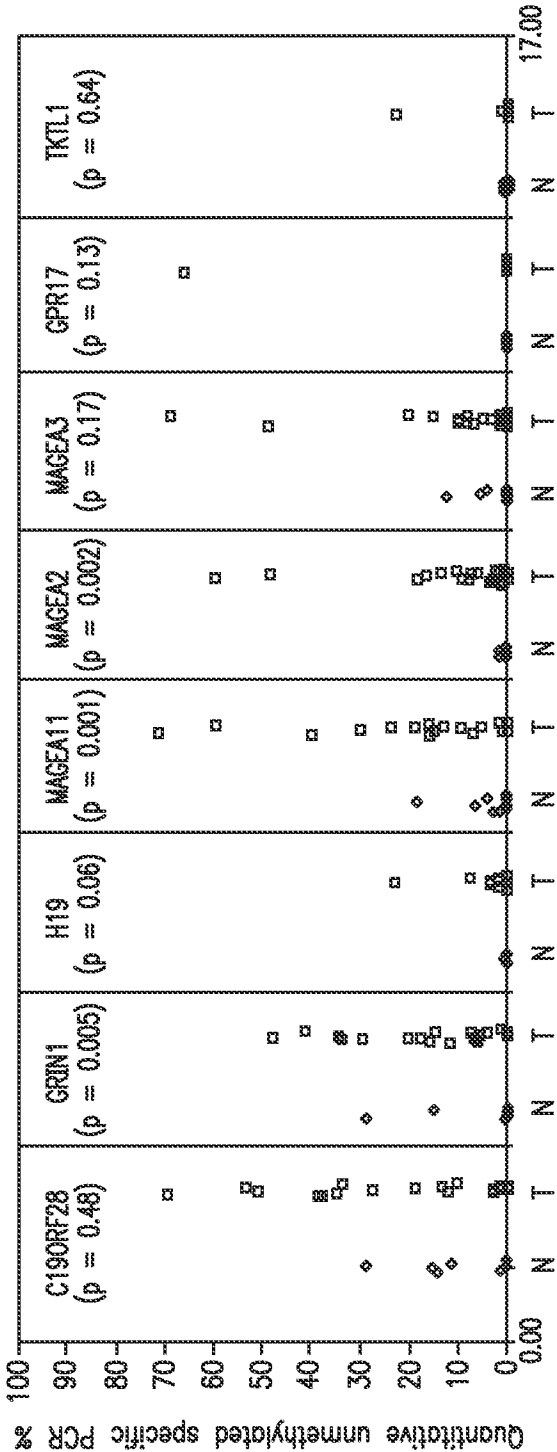
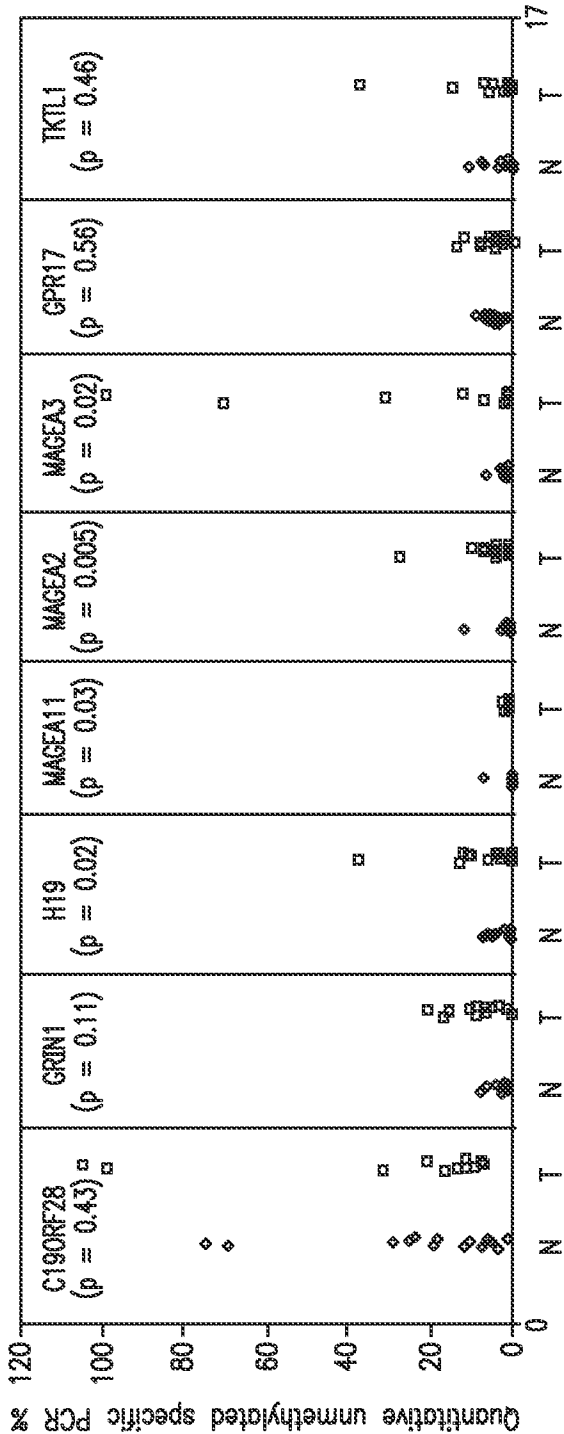
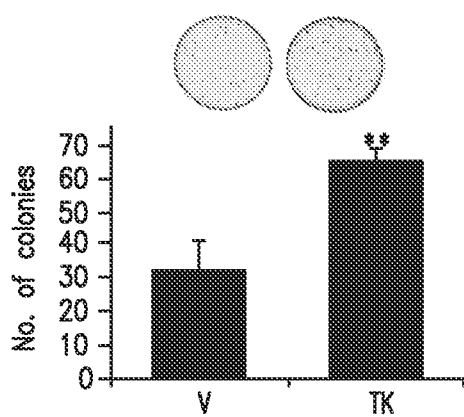
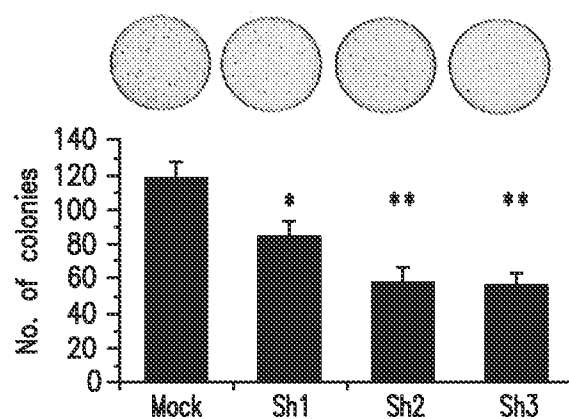
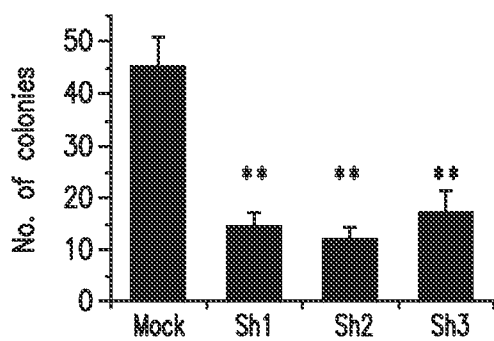
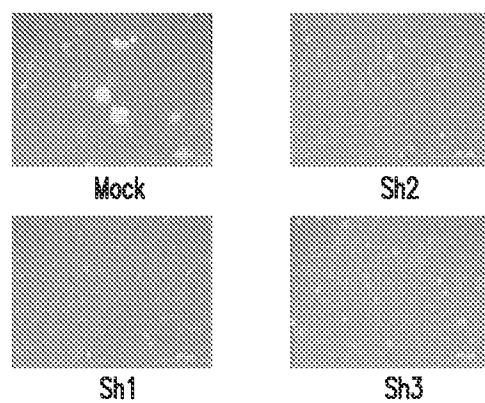


