



US 20160326593A1

(19) **United States**

(12) **Patent Application Publication**
Clement et al.

(10) **Pub. No.: US 2016/0326593 A1**

(43) **Pub. Date: Nov. 10, 2016**

(54) **COMPOSITIONS AND METHODS FOR
DIAGNOSING, EVALUATING AND
TREATING CANCER**

Related U.S. Application Data

(60) Provisional application No. 61/908,316, filed on Nov. 25, 2013.

(71) Applicants: **The Broad Institute Inc.**, Cambridge, MA (US); **Dana-Farber Cancer Institute, Inc.**, Boston, MA (US); **The General Hospital Corporation**, Boston, MA (US); **President and Fellows of Harvard College**, Cambridge, MA (US)

Publication Classification

(51) **Int. Cl.**
C12Q 1/68 (2006.01)
G06F 19/22 (2006.01)
(52) **U.S. Cl.**
CPC **C12Q 1/6886** (2013.01); **G06F 19/22** (2013.01); **C12Q 2600/154** (2013.01); **C12Q 2600/106** (2013.01); **C12Q 2600/156** (2013.01); **C12Q 2600/118** (2013.01); **C12Q 2600/136** (2013.01)

(72) Inventors: **Mark Kendell Clement**, Cambridge, MA (US); **Gad GETZ**, Cambridge, MA (US); **Dan-Avi LANDAU**, Cambridge, MA (US); **Alexander MEISSNER**, Brookline, MA (US); **Catherine Ju-Ying WU**, Brookline, MA (US)

ABSTRACT

(57) The present invention relates to methods of determining a cancer treatment prognosis for a subject in need thereof by evaluating epigenetic and genetic changes within a tumor sample from the subject. The present invention further provides methods of treating cancer in a subject by evaluating epigenetic and genetic changes within a tumor sample from the subject. In addition, the present invention provides methods of screening test agents to identify agents that decrease tumor cell plasticity.

(21) Appl. No.: **15/038,504**

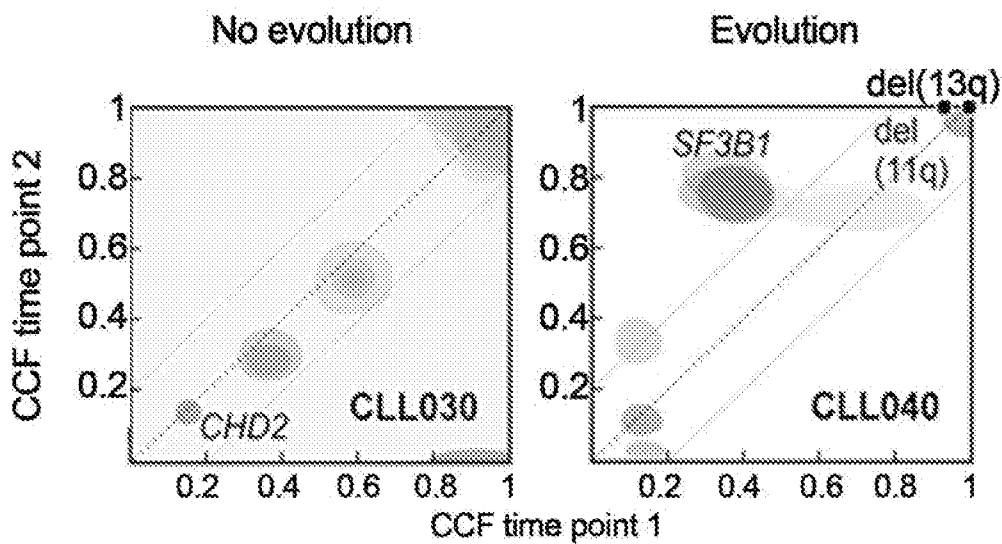
(22) PCT Filed: **Nov. 24, 2014**

(86) PCT No.: **PCT/US2014/067146**

§ 371 (c)(1),

(2) Date: **May 23, 2016**

FIG. 1A



CLLs	No evolution	Evolution
No intervening treatment	5/6	1/6
+ intervening treatment	2/12	10/12

FIG. 1B

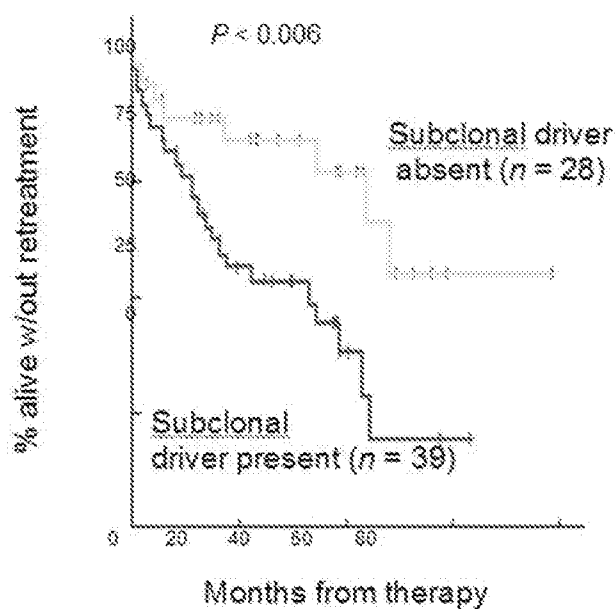


FIG. 2A

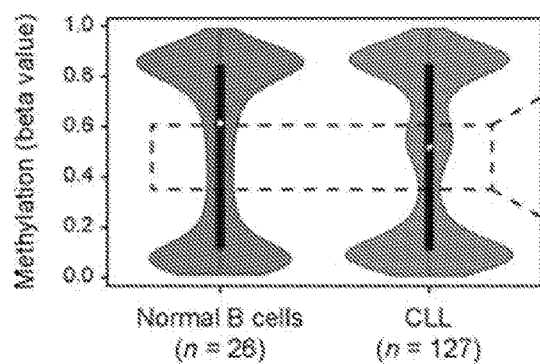


FIG. 2B

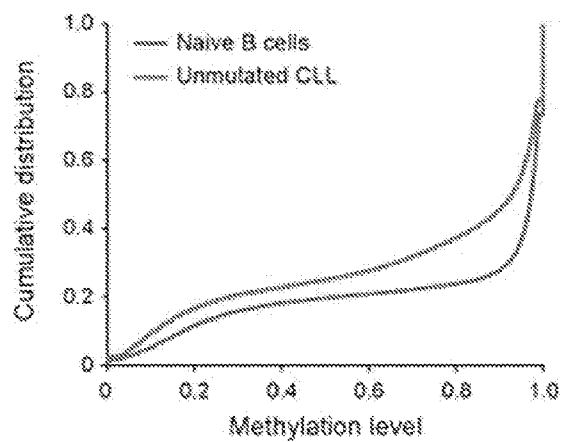
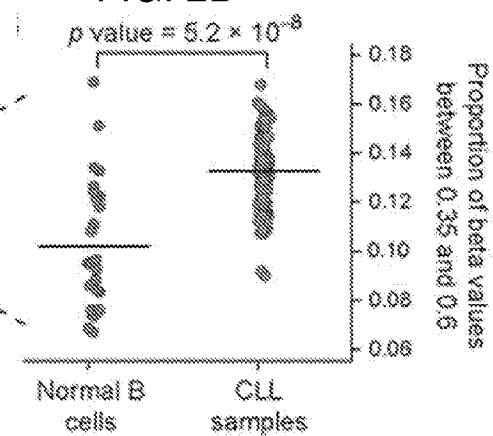