FIG. 3f



11 / 49

**FIG. 4a****FIG. 4b****FIG. 4c****FIG. 4d**

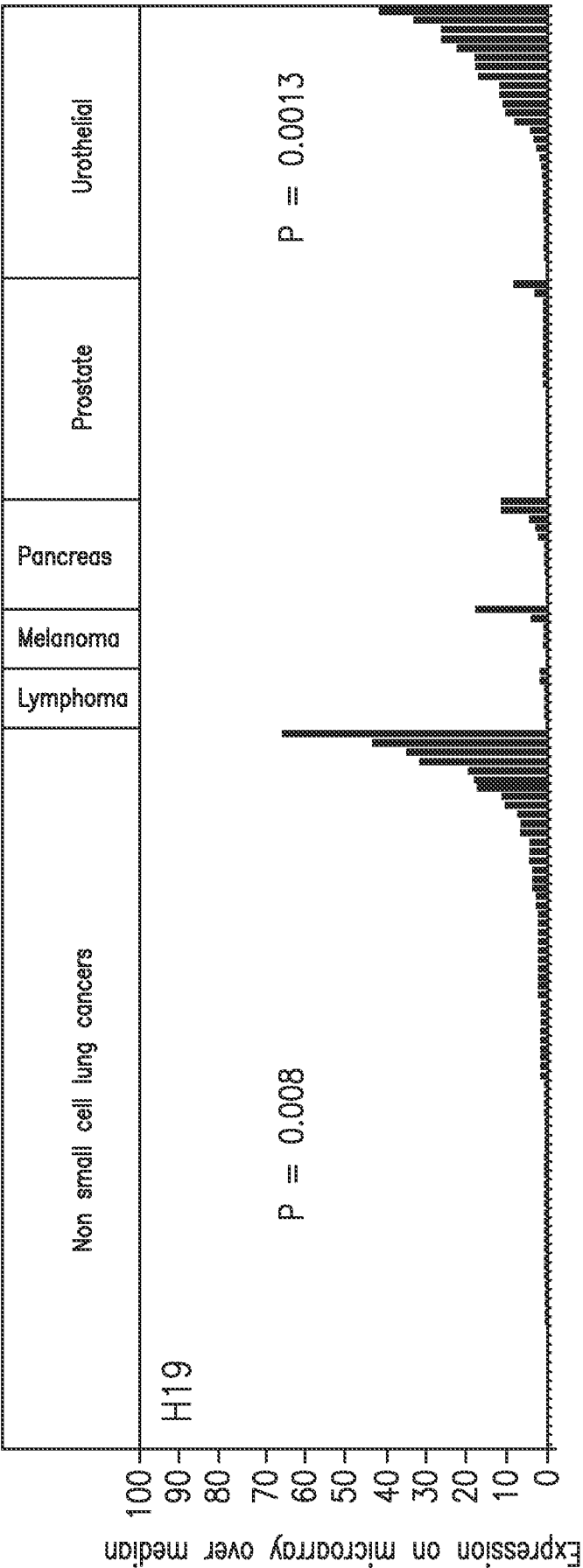


FIG. 5a

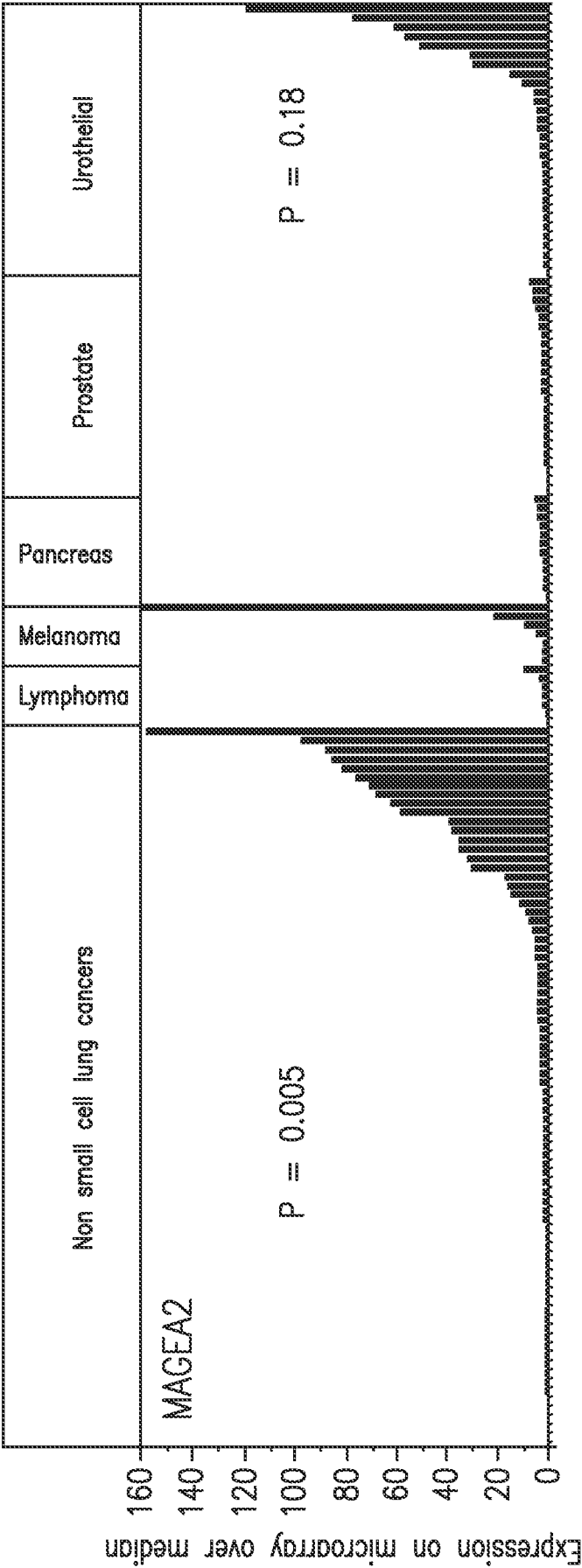


FIG. 5b

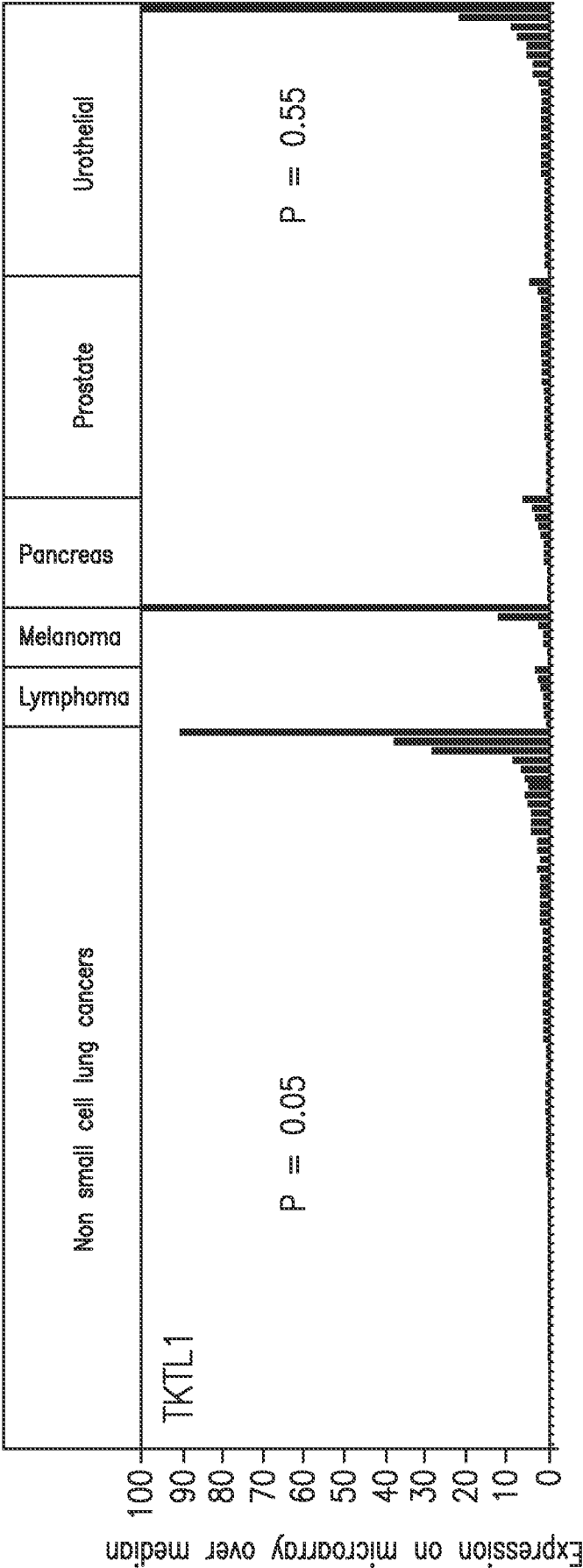


FIG. 5c

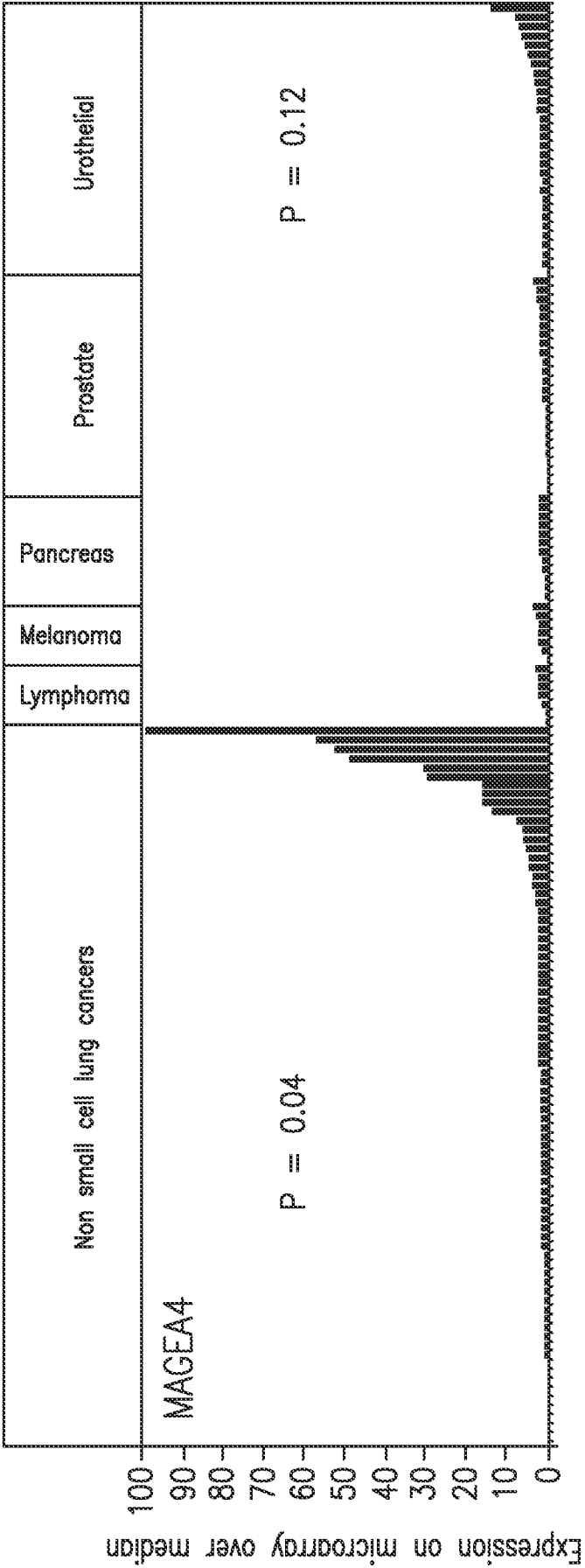


FIG. 5d

16 / 49

	MAGEA11	MAGEA2	MAGEA3	MAGEA4	TKTL1	GPR17	GRIN1
MAGEA11	<0.001	0.068	0.012	0.003	0.094	0.152	0.500
MAGEA2	0.068	<0.001	<0.001	0.064	0.811	0.887	0.336
MAGEA3	0.012	<0.001	<0.001	<0.001	0.400	0.125	0.105
MAGEA4	0.003	0.064	<0.001	<0.001	0.574	0.872	0.204
TKTL1	0.094	0.811	0.400	0.574	<0.001	0.027	0.230
GPR17	0.152	0.887	0.125	0.872	0.027	<0.001	0.047
GRIN1	0.500	0.336	0.105	0.204	0.230	0.047	<0.001

FIG. 6a

17 / 49

	H19	MAGEA11	MAGEA2	MAGEA3	MAGEA4	TKTL1	GPR17	GRIN1
H19	<0.001	0.429	0.003	0.103	0.755	0.729	0.735	0.346
MAGEA11	0.429	<0.001	0.031	0.002	0.017	0.681	0.077	0.405
MAGEA2	0.003	0.031	<0.001	<0.001	<0.001	0.524	0.340	0.565
MAGEA3	0.103	0.002	<0.001	<0.001	<0.001	0.607	0.237	0.356
MAGEA4	0.755	0.017	<0.001	<0.001	<0.001	0.473	0.352	0.391
TKTL1	0.729	0.681	0.524	0.607	0.473	<0.001	0.957	0.780
GPR17	0.735	0.077	0.340	0.237	0.352	0.957	<0.001	0.195
GRIN1	0.346	0.405	0.565	0.356	0.391	0.780	0.195	<0.001

FIG. 6b

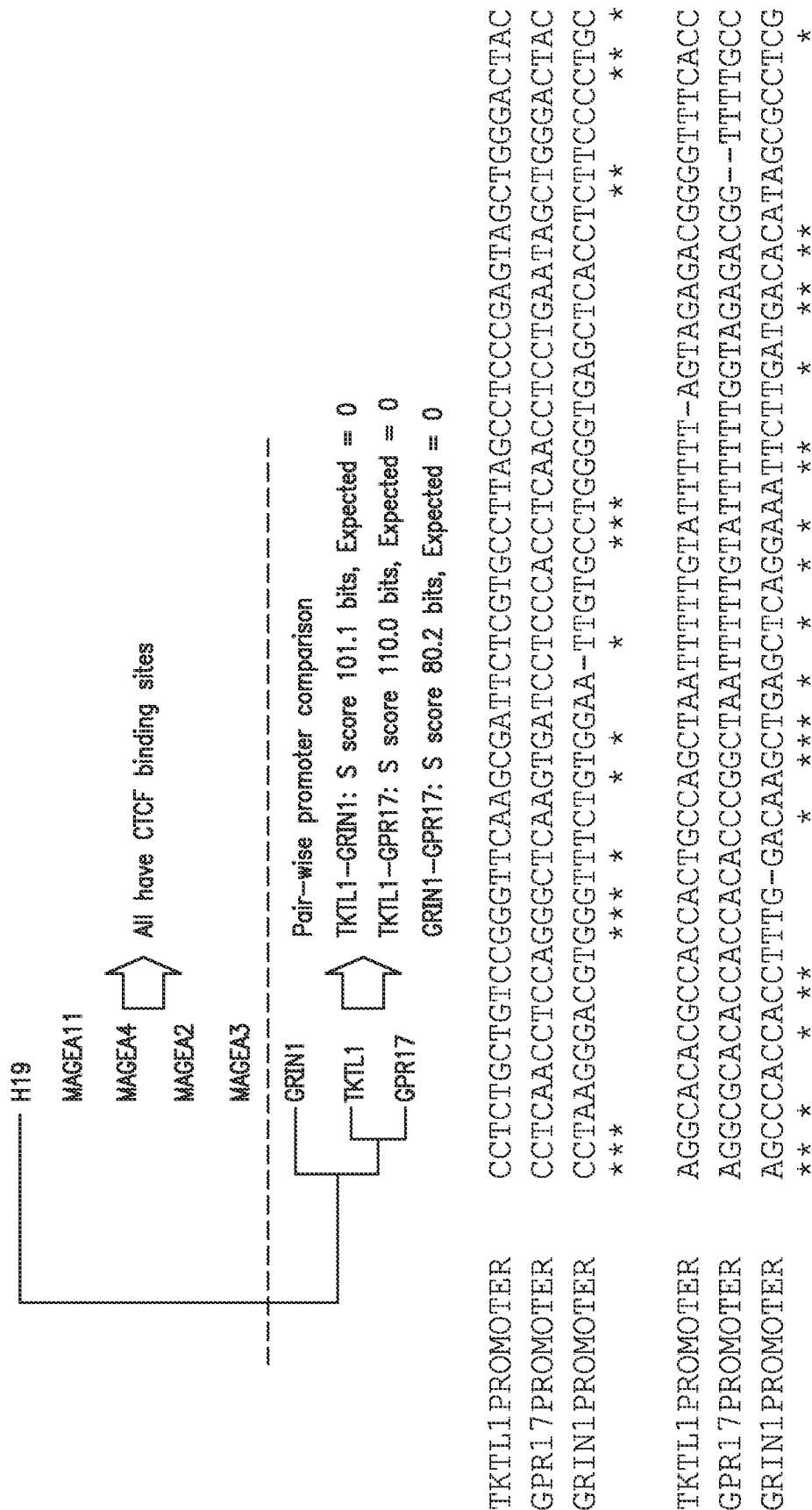


FIG. 6c

19 / 49

	H19	MAGEA11	MAGEA2	MAGEA3	TKTL1	GPR17	GRIN1
H19	<0.001	0.001	0.001	<0.001	0.938	0.725	0.105
MAGEA11	0.001	<0.001	0.098	<0.001	0.802	0.179	0.005
MAGEA2	0.001	0.098	<0.001	0.023	0.760	0.712	0.469
MAGEA3	<0.001	<0.001	0.023	<0.001	0.849	<0.001	0.010
TKTL1	0.938	0.802	0.760	0.849	<0.001	0.849	0.542
GPR17	0.725	0.179	0.712	0.001	0.849	<0.001	0.889
GRIN1	0.105	0.005	0.469	0.010	0.542	0.889	<0.001

FIG. 6d

20 / 49

	H19	MAGEA11	MAGEA2	MAGEA3	GPR17	GRIN1	TKTL1
H19	<0.001	0.049	0.343	0.039	0.058	0.002	0.272
MAGEA11	0.049	<0.001	0.109	0.028	0.329	0.388	0.389
MAGEA2	0.343	0.109	<0.001	0.013	0.316	0.484	0.224
MAGEA3	0.039	0.028	0.013	<0.001	<0.001	0.021	0.073
GPR17	0.058	0.329	0.316	0.001	<0.001	<0.001	0.369
GRIN1	0.002	0.388	0.484	0.021	<0.001	<0.001	0.734
TKTL1	0.272	0.389	0.224	0.073	0.369	0.734	<0.001

FIG. 6e

21 / 49

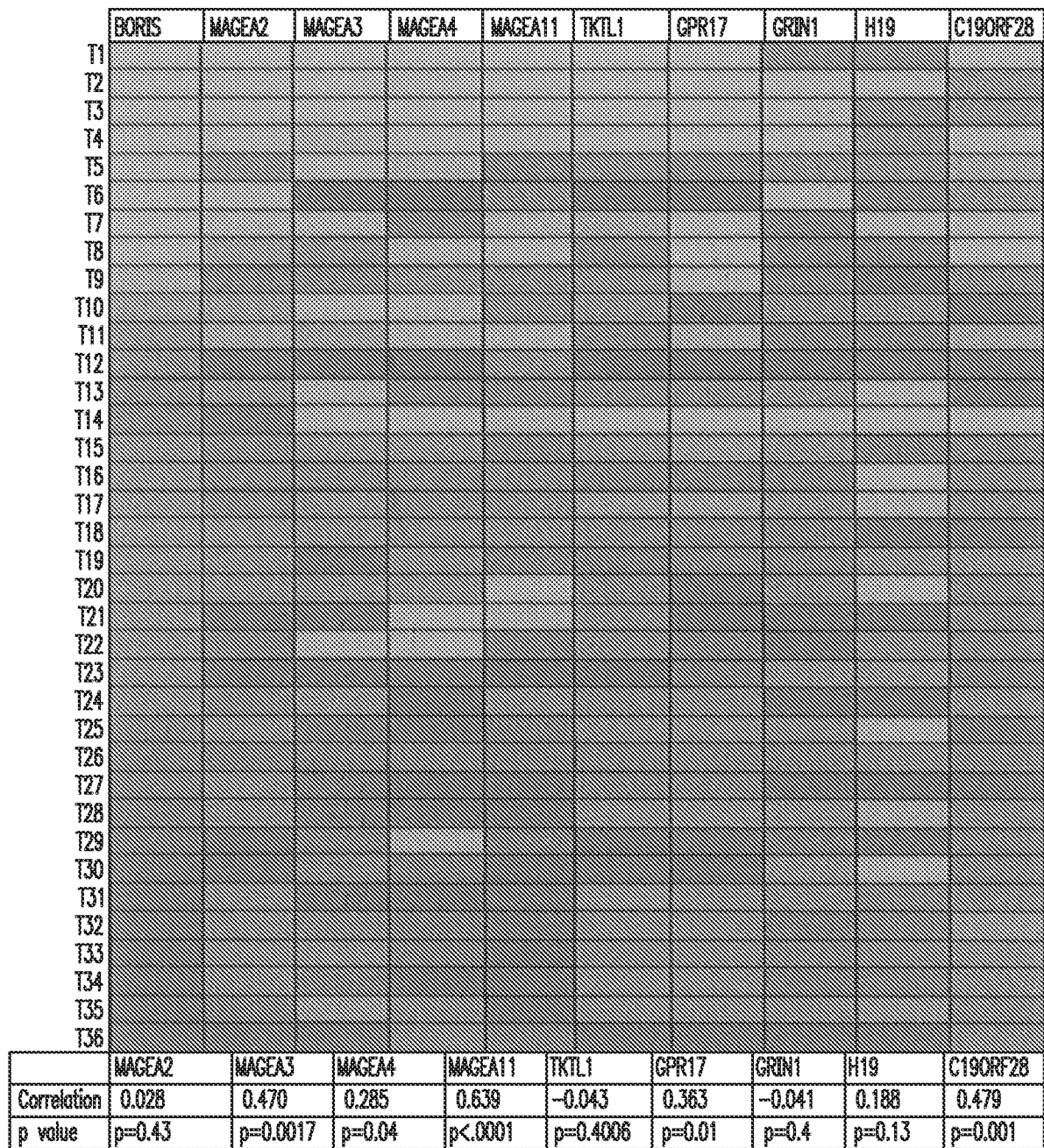
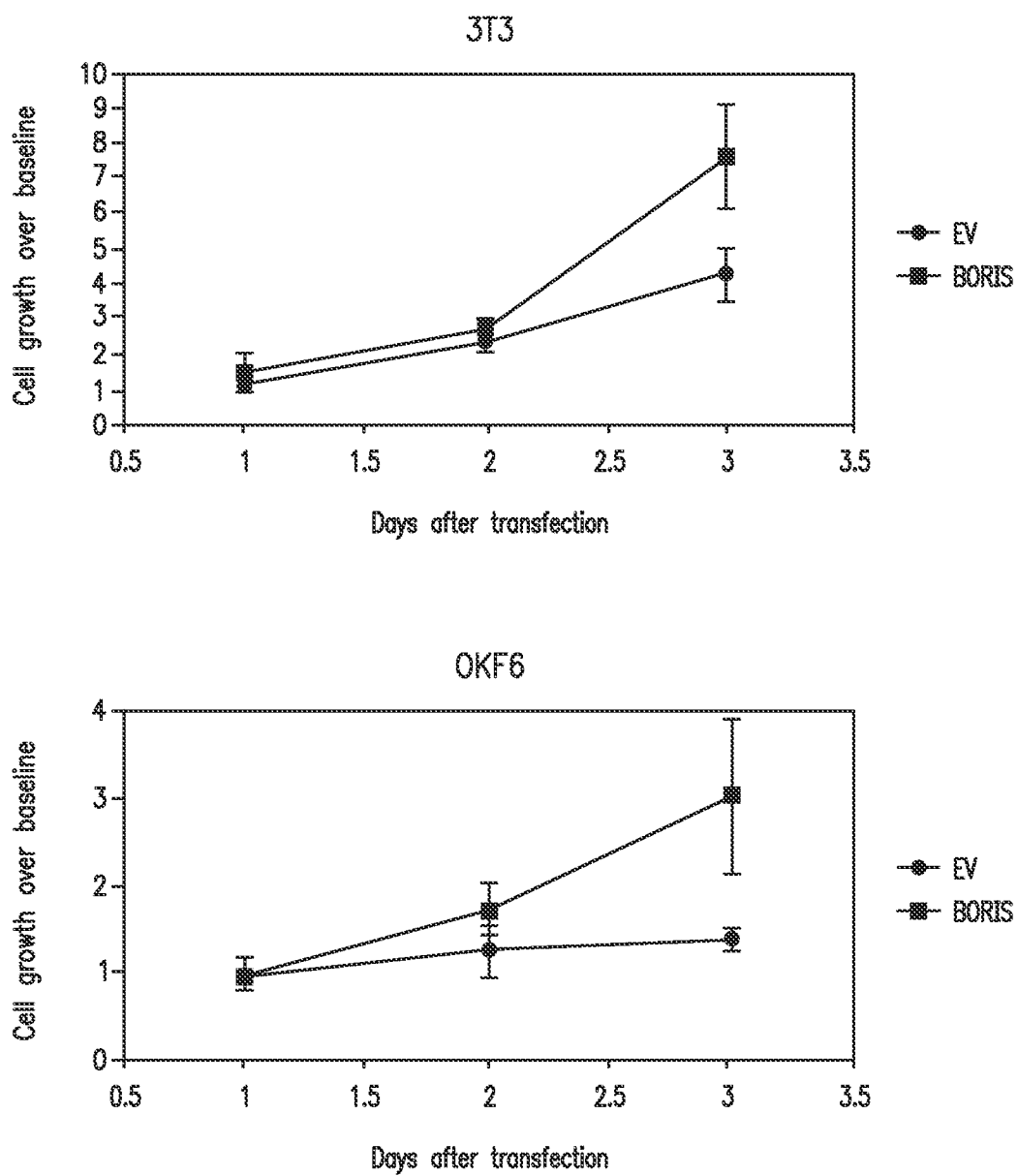
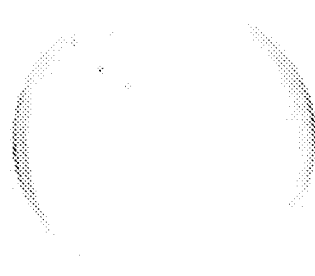
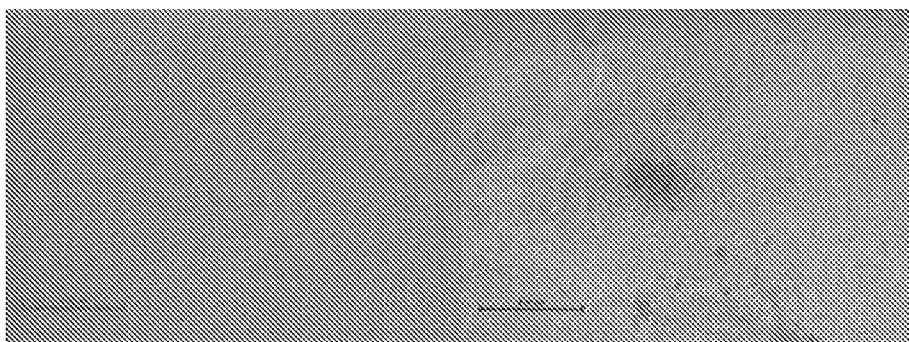
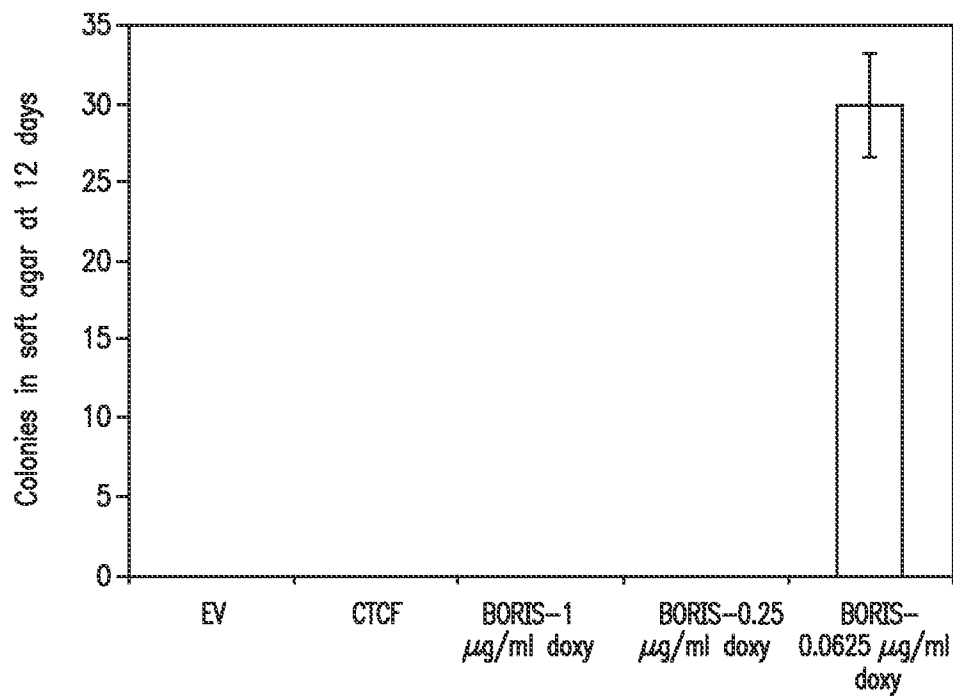


FIG. 7a

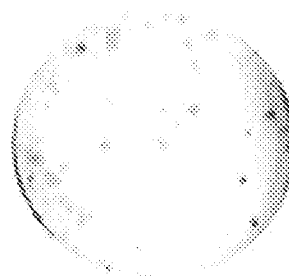
22 / 49

**FIG. 7b**

23 / 49



Transfected with
EV



Transfected with
BORIS and Rx
doxy 0.0625
 $\mu\text{g/ml}$

FIG. 7c

24 / 49

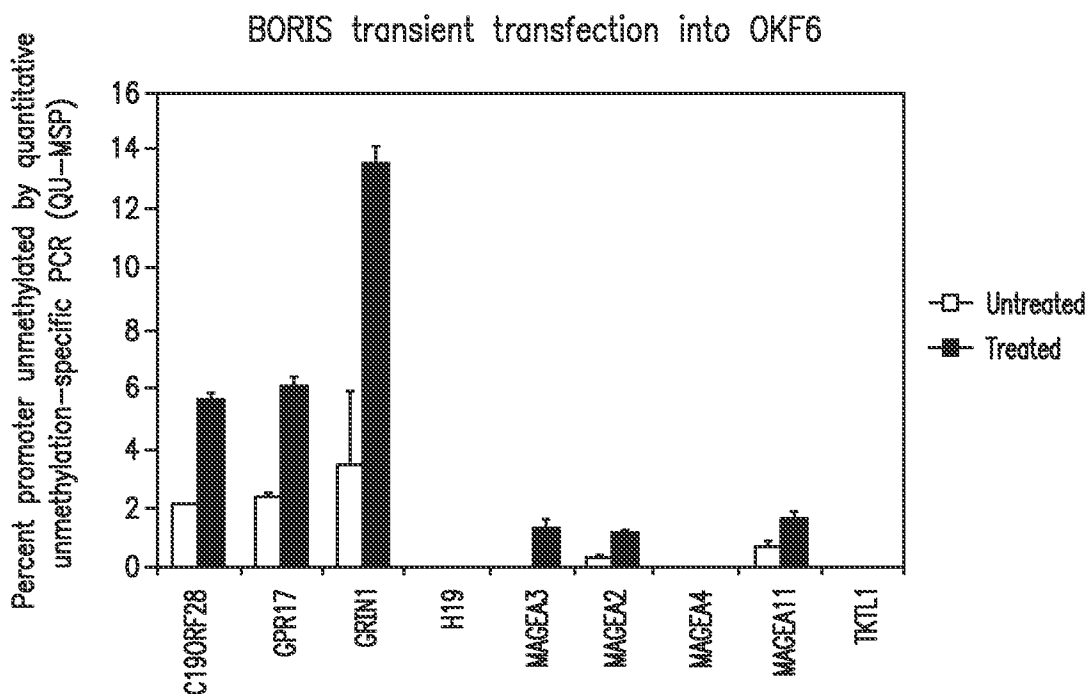


FIG. 7d

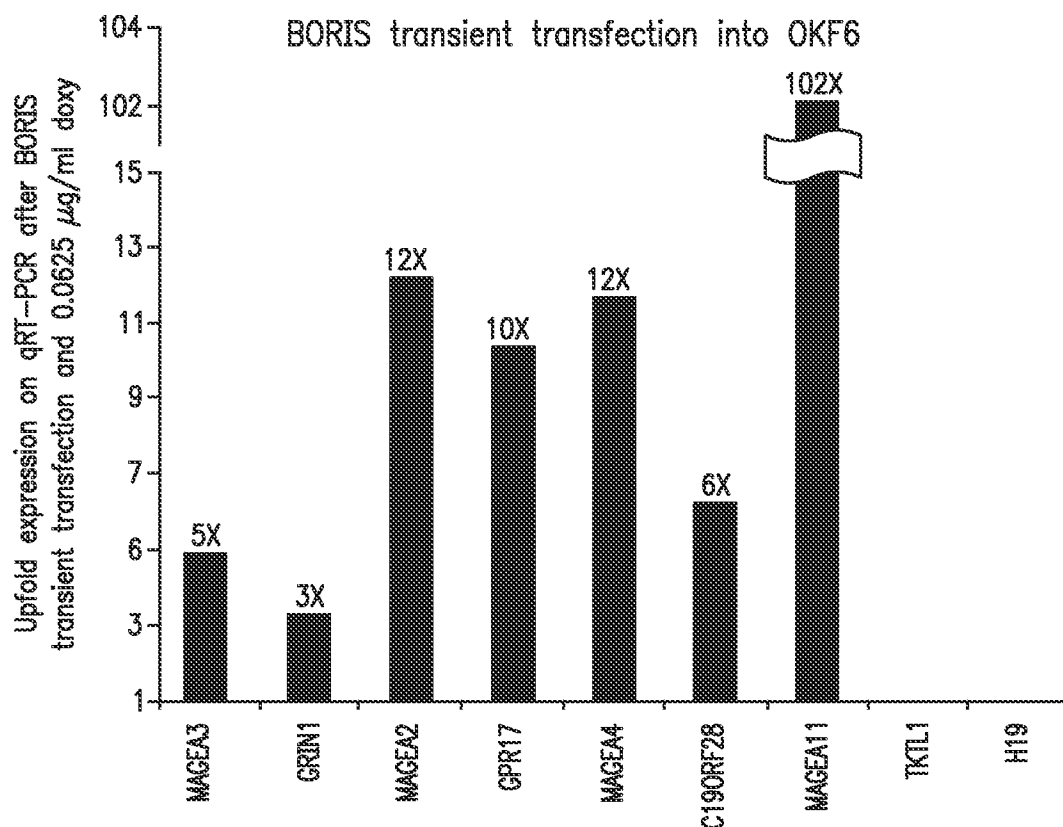
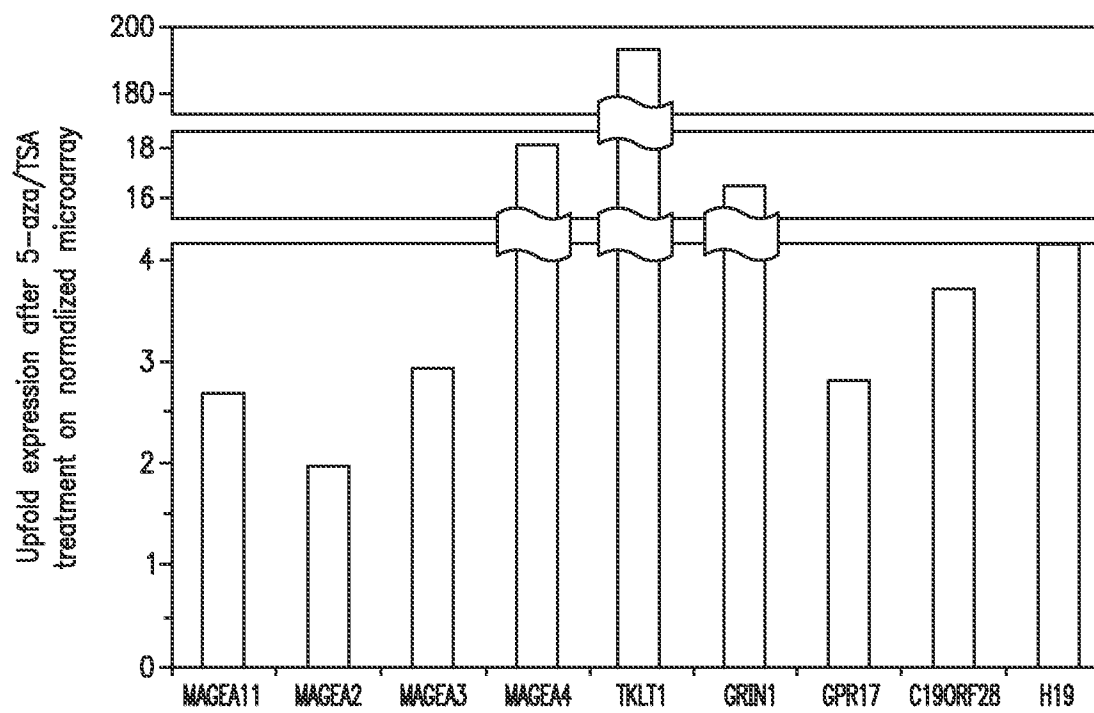