FIG. 2C

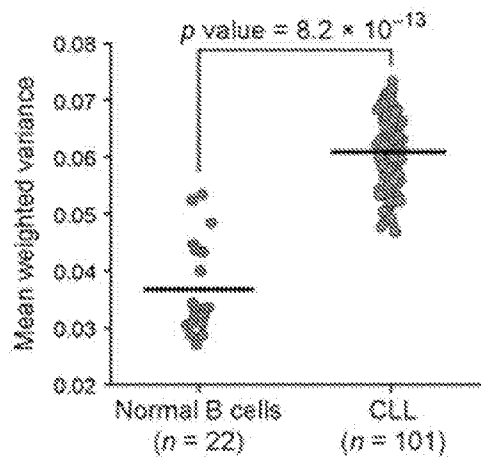


FIG. 2D

FIG. 3A

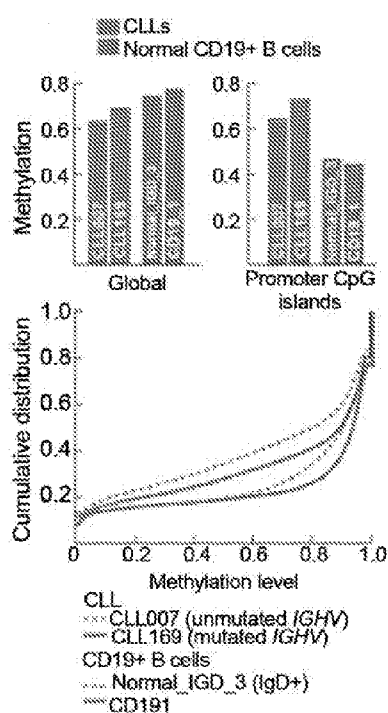


FIG. 3B

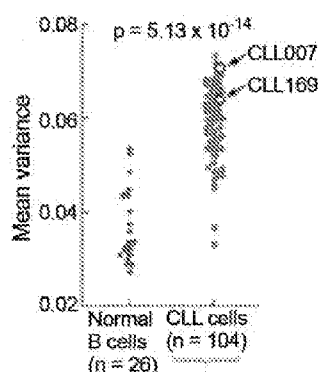


FIG. 3C

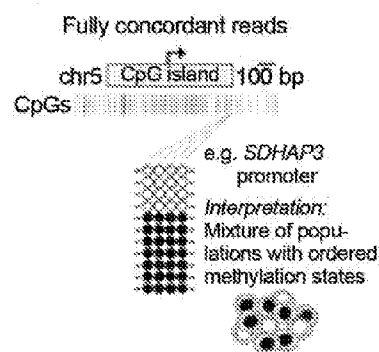
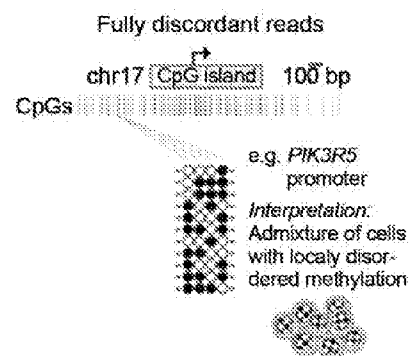
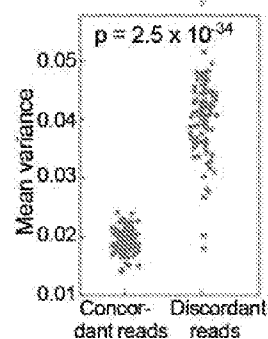