FIG. 7e

25 / 49

**FIG. 8**

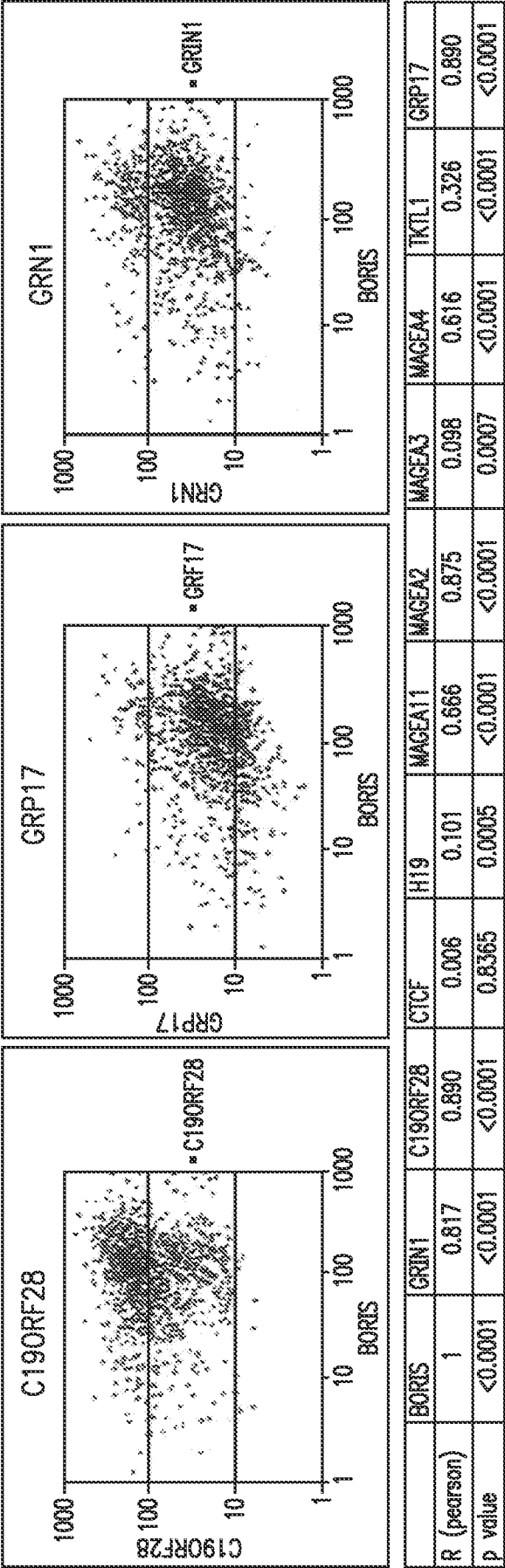


FIG. 9

27 / 49

Integrative Epigenetic Screening Strategy

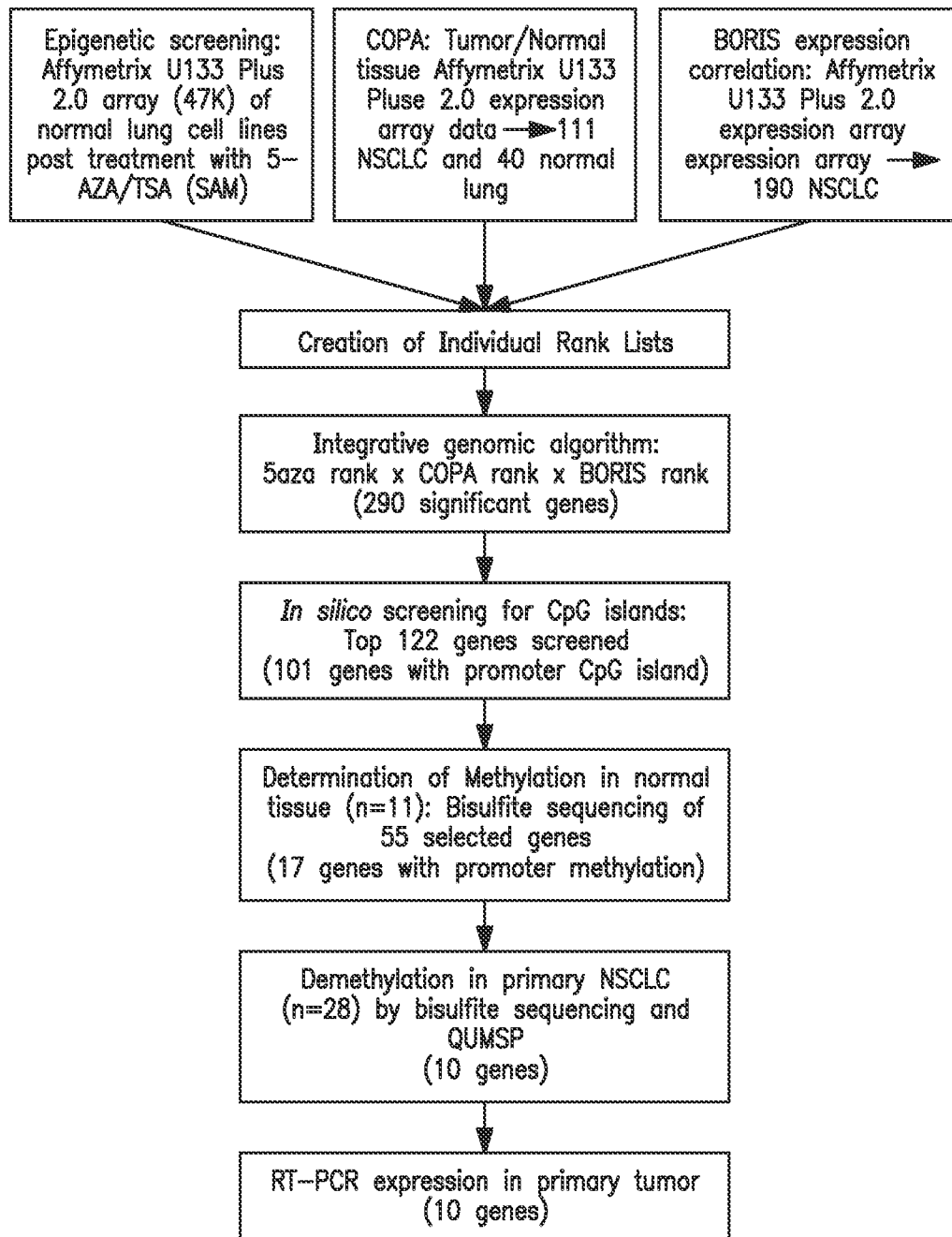


FIG. 10a

28 / 49

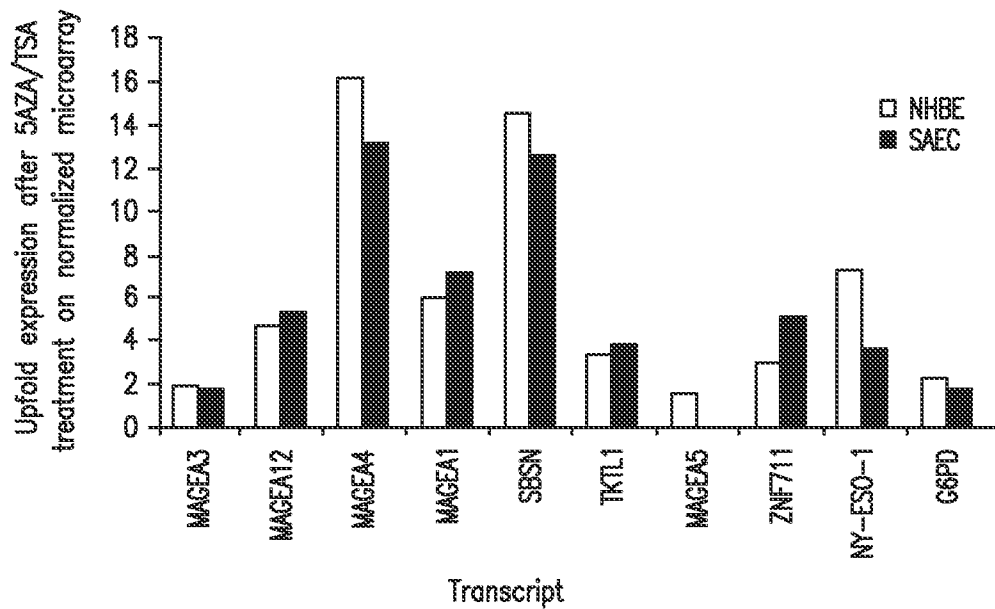


FIG. 10b

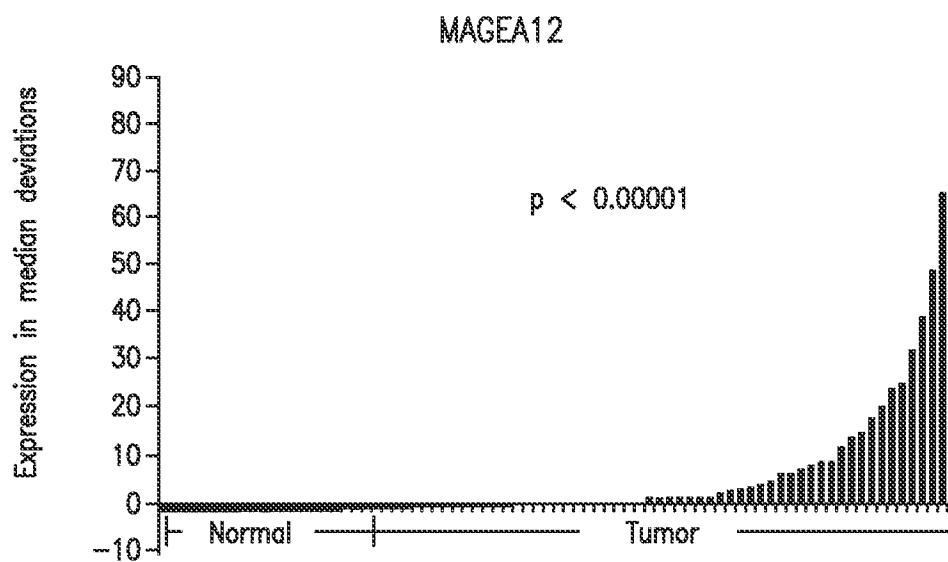


FIG. 10c

29 / 49

	Tumor																Normal																											
Gene	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	P Value				
MAGEA3																																									p=0.0067			
MAGEA12																																									p=0.0001			
MAGEA4																										ND															p=0.0378			
MAGEA1																		ND																							p=0.0001			
SBSN																																									p=0.0067			
TKTL1																					ND																				p=0.2949			
MAGEA5	ND					ND		ND										ND		ND																						p=0.0172		
ZNF711								ND										ND			ND																						p=0.0001	
NY-ESO-1	ND	ND	ND	ND	ND	ND	ND	ND										ND																									p=0.0002	
G6PD	ND	ND	ND	ND	ND	ND	ND	ND										ND																										p=0.0014

FIG. 11

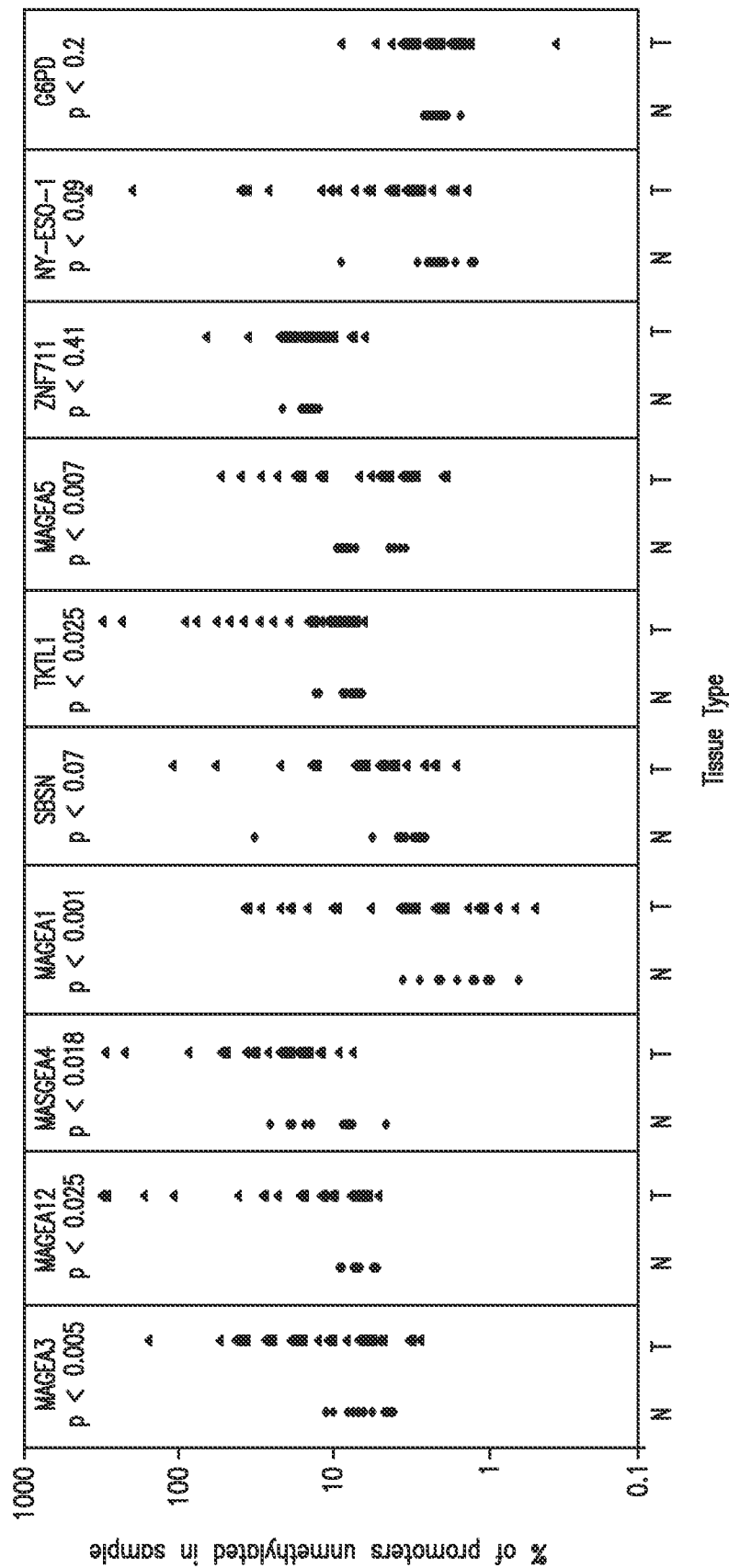


FIG. 12a

31 / 49

	MAGEA3	MAGEA12	MAGEA4	MAGEA1	SBSN	TKTL1	MAGEA5	ZNF711	NY-ESO-1	G6PD
MAGEA3		<0.004	<0.004	<0.004	<0.004	0.004	<0.004	<0.004	0.02	0.02
MAGEA12			<0.004	0.112	<0.004	<0.004	0.004	0.04	<0.004	0.224
MAGEA4				0.024	0.004	<0.004	0.024	0.024	<0.004	0.068
MAGEA1					0.256	0.116	<0.004	<0.004	0.368	0.004
SBSN						0.004	0.012	<0.004	<0.004	0.024
TKTL1							0.028	0.108	<0.004	0.024
MAGEA5								<0.004	0.1	<0.004
ZNF711									0.652	<0.004
NY-ESO-1										0.472
G6PD										

FIG. 12b

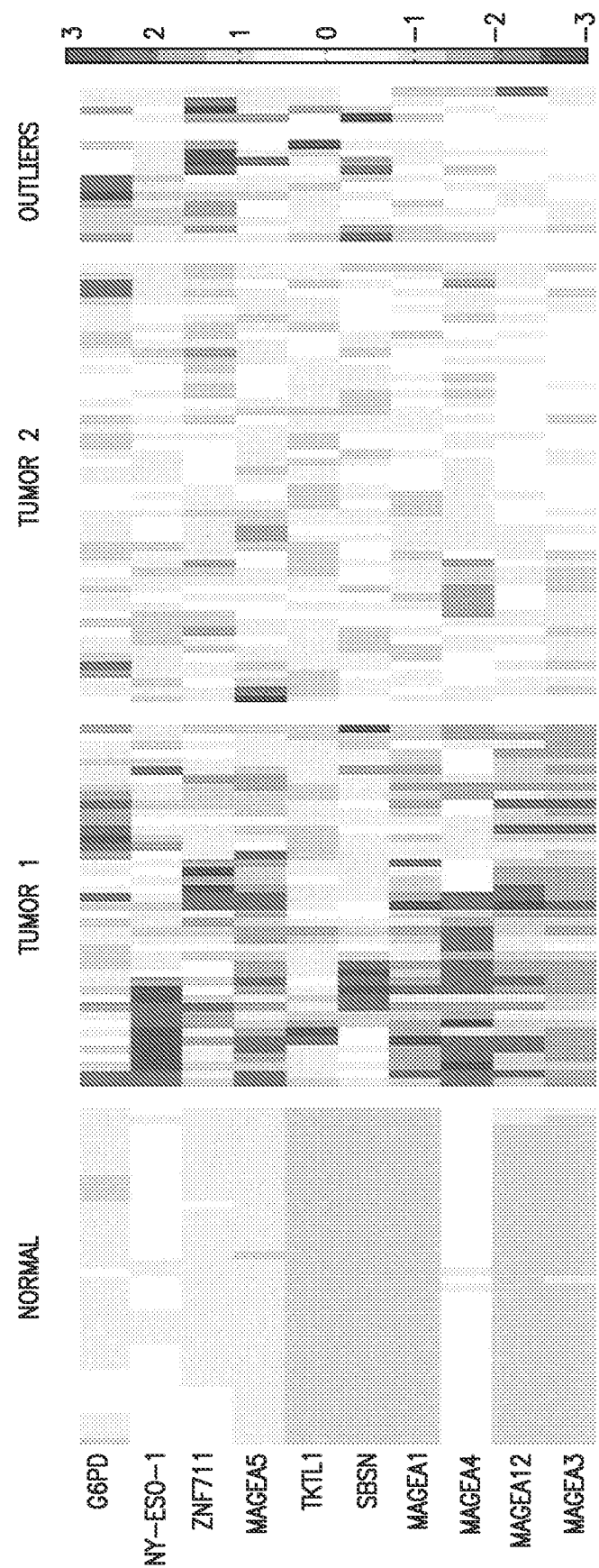


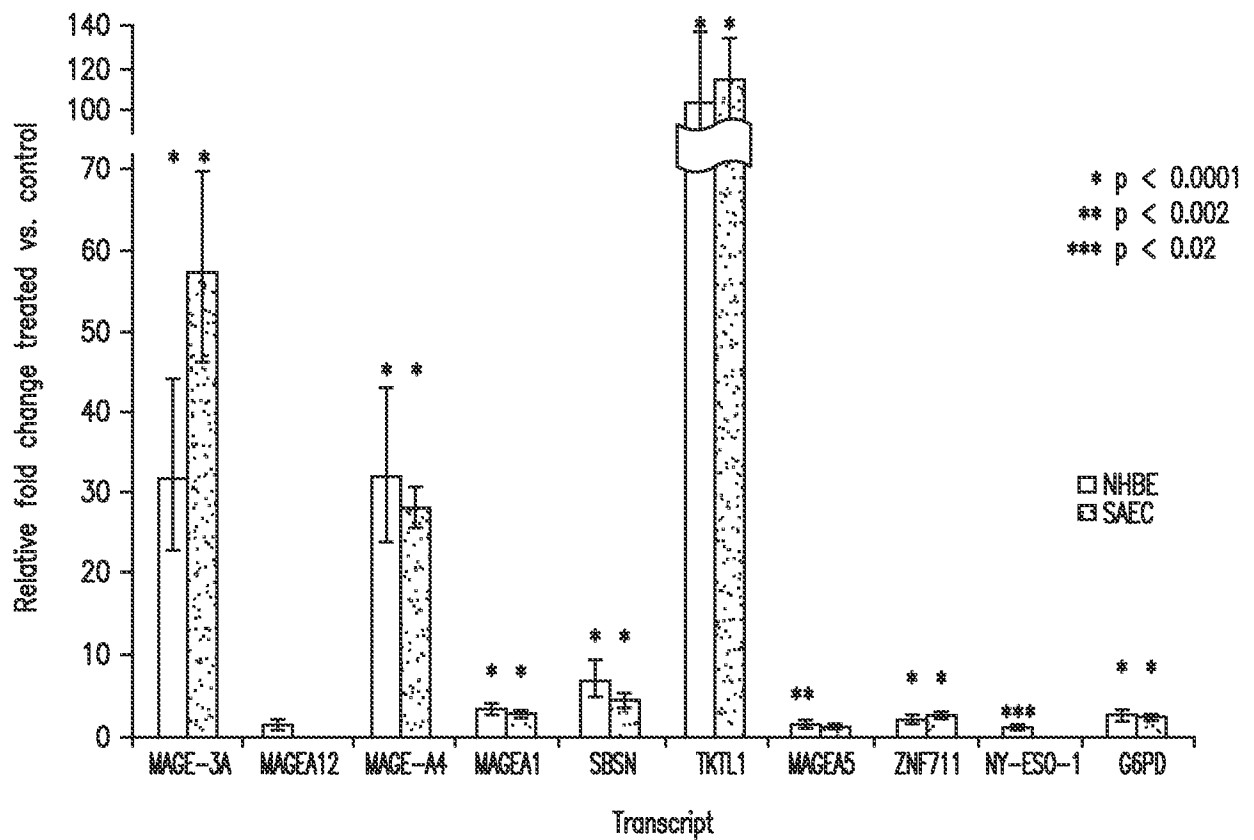
FIG. 13a

33 / 49

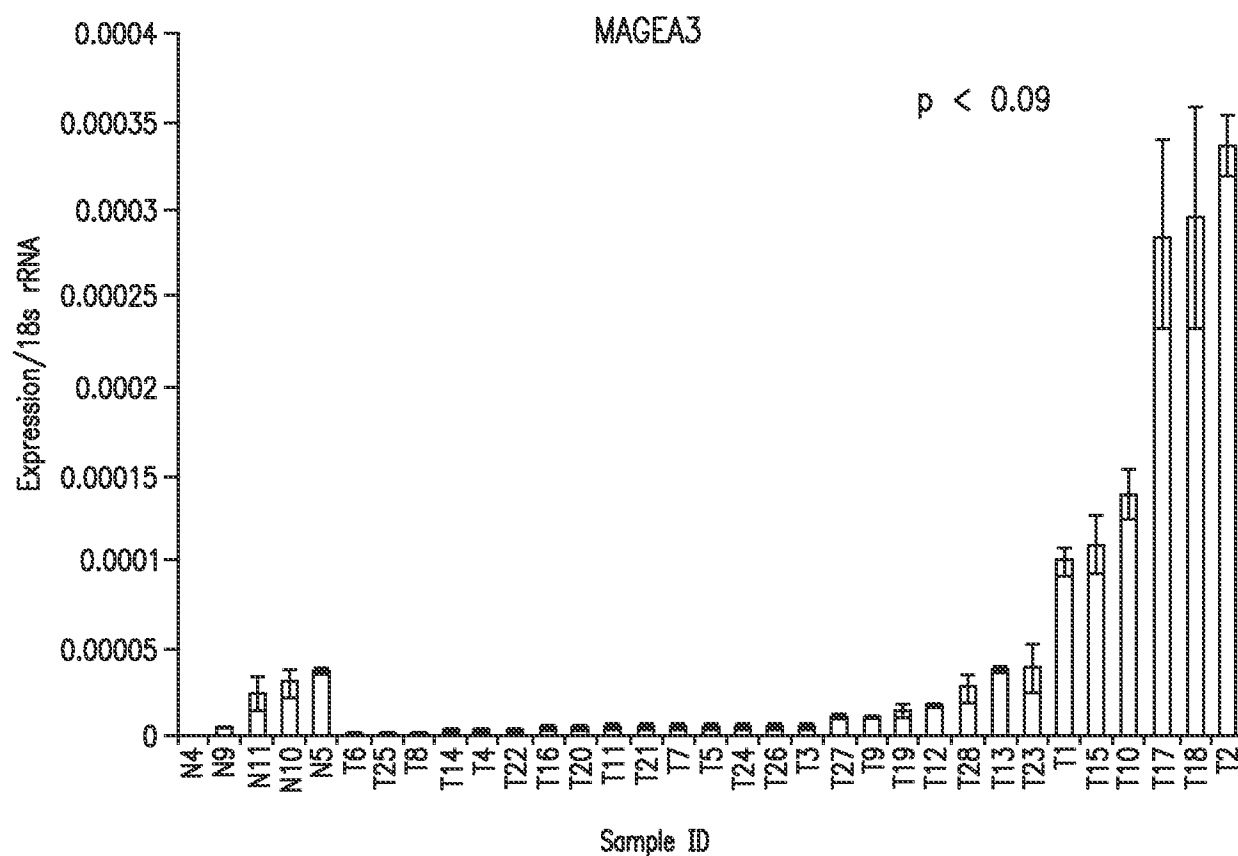
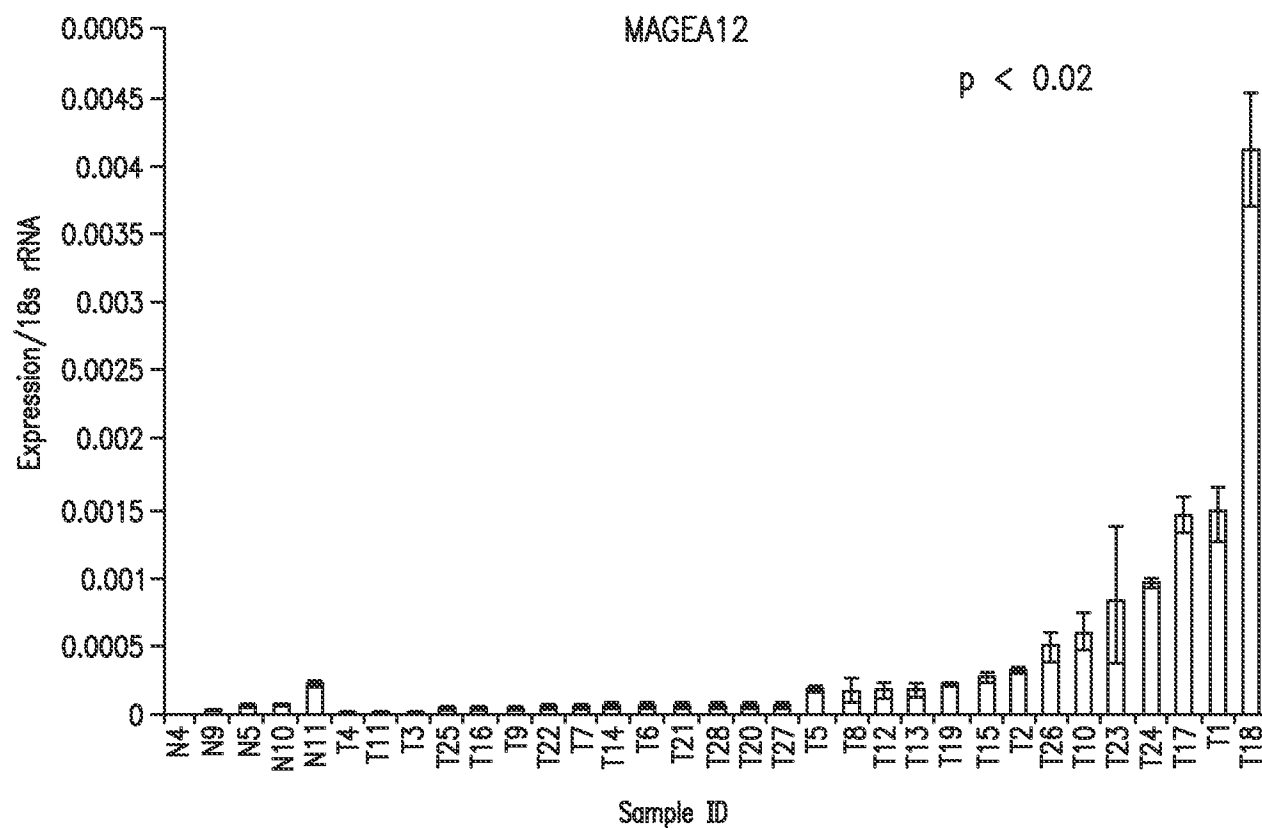
	MAGEA3	MAGEA12	MAGEA4	MAGEA1	SBSN	TKTL1	MAGEA5	ZNF711	NY-ESO-1	G6PD
MAGEA3	1.04E-61	6.40E-13	2.28E-37	2.28E-37	1.54E-11	3.73E-05	3.88E-36	0.00039094	9.32E-07	2.42E-07
MAGEA12	4.64E-63	3.20E-12	1.36E-41	1.36E-41	1.72E-14	5.62E-07	5.68E-35	6.15E-05	1.11E-05	4.92E-06
MAGEA4	1.57E-13	8.54E-13	1.72E-14	1.72E-14	0.0024618	0.076424	0	8.60E-05	4.31E-12	0.019159
MAGEA1	2.03E-38	9.05E-43	3.44E-15	3.44E-15	2.55E-13	4.08E-07	3.65E-26	0.001559	7.80E-08	2.85E-05
SBSN	4.79E-12	3.44E-15	0.00186	5.67E-14	5.94E-07	1.22E-06	3.81E-09	0.23356	0.012302	0.0032751
TKTL1	2.15E-05	2.50E-07	0.069631	1.72E-07	5.94E-07	0.00074306	0.00074306	0.013424	0.013351	0.68006
MAGEA5	4.31E-37	7.57E-36	0	5.68E-27	1.27E-09	0.00051189	0.00051189	0.001132	1.29E-08	4.17E-05
ZNF711	0.00026063	3.83E-05	5.54E-05	0.0011433	0.22318	0.011634	0.00080501	0.14475	0.15509	0.0033554
NY-ESO-1	4.35E-07	5.91E-06	1.24E-12	2.94E-08	0.010115	0.011274	4.60E-09	0.0026843	0.77627	0.77627
G6PD	9.66E-08	2.52E-06	0.01703	1.59E-05	0.0025473	0.66494	2.50E-05	0.0026843	0.77627	0.77627

FIG. 13b

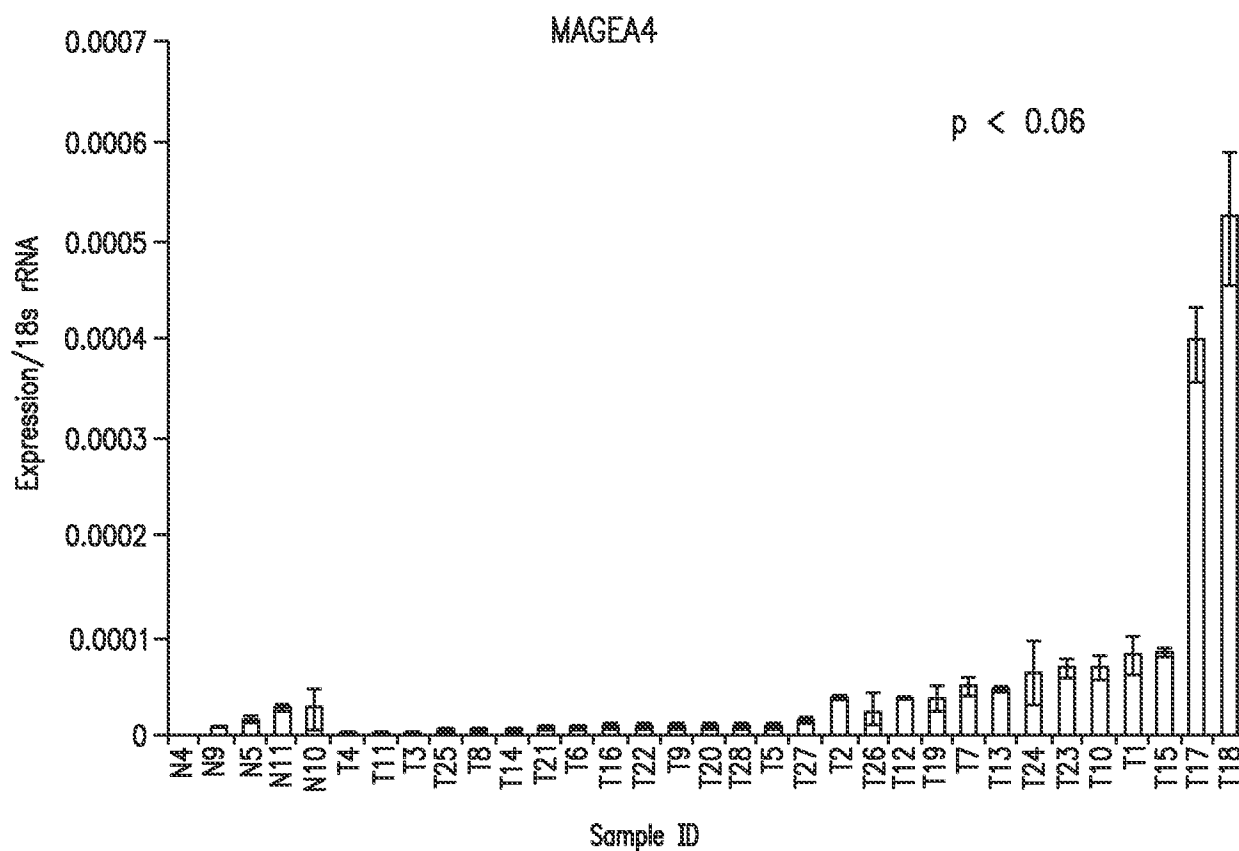
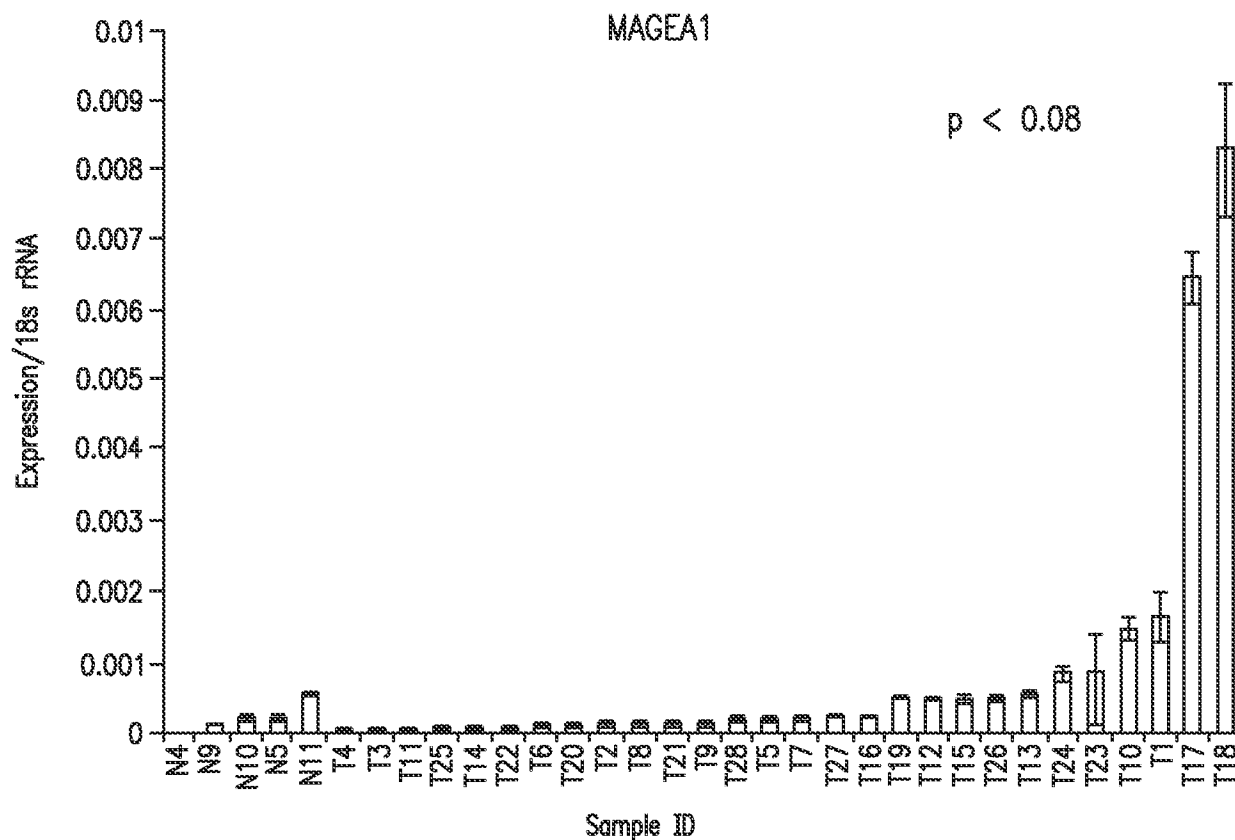
34 / 49