FIG. 3D



E

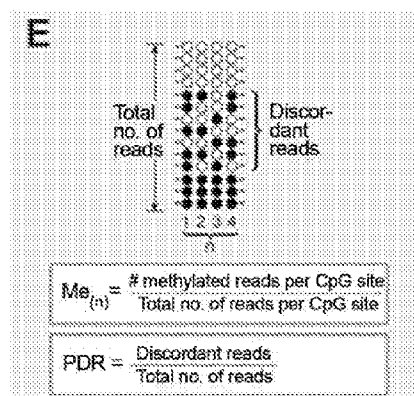


FIG. 3E

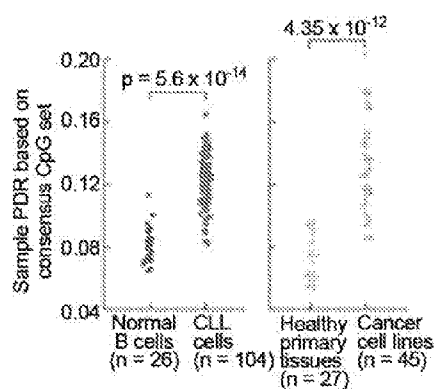


FIG. 3F

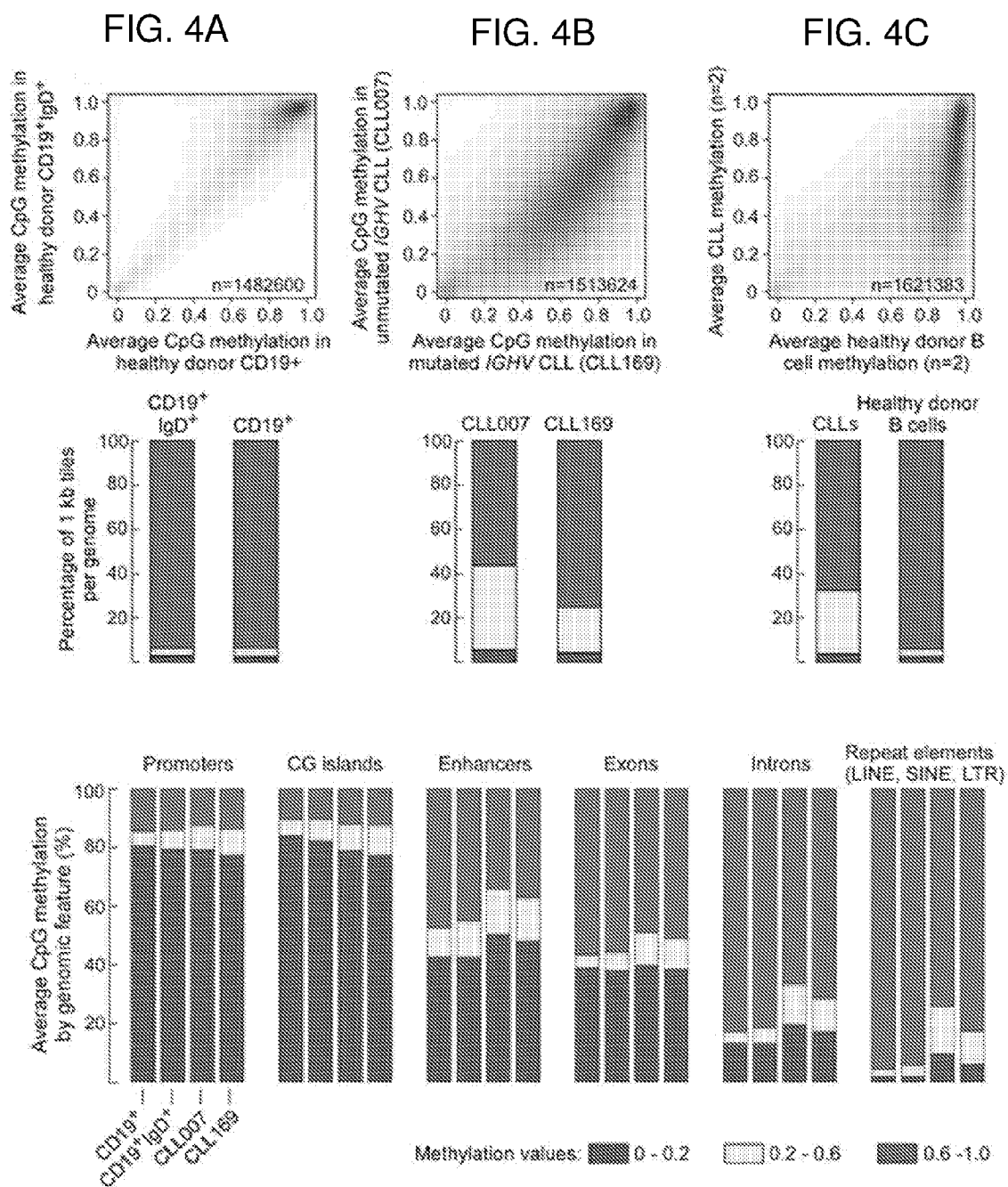


FIG. 4D

FIG. 4E

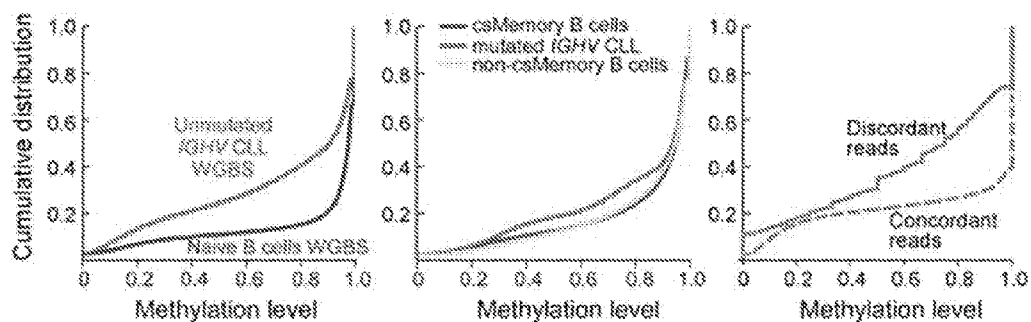


FIG. 4F

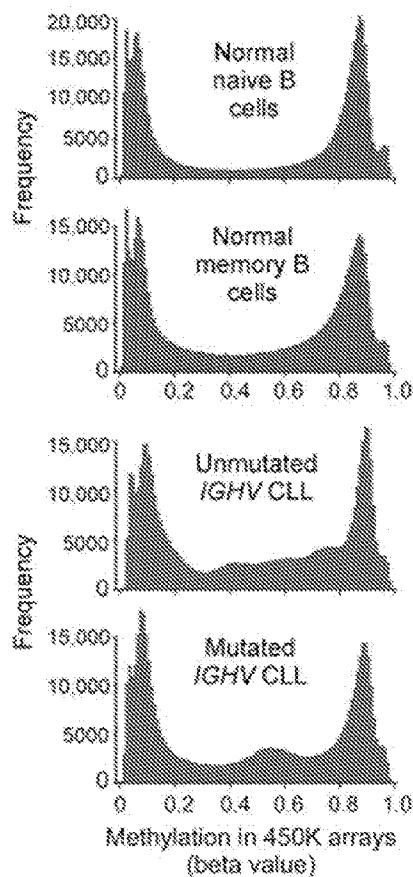


FIG. 4G

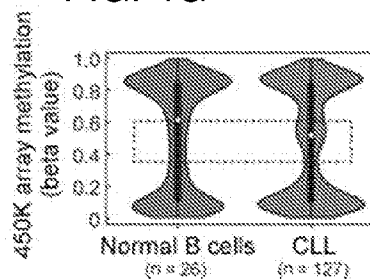


FIG. 4H

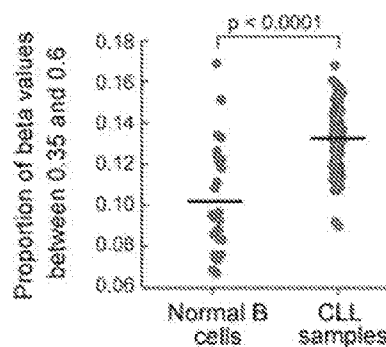


FIG. 4I

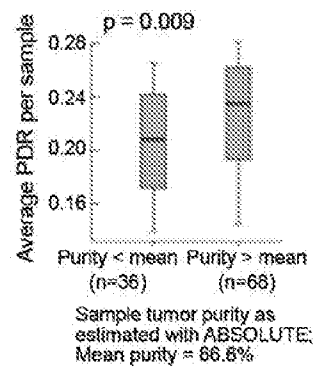


FIG. 4J

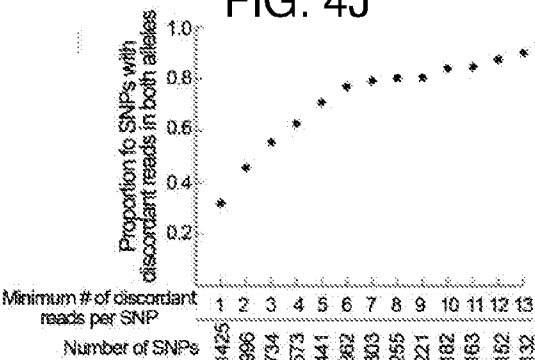


FIG. 4K

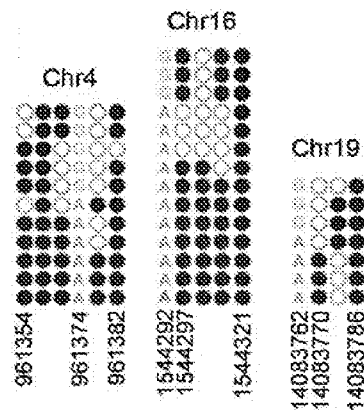


FIG. 4L

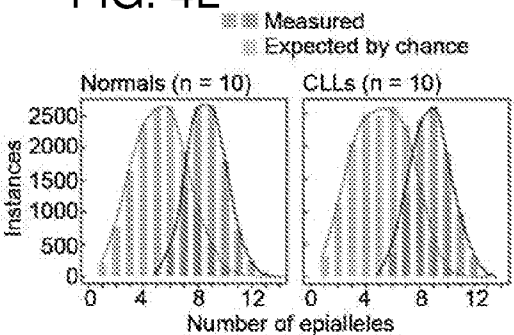


FIG. 4M

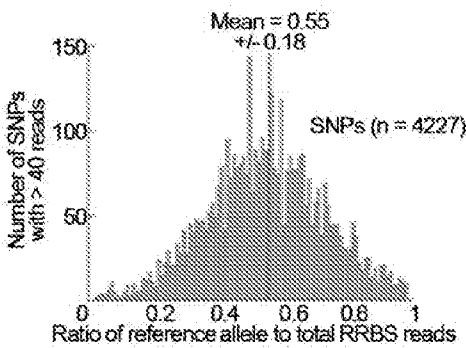


FIG. 4N

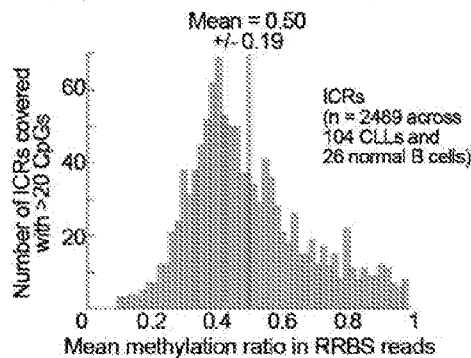


FIG. 4O

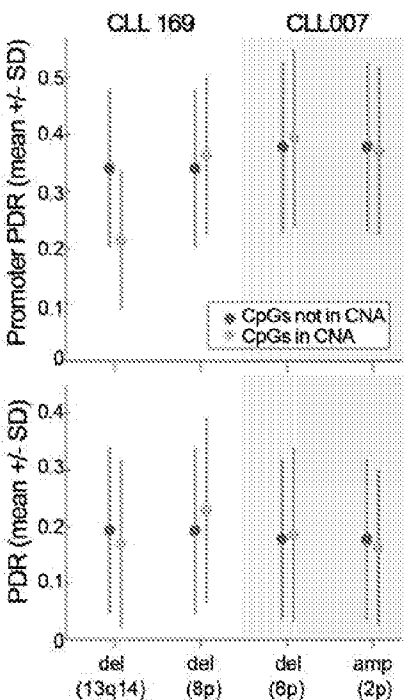


FIG. 5A

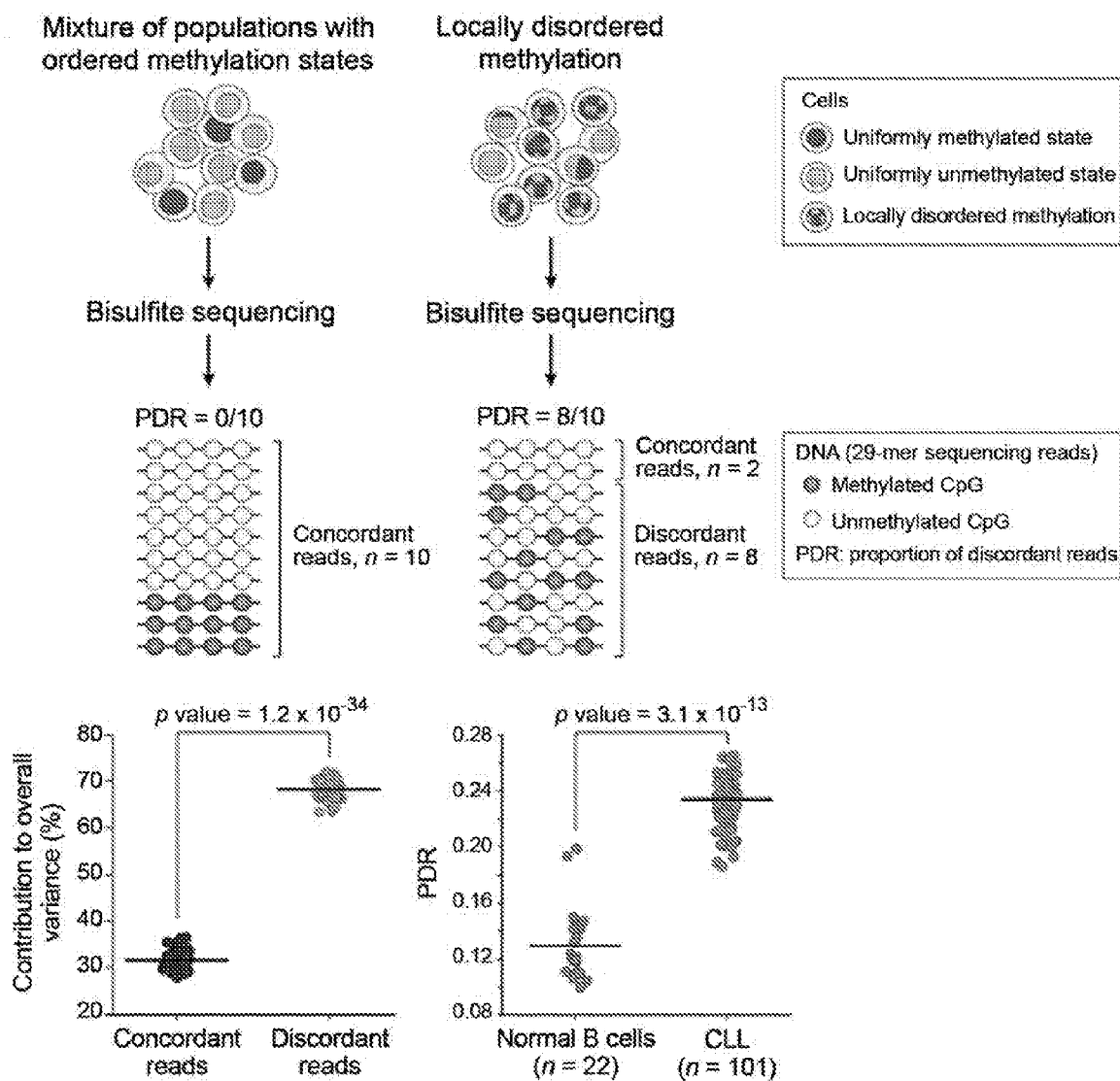


FIG. 5B

FIG. 5C

FIG. 6A

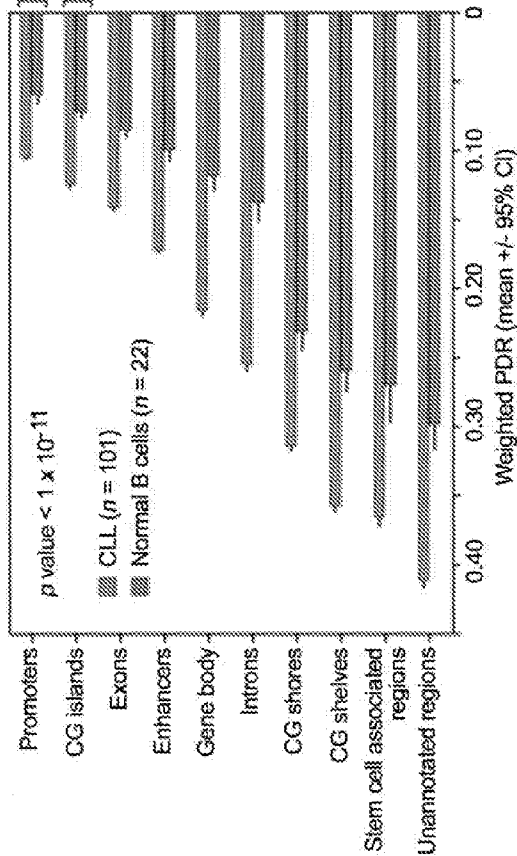


FIG. 6B

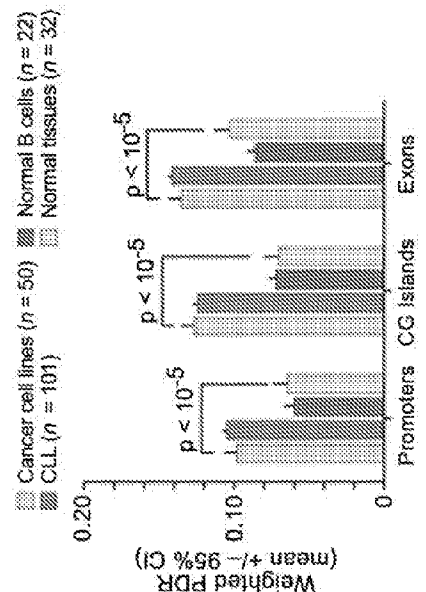
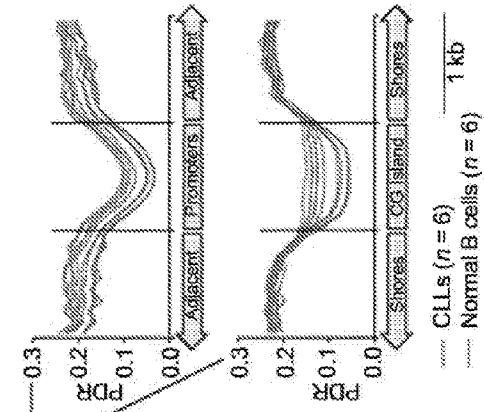