**FIG. 14**

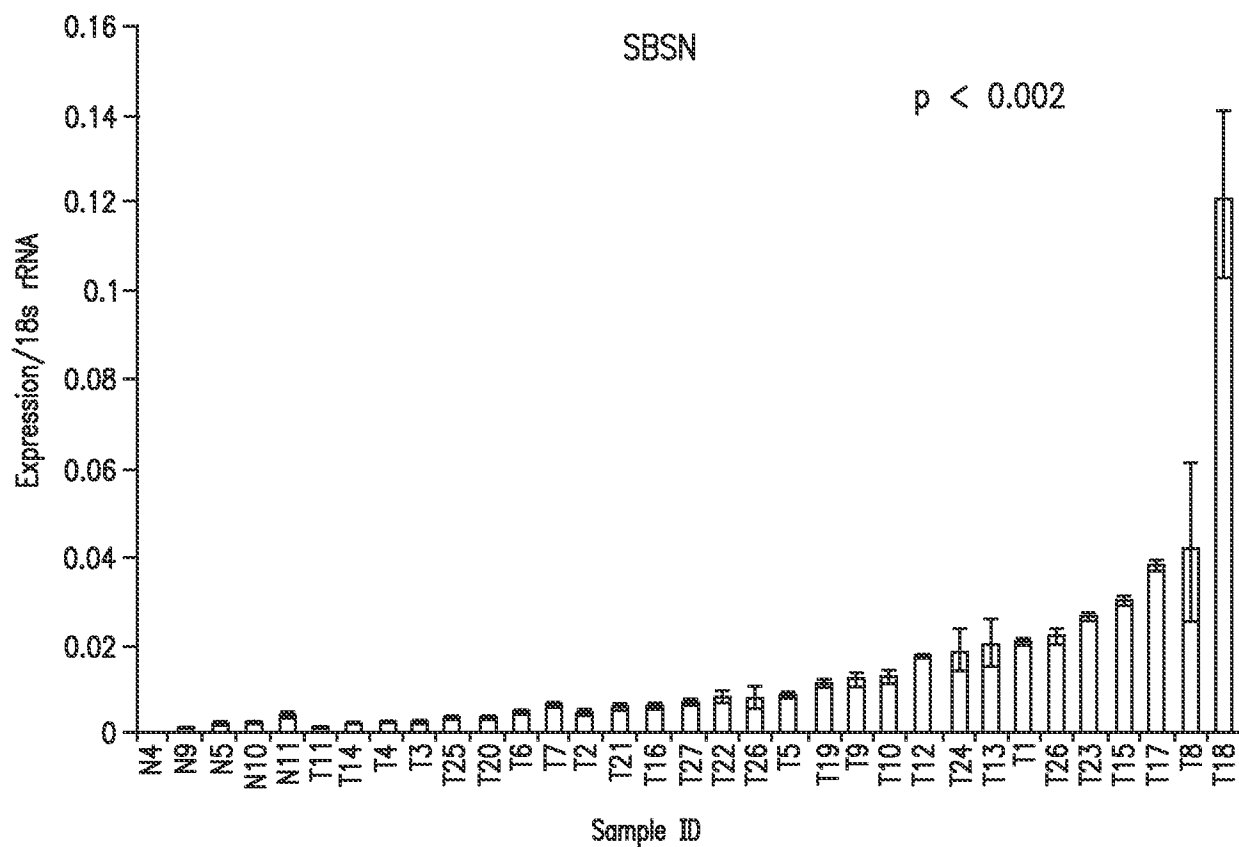
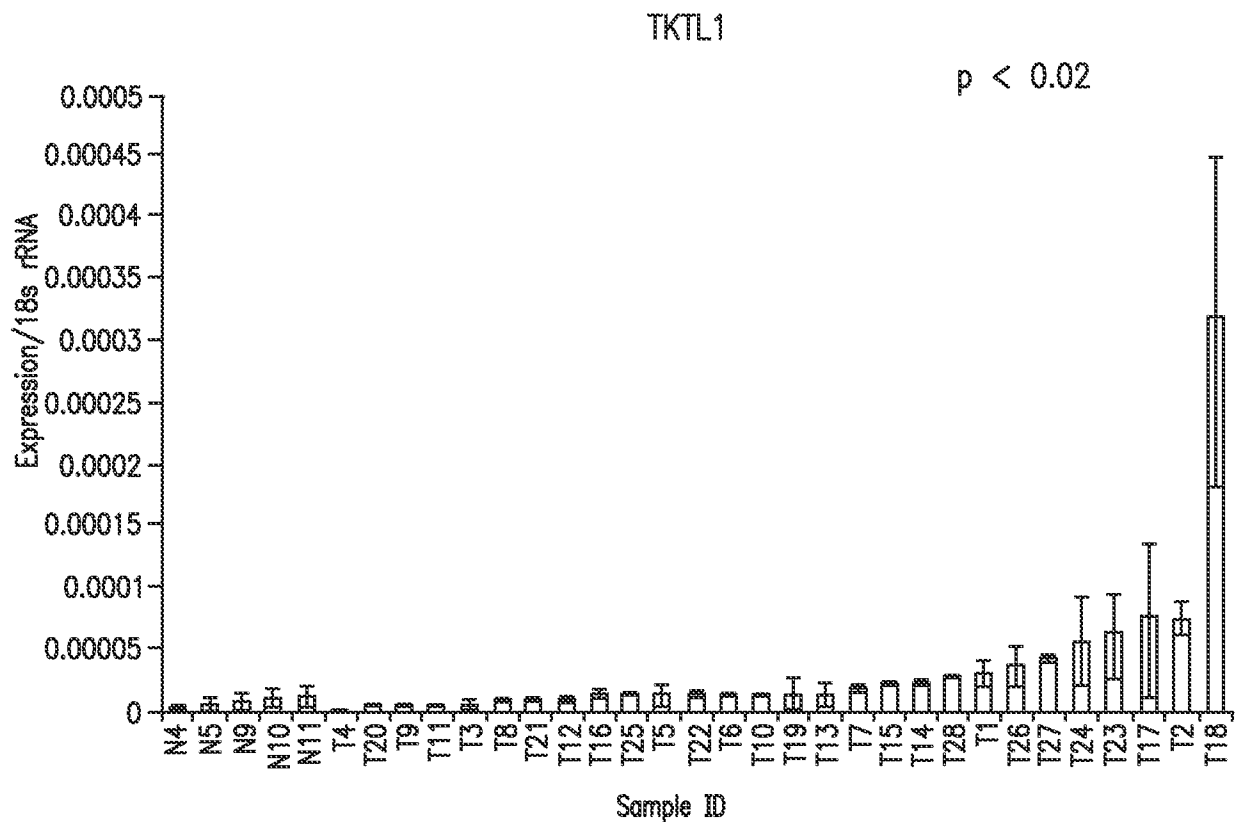
35 / 49

**FIG. 15a****FIG. 15b**

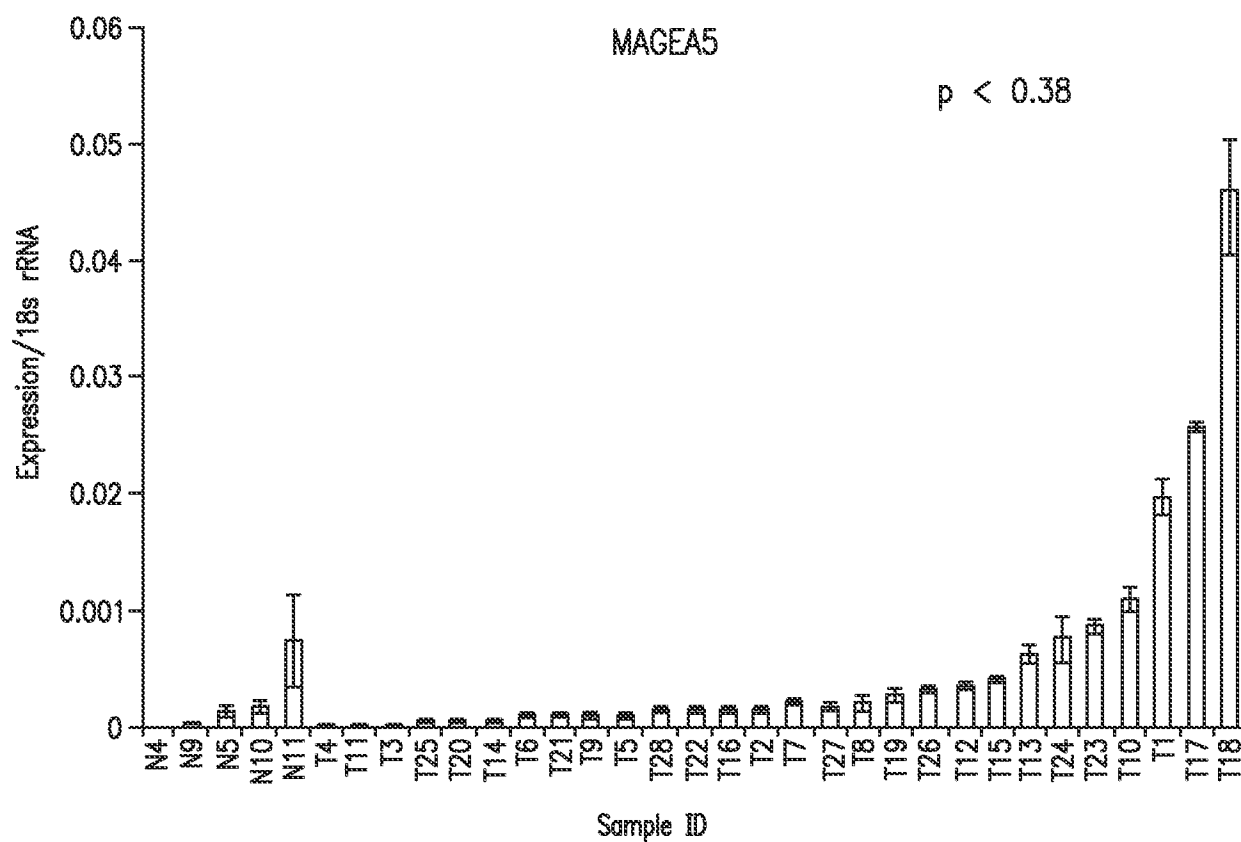
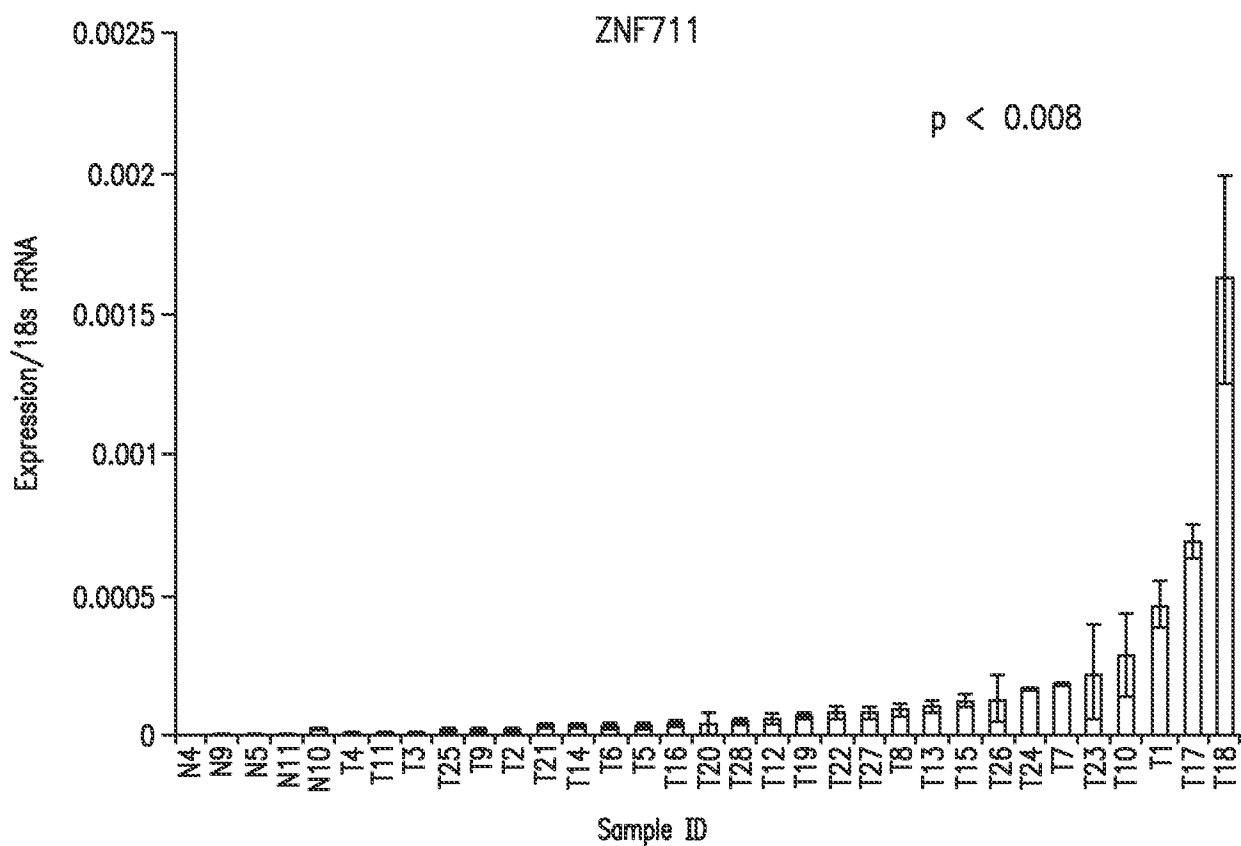
36 / 49