FIG. 6C

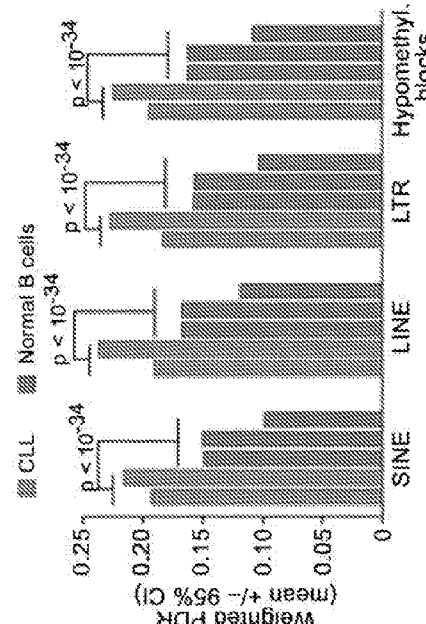


FIG. 6D

FIG. 7A

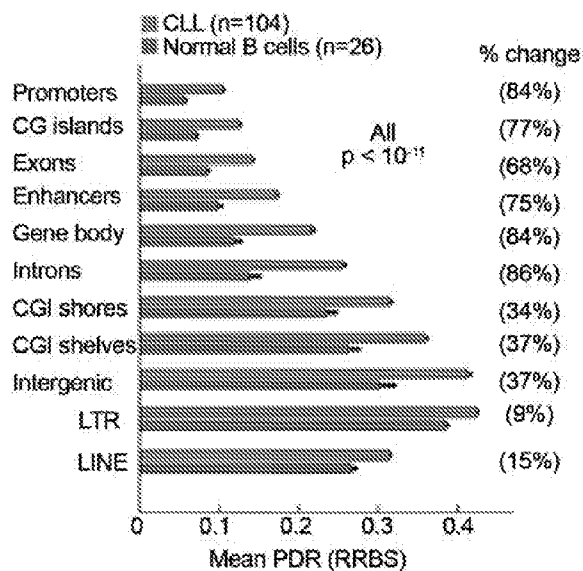


FIG. 7B

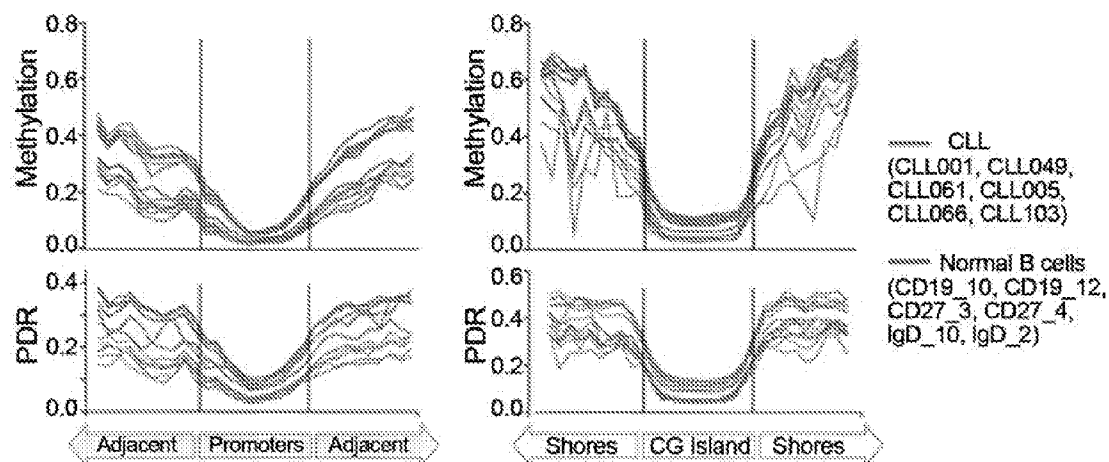
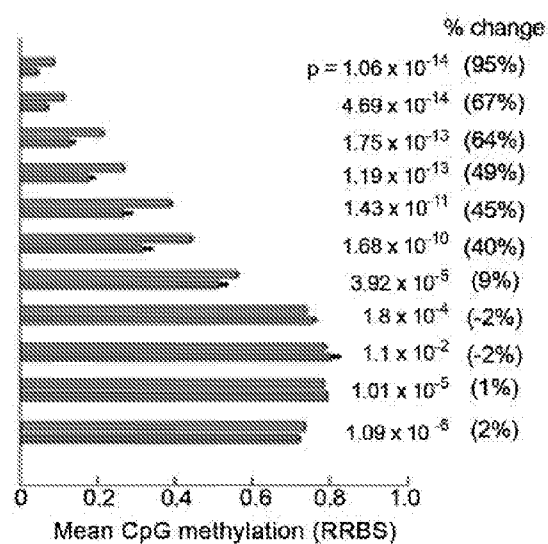


FIG. 7C

FIG. 8A

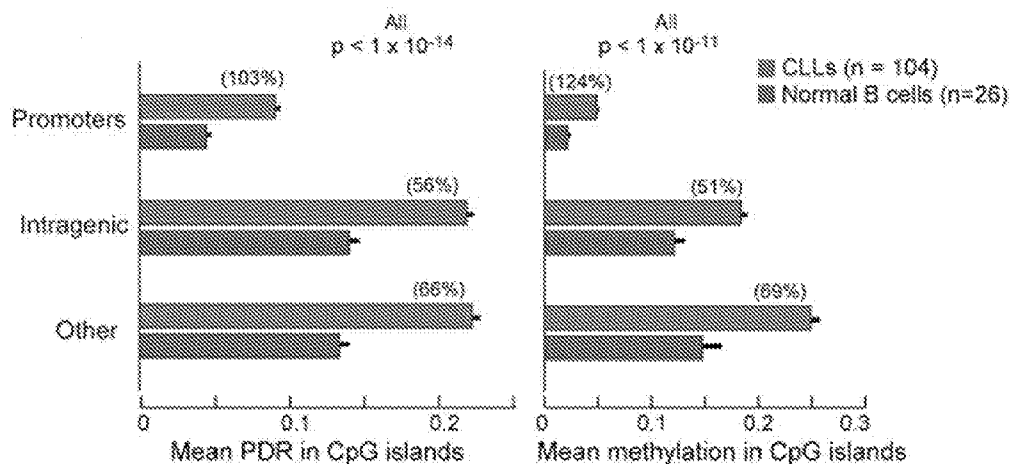


FIG. 8B

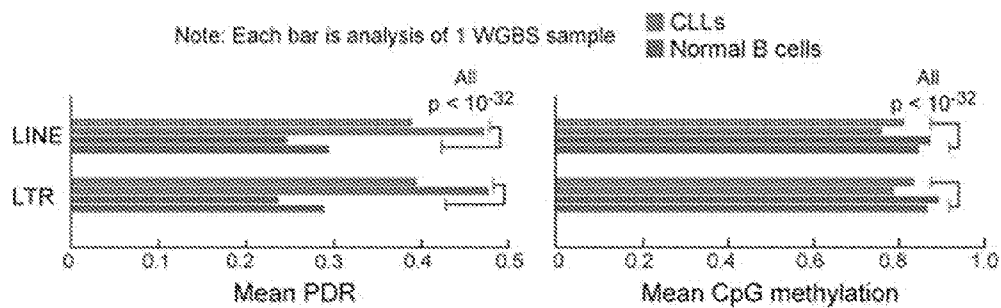


FIG. 8C

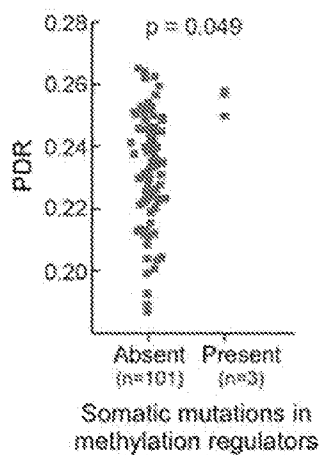


FIG. 9A

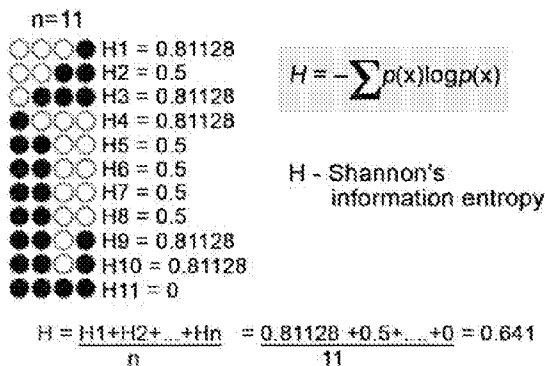


FIG. 9B

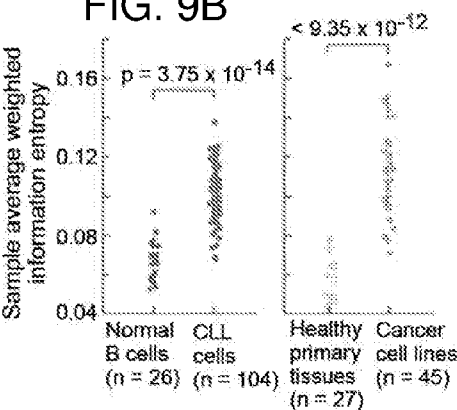


FIG. 9C

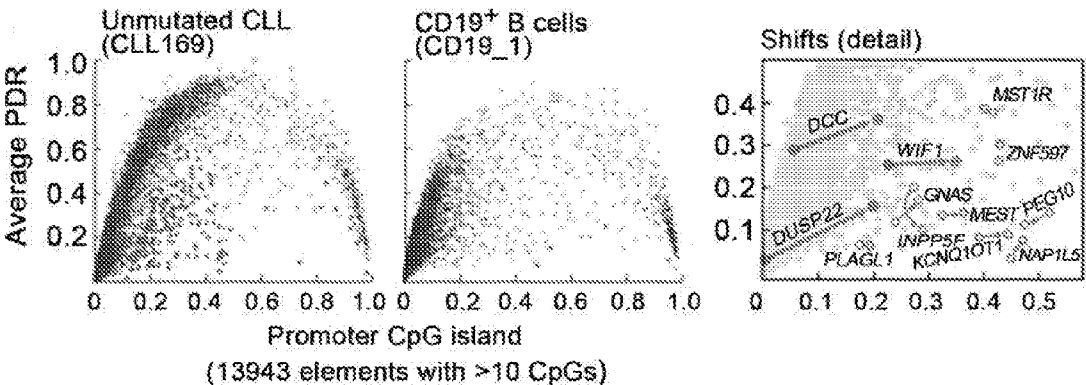
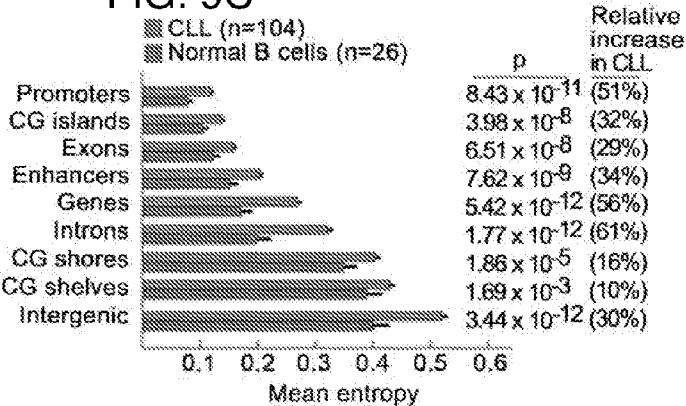