**FIG. 15c****FIG. 15d**

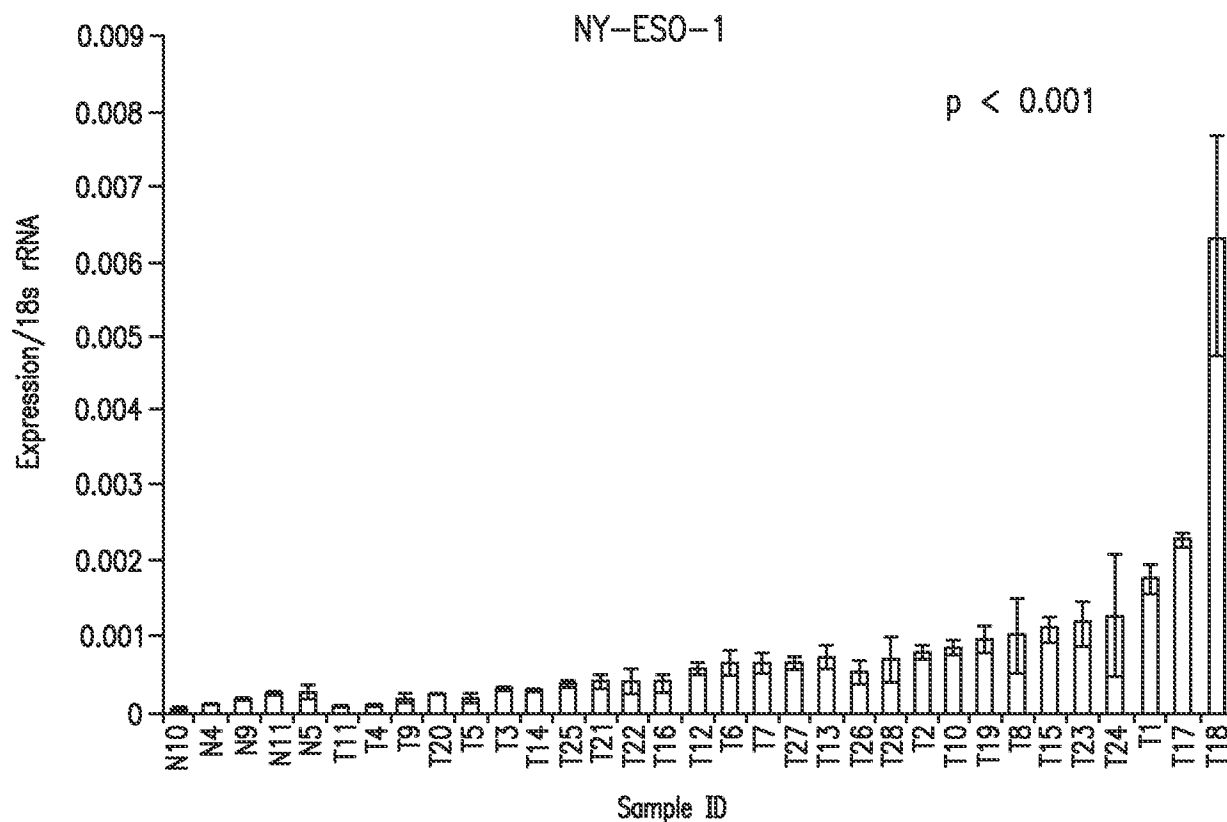
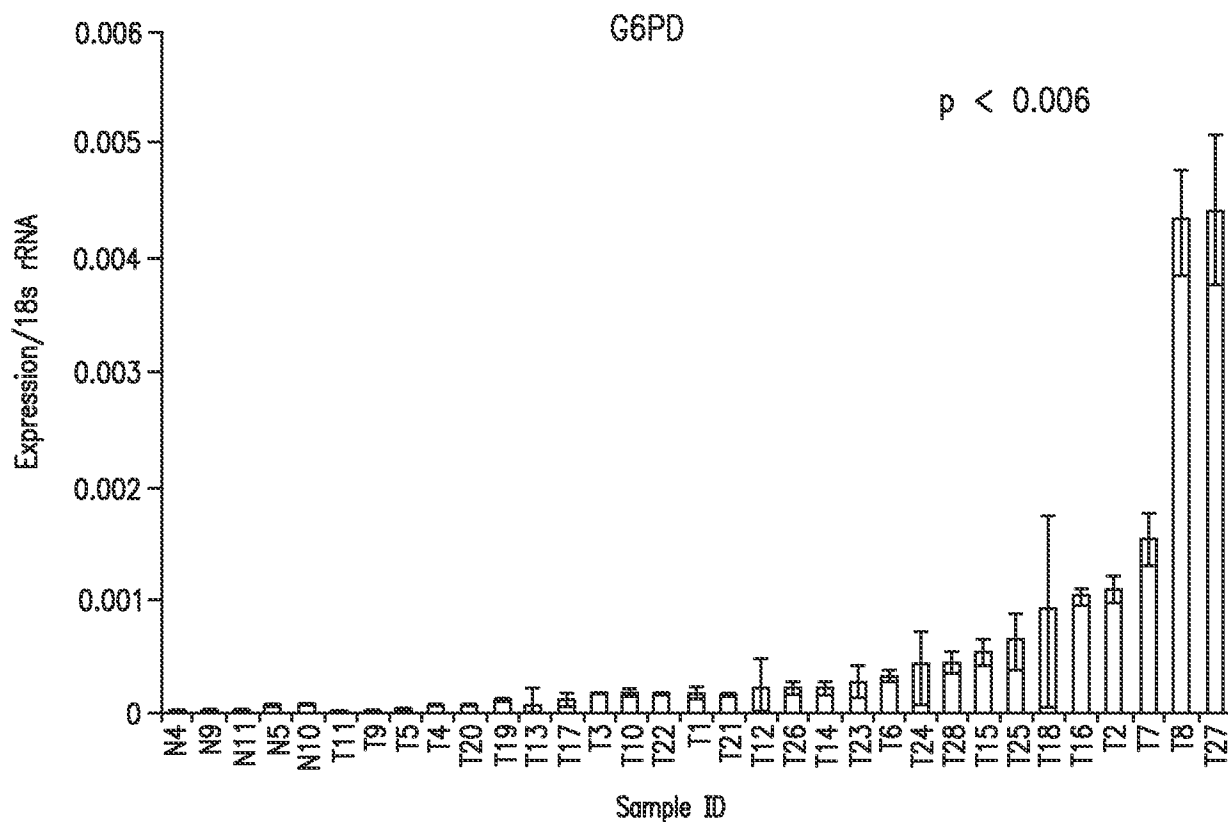
37 / 49

**FIG. 15e****FIG. 15f**

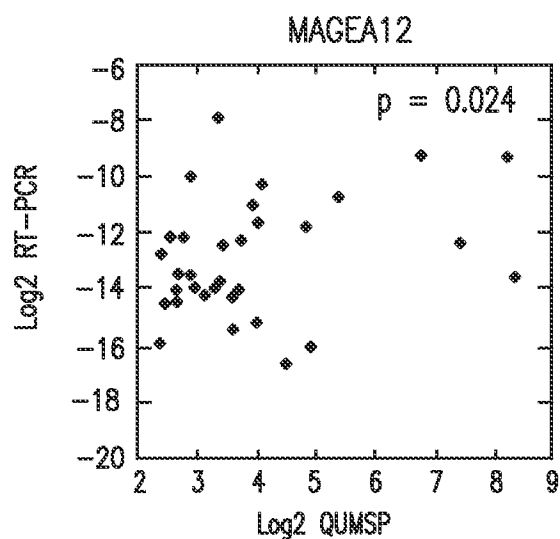
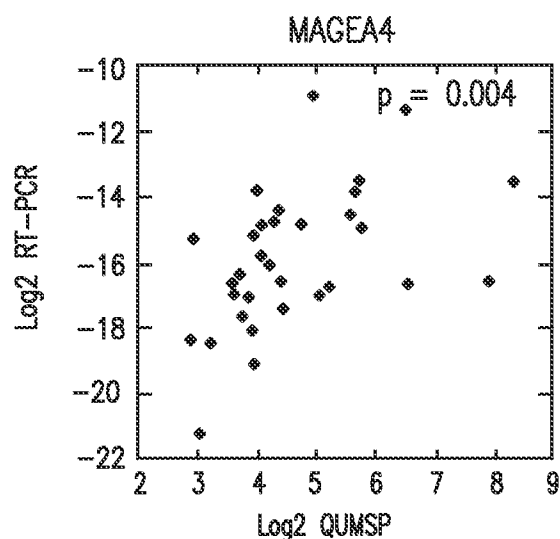
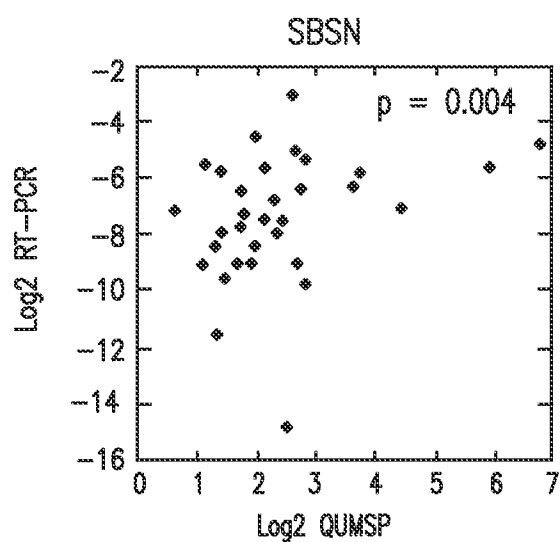
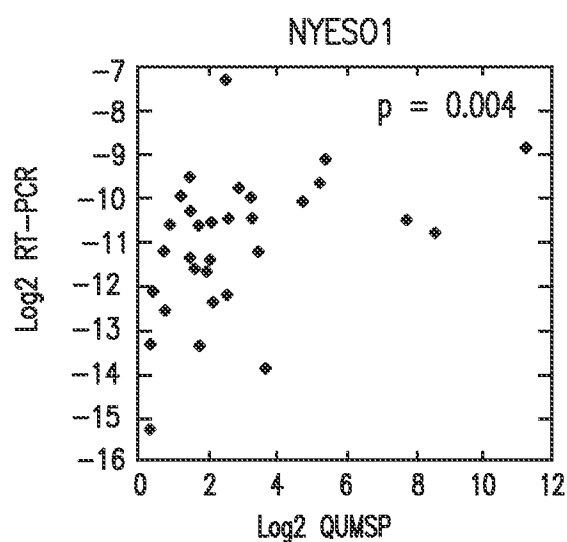
38 / 49

**FIG. 15g****FIG. 15h**

39 / 49

**FIG. 15i****FIG. 15j**

40 / 49

**FIG. 16a****FIG. 16b****FIG. 16c****FIG. 16d**

41 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
1	fibronectin type III and SPRY domain containing 1	NM_024333	yes	no	
2	melanoma antigen family A, 9	BC002351	yes	yes	no
3	hypothetical protein FLJ14627	AW299700	yes	no	
4	melanoma antigen family A, 3	BC000340	yes	yes	yes
5	melanoma antigen family A, 12	BC003408	yes	yes	yes
6	melanoma antigen family A, 4	AW438674	yes	yes	yes
7	melanoma antigen family A, 6	U10691	yes	yes	no
8	HRAS-like suppressor	NM_020386	yes	no	
9	HRAS-like suppressor	NM_020386	yes	no	
10	kinesin family member 1A	AL533416	no		
11	melanoma antigen family A, 1 (directs expression of antigen MZ2-E)	NM_004988	yes	yes	yes
12	Zic family member 2 (odd-paired homolog, Drosophila)	AF193855	yes	no	
13	hypothetical LOC388727	AI382195	yes		
14	T-cell lymphoma invasion and metastasis 1	U90902	yes	no	
15	Interleukin 17 receptor D	AW007080	yes	yes	no
16	cancer/testis antigen CT45-2 /// cancer/testis antigen CT45-5	U31738	yes	yes	no
17	chromosome 7 open reading frame 32	AI884867	yes	no	
18	keratin 23 (histone deacetylase inducible)	NM_015515	no		
19	melanoma antigen family B, 2	NM_002364	yes		
20	small proline-rich protein 1A	AI923984	no		
21	FXFD domain containing ion transport regulator 7	NM_022006	yes		
22	hypothetical protein FLJ38736	Z98443	yes	no	
23	keratin 6B	AI831452	no		
24	troponin T1, skeletal, slow	AJ011712	yes		
25	oxoglutarate dehydrogenase-like	NM_018245	yes	no	
26	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltransferase)	N22272	yes	no	
27	suprabasin	AI814274	yes	yes	yes
28	hypothetical protein FLJ10781	NM_018215	yes	no	
29	ribonuclease, RNase A family, 7	AJ131212	no		
30	tubulin, beta polypeptide paralog	AL533838	yes	no	
31	Clone IMAGE:110436 mRNA sequence	AW444944	no		
32	zinc finger DAZ interacting protein 3	BG502305	yes	no	
33	Ras association (RalGDS/AF-6) domain family 6	AW070877	yes		
34	Adaptor-related protein complex 1, sigma 2 subunit	AA205444	no		
35	transketolase-like 1	Z49258	yes	yes	yes
36	G antigen 4 /// G antigen 5 /// G antigen 6 /// G antigen 7B	NM_001476	yes	yes	no
37	synaptogyrin 3	NM_004209	yes	no	
38	hypothetical protein MGC23280	AI076793	yes	no	
39	reticulon 3	BE544689	yes	no	

FIG. 17

42 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
40	cytochrome P450, family 24, subfamily A, polypeptide 1	NM_000782	yes	no	
41	melanoma antigen family A, 5	AI200443	yes	yes	yes
42	protocadherin 19	AB037734	yes	no	
43	zinc finger protein 6 (ZNF711)	AU157017	yes	yes	yes
44	small proline-rich protein 2A	NM_006945	no		
45	LOC441335	AI343600	yes		
46	fibroblast growth factor 19	AF110400	yes	no	
47	potassium voltage-gated channel, subfamily H (eag-related), member 2	NM_000238	yes		
48	small proline-rich protein 3	BF575466	no		
49	integrin beta 1 binding protein (melusin) 2	NM_012278	yes	yes	no
50	similar to hypothetical protein A230046P18; cDNA sequence BC055759	AW188087	no		
51	calbindin 1, 28kDa	NM_004929	yes		
52	galactokinase 1	BG474736	yes		
53	CDNA FLJ35846 fis, clone TESTI2006877	AK093165	yes		
54	Ring finger protein 32	BE962709	yes	no	
55	Hypothetical protein FLJ13744	NM_025011	yes		
56	dimethylarginine dimethylaminohydrolase 2	NM_013974	yes		
57	G antigen 2 /// G antigen 4 /// G antigen 5 /// G antigen 6 /// G antigen 7 /// G antigen 7B /// G antigen 8	NM_001474	yes		
58	MARCKS-like 1	NM_023009	yes		
59	small nuclear RNA activating complex, polypeptide 3, 50kDa	NM_003084	yes	no	
60	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltransferase)	NM_003360	yes		
61	gb:BF446127 /DB_XREF=gi:11511265 /DB_XREF=7p32f11.x1 /CLONE=IMAGE:3647756 /FEA=EST /CNT=9 /TID=Hs.121866.0 /TIER=ConsEnd /STK=1 /UG=Hs.121866 /UG_TITLE=ESTs	BF446127	yes		
62	peripherin	NM_006262	yes		
63	retinol binding protein 1, cellular	NM_002899	yes		
64	DKFZP564I1171 protein	BF528646	yes		
65	cytidine deaminase	NM_001785	yes		
66	RAB3B, member RAS oncogene family	BC005035	yes		
67	protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regulated phosphoprotein, DARPP-32)	AK024593	yes	no	
68	fatty acid binding protein 7, brain	NM_001446	no		
69	dimethylarginine dimethylaminohydrolase 2	AJ012008	yes		
70	Zinc finger and BTB domain containing 11	BC003066	yes		
71	keratin 17	NM_000422	no		
72	cancer/testis antigen 1B (NY-ESO1)/// cancer/testis antigen 1B /// cancer/testis antigen 1A /// cancer/testis antigen 1A, NY-ESO-1, accession # AF038567	AF038567	yes	yes	yes
73	opposite strand transcription unit to STAG3	AL831967	yes		

FIG. 17 (Continued)

43 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
74	hypothetical protein FLJ14503	AW237462	no		
75	keratin 6B	L42612	no		
76	Homo sapiens, clone IMAGE:5229457, mRNA	BC037977	no		
77	matrix metalloproteinase 13 (collagenase 3) /// matrix metalloproteinase 13 (collagenase 3)	NM_002427	no		
78	ovary-specific acidic protein	AF329088	yes	no	
79	hypothetical protein MGC50844	AK000208	yes		
80	myosin VB	AI991160	yes		
81	keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner)	BC002690	yes		
82	adducin 2 (beta)	NM_017488	yes		
83	Homo sapiens, clone IMAGE:5297356, mRNA	AW665239	yes		
84	hypothetical protein FLJ20489	NM_017842	yes		
85	coagulation factor C homolog, coxlin (Limulus polyphemus)	AA669336	yes		
86	SET and MYND domain containing 3	NM_022743	no		
87	proenkephalin	NM_006211	yes		
88	enolase 2 (gamma, neuronal)	NM_001975	yes		
89	neuronal cell adhesion molecule	NM_005010	yes	no	
90	homeo box A10	AI375919	yes		
91	neurofilament, heavy polypeptide 200kDa	X15306	yes		
92	amine oxidase (flavin containing) domain 1	BE348688	yes		
93	lactotransferrin	NM_002343	yes		
94	neurotensin	NM_006183	no		
95	calbindin 1, 28kDa	AW014927	yes		
96	gb:AI692575 /DB_XREF=gi:4969915 /DB_XREF=wd73h12.x1 /CLONE=IMAGE:2337287 /FEA=EST /CNT=20 /TID=Hs.59761.0 /TIER=Stack /STK=13 /UG=Hs.59761 /UG_TITLE=ESTs, Weakly similar to DAP1_HUMAN DEATH-ASSOCIATED PROTEIN 1 (H.sapiens)	AI692575	no		
97	interleukin 21 receptor	AF269133	yes		
98	keratin 13	NM_002274	no		
99	galanin receptor 2	AF080586	yes	no	
100	GRB2-associated binding protein 2	NM_012296	yes		
101	peptidyl arginine deiminase, type III	NM_016233	-		
102	chromosome 6 open reading frame 148	NM_030568	-		
103	synaptotagmin XI	AA626780	yes		
104	matrix metalloproteinase 10 (stromelysin 2)	NM_002425	yes	yes	no
105	glucose-6-phosphate dehydrogenase	NM_000402	yes	yes	yes
106	neurofilament, heavy polypeptide 200kDa	NM_021076	-		
107	zinc finger protein 38	AL136865	-		
108	gamma-glutamyltransferase-like 3	AL049709	-		
109	GDP-mannose 4,6-dehydratase	NM_001500	-		
110	Homo sapiens, clone IMAGE:4837650, mRNA, partial cds	AI862542	-		

FIG. 17 (Continued)

SUBSTITUTE SHEET (RULE 26)

44 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
111	glycine C-acetyltransferase (2-amino-3-ketobutyrate coenzyme A ligase)	NM_014291	-		
112	adenosine deaminase	X02189	-		
113	similar to hepatocellular carcinoma-associated antigen HCA557b	AI625022	yes	no	
114	brevican	AI739071	-		
115	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	BF434335	-		
116	SRY (sex determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	NM_000346	yes	no	
117	Hypothetical protein FLJ35801	BG260069	-		
118	Kell blood group precursor (McLeod phenotype)	NM_021083	-		
119	keratin associated protein 2-1 /// keratin associated protein 2-4	BC012486	-		
120	actin like protein	BF594459	-		
121	pleckstrin homology domain containing, family G (with RhoGef domain) member 4	AI056683	-		
122	NSE1 /// hypothetical LOC400944	AI601101	-		
123	calmegin	NM_004362	-		
124	hypothetical LOC151760	BC038577	-		
125	potassium voltage-gated channel, subfamily H (eag-related), member 2	AB044806	-		
126	small proline-rich protein 1B (cornifin)	NM_003125	no		
127	UL16 binding protein 2	AA831769	-		
128	hypothetical protein MGC4825	AF061264	-		
129	Transcribed locus, weakly similar to XP_512006.1 similar to FLJ46489 protein [Pan troglodytes]	AA629068	-		
130	quiescin Q6	NM_002826	yes		
131	hypothetical protein LOC196394	BG031897	-		
132	protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), gamma isoform	AI669212	-		
133	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	S36219	-		
134	CDNA clone IMAGE:4620359, partial cds	BC023568	-		
135	hypothetical LOC344595	AA463827	-		
136	hypothetical protein FLJ32942	BG150433	-		
137	Ras association (RalGDS/AF-6) domain family 6	AI167789	yes	no	
138	Hypothetical protein FLJ10385	U58658	-		
139	ephrin-A3	AW189015	yes	no	
140	homeo box A1	S79910	-		
141	melanoma antigen family F, 1	NM_022149	yes	no	
142	hydroxysteroid (17-beta) dehydrogenase 2	NM_002153	-		
143	lipidosin	BE856376	-		
144	hypothetical protein FLJ23322	NM_024955	-		

FIG. 17 (Continued)

45 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
145	gb:BF241692 /DB_XREF=gi:11155618 /DB_XREF=601879820F1 /CLONE=IMAGE:4108469 /FEA=EST /CNT=9 /TID=Hs.124839.0 /TIER=ConsEnd /STK=0 /UG=Hs.124839 /UG_TITLE=ESTs	BF241692	-		
146	hypothetical protein LOC112868	BF447901	-		
147	protein phosphatase 1, regulatory (inhibitor) subunit 1A	NM_006741	-		
148	B-box and SPRY domain containing	AJ276691	-		
149	endothelin 1	BC036851	yes	no	
150	hypothetical protein LOC285590	BC038561	-		
151	distal-less homeobox 4	NM_001934	-		
152	transcription factor Dp-2 (E2F dimerization partner 2)	NM_006286	-		
153	gb:Nm_022443.1 /DB_XREF=gi:11967974 /GEN=MLF1 /FEA=FLmRNA /CNT=64 /TID=Hs.85195.0 /TIER=ConsEnd /STK=0 /UG=Hs.85195 /LL=4291 /DEF=Homo sapiens myeloid leukemia factor 1 (MLF1), mRNA. /PROD=myeloid leukemia factor 1 /FL=gb:Nm_022443.1	NM_022443	-		
154	dispatched homolog 2 (Drosophila)	AB051529	-		
155	gb:BF511718 /DB_XREF=gi:11595016 /DB_XREF=UI-H-BI4-aom-f-07-0-UI.s1 /CLONE=IMAGE:3085572 /FEA=FLmRNA /CNT=40 /TID=Hs.99034.0 /TIER=Stack /STK=19 /UG=Hs.99034 /LL=8153 /UG_GENE=RHO7 /UG_TITLE=GTP-binding protein Rho7 /FL=gb:Nm_005440.2	BF511718	-		
156	WNK lysine deficient protein kinase 3	H06509	-		
157	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)	BF725250	-		
158	LOC441297	AI870903	-		
159	KIAA1967	BC003172	-		
160	neurofilament, light polypeptide 68kDa	AL537457	-		
161	immunoglobulin superfamily, member 9	AB037776	-		
162	family with sequence similarity 50, member B	NM_012135	-		
163	protein phosphatase 1J (PP2C domain containing)	AW117553	-		
164	dimethylarginine dimethylaminohydrolase 2	AK026191	-		
165	brain-specific angiogenesis inhibitor 2	NM_001703	-		
166	hairy and enhancer of split 2 (Drosophila)	AK023754	-		
167	Homeo box D13	AI971104	-		
168	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	BE613133	-		
169	adenosine deaminase	NM_000022	-		
170	membrane protein, palmitoylated 2 (MAGUK p55 subfamily member 2)	NM_005374	-		

FIG. 17 (Continued)

46 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
171	quinolinate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (carboxylating))	NM_014298	-		
172	cholinergic receptor, nicotinic, alpha polypeptide 3	BC000513	-		
173	Hypothetical protein FLJ14627	AL120375	-		
174	Insulin-like growth factor binding protein-like 1	AL522781	-		
175	RAB3B, member RAS oncogene family	NM_002867	yes	no	
176	keratin 8 /// keratin 8	U76549	-		
177	keratin 17	Z19574	-		
178	acid phosphatase-like 2	AW069729	-		
179	glutamate decarboxylase 1 (brain, 67kDa)	NM_000817	-		
180	transmembrane protease, serine 11E	NM_014058	-		
181	hypothetical protein LOC112868	AI935915	-		
182	cytochrome P450, family 4, subfamily F, polypeptide 3	NM_000896	-		
183	histone deacetylase 5	NM_005474	yes	no	
184	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5	BC001875	-		
185	RNA-binding region (RNP1, RRM) containing 1 /// RNA-binding region (RNP1, RRM) containing 1	AL109955	-		
186	cadherin 2, type 1, N-cadherin (neuronal)	M34064	-		
187	cyclin-dependent kinase 5	NM_004935	yes	no	
188	v-myb myeloblastosis viral oncogene homolog (avian)	NM_005375	yes	no	
189	Zinc finger protein 192	AI920953	-		
190	Transmembrane, prostate androgen induced RNA	AI821781	-		
191	Wiskott-Aldrich syndrome-like	BF438330	yes	no	
192	Full-length cDNA clone CS0DK010YC18 of HeLa cells Cot 25-normalized of Homo sapiens (human)	AI521166	-		
193	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	NM_004181	-		
194	dopa decarboxylase (aromatic L-amino acid decarboxylase)	NM_000790	-		
195	ankyrin repeat and SOCS box-containing 2	AI872284	-		
196	Full length insert cDNA clone YR23D07	AI919519	-		
197	Doublecortin and CaM kinase-like 1	AI129626	-		
198	nicotinamide nucleotide adenyltransferase 2	AF288395	-		
199	growth differentiation factor 11	AF028333	-		
200	lipid phosphate phosphatase-related protein type 2	NM_022737	-		
201	kinesin light chain 2-like	AI279514	-		
202	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1	AI220427	-		

FIG. 17 (Continued)

47 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
203	SH3-domain binding protein 4	AF015043	-		
204	Neurofilament, light polypeptide 68kDa	BF055311	-		
205	zinc finger protein 42	AF450454	-		
206	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	U07821	-		
207	glycine amidinotransferase (L-arginine:glycine amidinotransferase)	X86401	-		
208	normal mucosa of esophagus specific 1	AF228422	-		
209	testis expressed sequence 15	AL133653	yes	no	
210	Fraser syndrome 1	AL157471	-		
211	cancer/testis antigen 1B	AJ275978	yes		
212	hypothetical gene supported by BC013438	AI569804	-		
213	NAD(P)H dehydrogenase, quinone 2	NM_000904	yes		
214	Full-length cDNA clone CS0DF012YG01 of Fetal brain of Homo sapiens (human)	BG260087	-		
215	DNA segment on chromosome 4 (unique) 234 expressed sequence	NM_014392	-		
216	hypothetical protein FLJ23033	NM_024686	-		
217	Phosphatidylinositol-specific phospholipase C, X domain containing 2	AW170015	-		
218	small proline-rich protein 1A	NM_005987	-		
219	gb:AW274658 /DB_XREF=gi:6661688 /DB_XREF=xv32b10.x1 /CLONE=IMAGE:2814811 /TID=Hs.2.28714.1 /CNT=12 /FEA=mRNA /TIER=ConsEnd /STK=1 /UG=Hs.28714 /UG_TITLE=Homo sapiens, clone IMAGE:5284550, mRNA	AW274658	-		
220	tumor suppressor candidate 3	AI884858	-		
221	signal transducer and activator of transcription 1, 91kDa	NM_007315	-		
222	chromosome X open reading frame 48	NM_017863	-		
223	B-box and SPRY domain containing	NM_017688	-		
224	triple functional domain (PTPRF interacting)	BF223718	-		
225	acheron	BF792126	-		
226	tubulin, beta 4	AL567012	-		
227	guanylate cyclase activator 1A (retina)	L36861	-		
228	hypothetical protein FLJ22595	NM_025047	-		
229	cornifelin /// cornifelin	AB049591	-		
230	glutathione peroxidase 2 (gastrointestinal)	NM_002083	-		
231	glutaminase 2 (liver, mitochondrial)	NM_013267	-		
232	SLAM family member 9	NM_033438	-		
233	KIAA1211 protein	BE855799	-		
234	Retinol binding protein 1, cellular	AI802099	-		
235	hypothetical protein FLJ38736	AW299463	-		
236	nerve growth factor receptor (TNFR superfamily, member 16)	NM_002507	-		
237	Similar to expressed sequence AI836003	AI971535	-		
238	hypothetical protein LOC256021	AK055439	-		
239	chromogranin B (secretogranin 1)	NM_001819	-		

FIG. 17 (Continued)

48 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
240	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	NM_002575	-		
241	hypothetical protein FLJ30707	AW450586	-		
242	translocase of outer mitochondrial membrane 70 homolog A (yeast)	NM_014820	yes	no	
243	DKFZP434B0335 protein	BE616972	-		
244	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	AI760760	-		
245	dehydrogenase/reductase (SDR family) X-linked	AL547782	-		
246	glucosidase, beta, acid 3 (cytosolic)	NM_020973	-		
247	serine protease inhibitor, Kazal type 1	NM_003122	-		
248	Ribosomal protein L4	BF431260	-		
249	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	NM_000962	-		
250	Myeloid leukemia factor 1	AI911434	-		
251	hypothetical protein LOC254128	BF939830	-		
252	hypothetical protein LOC283901	AK095480	-		
253	protein phosphatase 1H (PP2C domain containing)	AB032983	-		
254	carnitine palmitoyltransferase 1C	AL565745	-		
255	cytochrome P450, family 39, subfamily A, polypeptide 1	AI796334	-		
256	carboxypeptidase E	AI922855	-		
257	glutathione peroxidase 7	AA406605	-		
258	ovostatin 2	AW594320	yes		
259	WD repeat domain 9	AI638279	-		
260	phospholipase D1, phosphatidylcholine-specific	U38545	-		
261	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)	BE866412	-		
262	phorbol-12-myristate-13-acetate-induced protein 1	NM_021127	-		
263	piggyBac transposable element derived 5	NM_024554	-		
264	Kringle containing transmembrane protein 1	BF221745	-		
265	tripartite motif-containing 37	AK022701	-		
266	calmodulin-like 5	NM_017422	-		
267	chromosome 9 open reading frame 157	BC031861	-		
268	nuclear factor (erythroid-derived 2)-like 3	NM_004289	-		
269	gb:AI436290 /DB_XREF=gi:4309151 /DB_XREF=th81c01.x1 /CLONE=IMAGE:2125056 /FEA=EST /CNT=3 /TID=Hs.164162.0 /TIER=ConsEnd /STK=3 /UG=Hs.164162 /UG_TITLE=ESTs, Moderately similar to AF161552.1 HSPC067 (H.sapiens)	AI436290	-		
270	myo-inositol 1-phosphate synthase A1	AL137749	-		
271	PTEN induced putative kinase 1	BF432478	-		

FIG. 17 (Continued)

49 / 49

Combined Rank	Gene Name	Accession	Promoter CpG Island	Methylated in Normals	Demethylated in Tumors
272	gb:BE615699 /DB_XREF=gi:9897298 /DB_XREF=601279805F1 /CLONE=IMAGE:3621960 /FEA=EST /CNT=17 /TID=Hs.7531.1 /TIER=Stack /STK=15 /UG=Hs.7531 /LL=23353 /UG_GENE=KIAA0810 /UG_TITLE=KIAA0810 protein	BE615699	-		
273	proopiomelanocortin (adrenocorticotropin/ beta-lipotropin/ alpha-melanocyte stimulating hormone/ beta-melanocyte stimulating hormone/ beta-endorphin)	NM_000939	-		
274	testes-specific protease 50	NM_013270	yes	no	
275	IGF-II mRNA-binding protein 3	NM_006547	-		
276	homeo box D10	AW299531	-		
277	elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 3	AF292387	-		
278	KIAA1718 protein	AK026372	-		
279	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	NM_000962	-		
280	cortixin 1	BF982289	-		
281	progesterone-associated endometrial protein (placental protein 14, pregnancy-associated endometrial alpha-2-globulin, alpha uterine protein)	NM_002571	-		
282	ubiquitin specific protease 2	AW274034	-		
283	Transcription factor CP2-like 3	AI224578	-		
284	gb:AL049569 /DB_XREF=gi:5263031 /FEA=FLmRNA /CNT=48 /TID=Hs.33455.0 /TIER=Stack /STK=8 /UG=Hs.33455 /LL=11240 /UG_GENE=PD12 /UG_TITLE=peptidyl arginine deiminase, type II /DEF=Human DNA sequence from clone RP1-37C10 on chromosome 1p35.2-35.21. Contains th	AL049569	-		
285	cytoplasmic linker 2 /// cytoplasmic linker 2	BC006259	-		
286	gb:AI743489 /DB_XREF=gi:5111777 /DB_XREF=wf63g11.x2 /CLONE=IMAGE:2360324 /FEA=EST /CNT=13 /TID=Hs.294133.0 /TIER=ConsEnd /STK=7 /UG=Hs.294133 /UG_TITLE=ESTs	AI743489	-		
287	hypothetical locus LOC283861	BC009880	-		
288	G protein-coupled receptor 160	BC000181	-		
289	inorganic pyrophosphatase 2	BF446912	-		
290	glutamate decarboxylase 1 (brain, 67kDa)	NM_013445	-		

FIG. 17 (Continued)