FIG. 9D

FIG. 9E

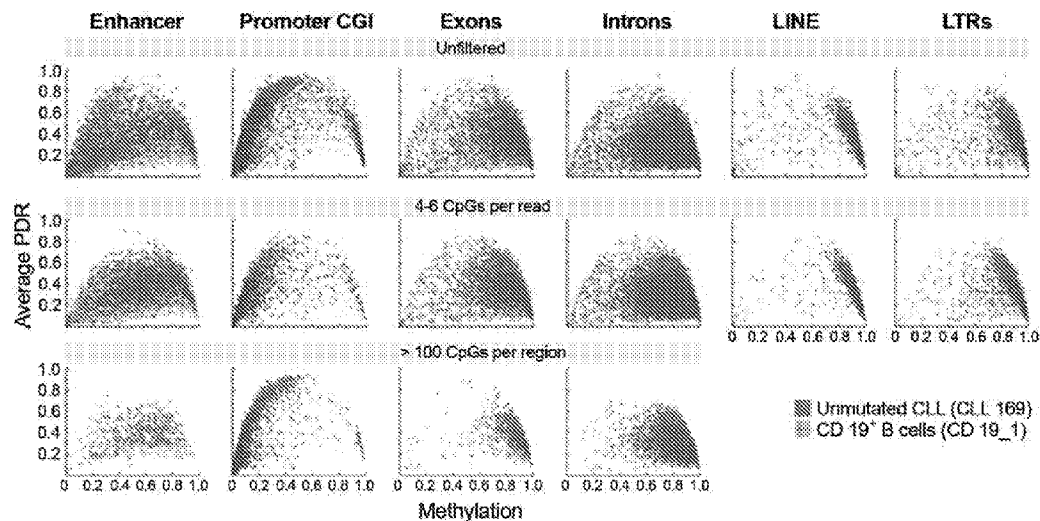


FIG. 9F

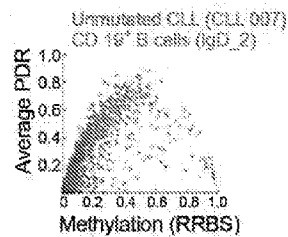


FIG. 9G

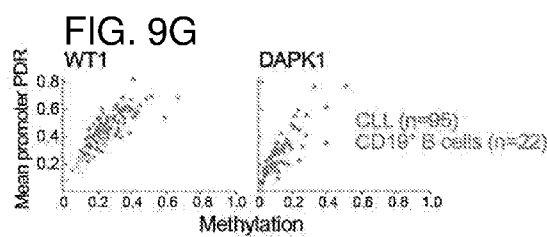


FIG. 9H

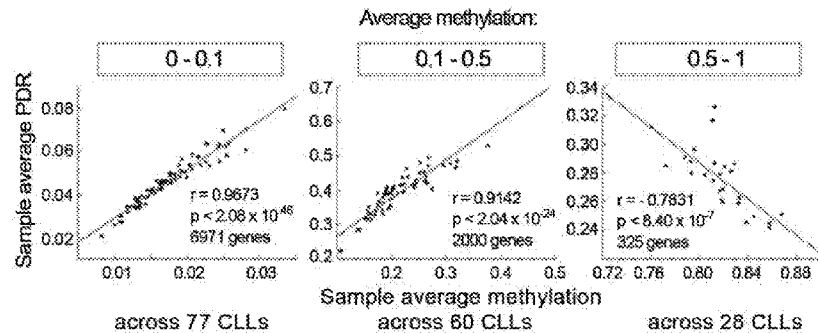


FIG. 10A

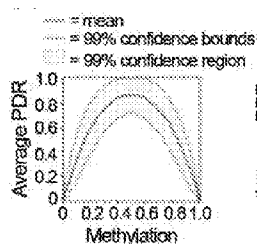


FIG. 10B

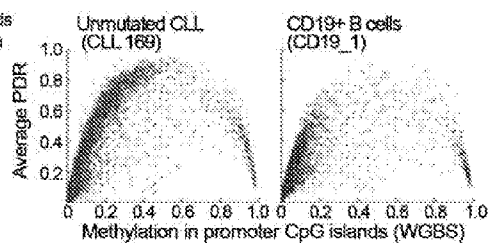


FIG. 10D

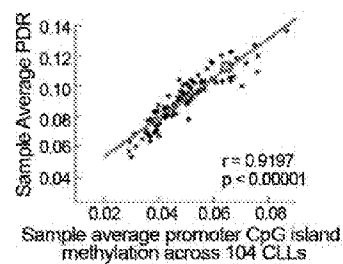


FIG. 10C

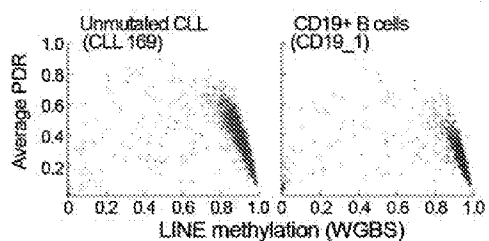


FIG. 10E

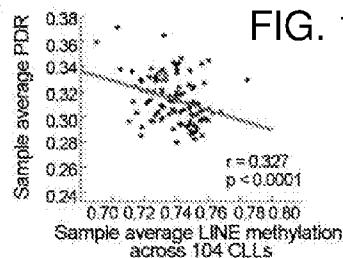


FIG. 10F

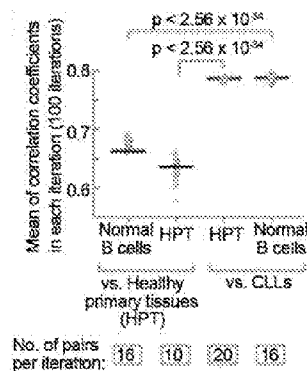
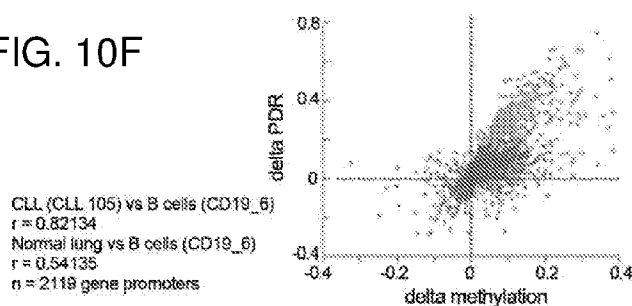


FIG. 10G

FIG. 11A

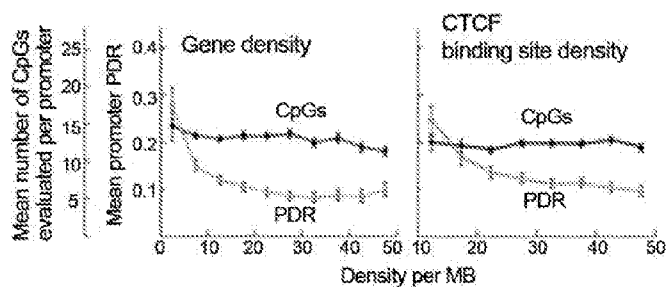


FIG. 11B

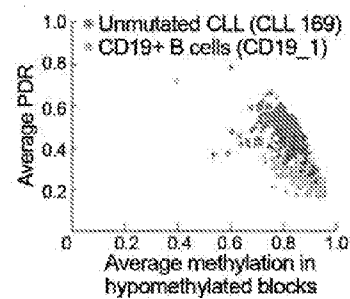


FIG. 11C

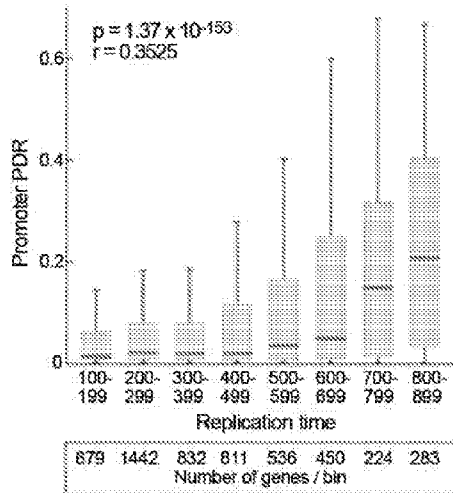


FIG. 11D

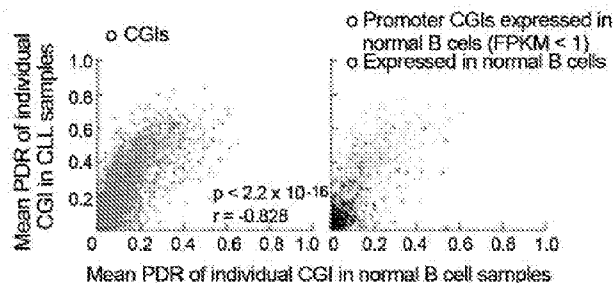
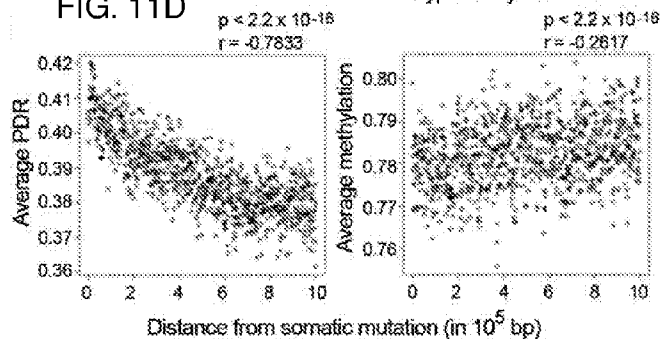
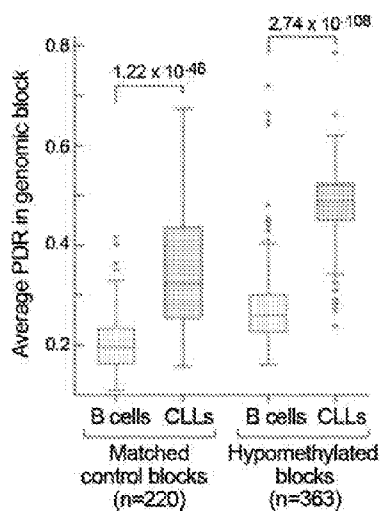


FIG. 11E

FIG. 12A



PDR and somatic mutations

Methylation and somatic mutations

FIG. 12B

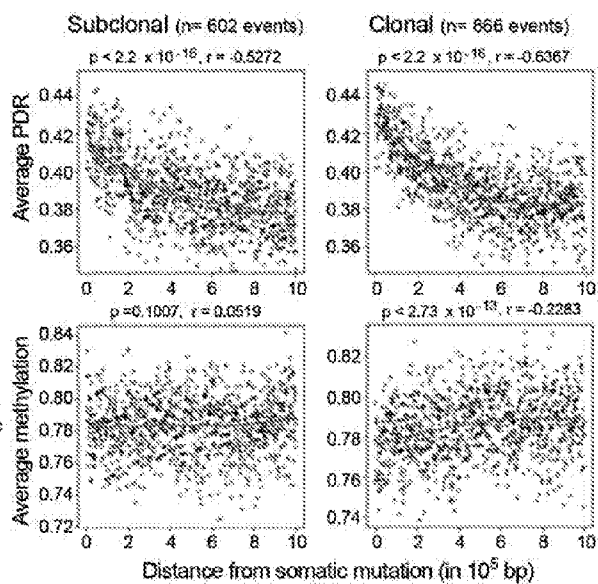


FIG. 12C

FIG. 13A

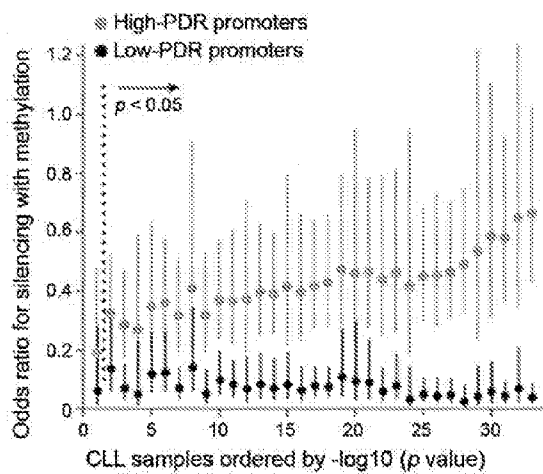


FIG. 13B

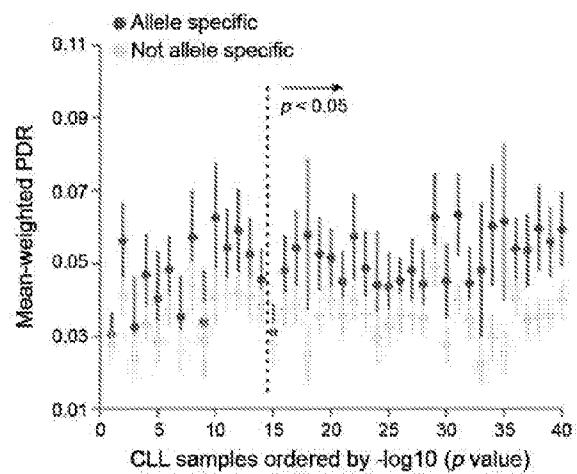


FIG. 14A

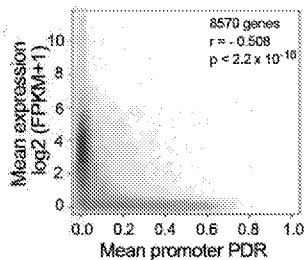


FIG. 14B

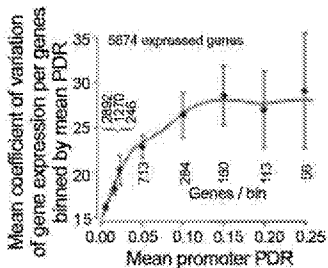


FIG. 14C

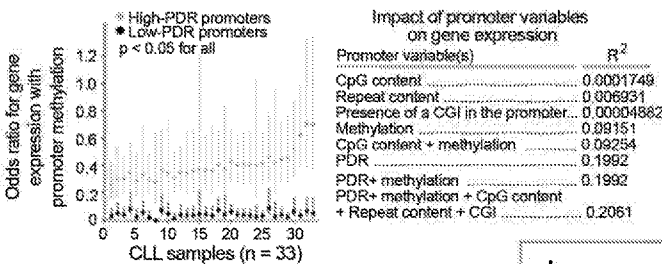
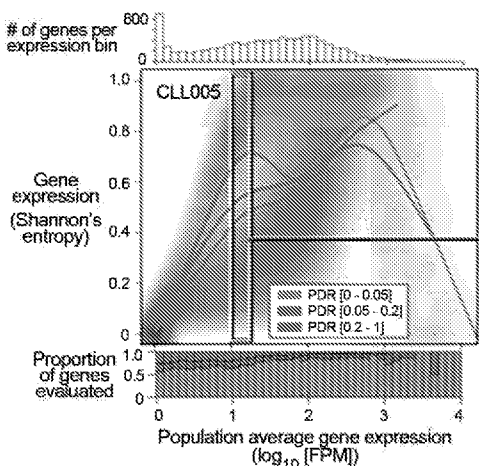


FIG. 14D



E

FIG. 14E

Sample	# single cells evaluated	p value for PDR in a model predicting gene expression entropy based on:	
		PDR	Population avg. gene expression entropy based on:
CLL005	84	< 2 x 10 ⁻¹⁸	Transcript length
CLL032	75	1.15 x 10 ⁻¹⁴	
CLL096	70	< 2 x 10 ⁻¹⁶	
CLL146	81	1.39 x 10 ⁻⁵	

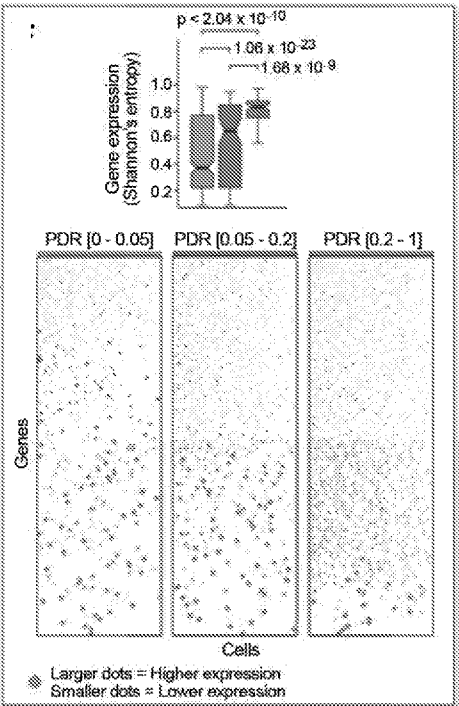


FIG. 14F

FIG. 15A

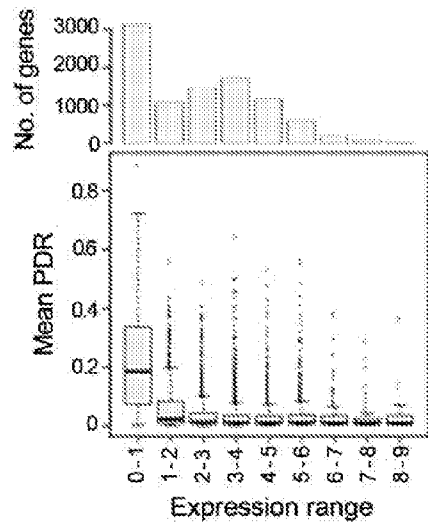


FIG. 15B

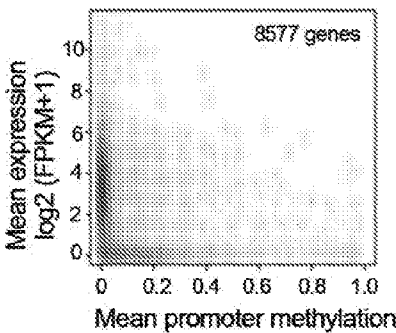


FIG. 15C

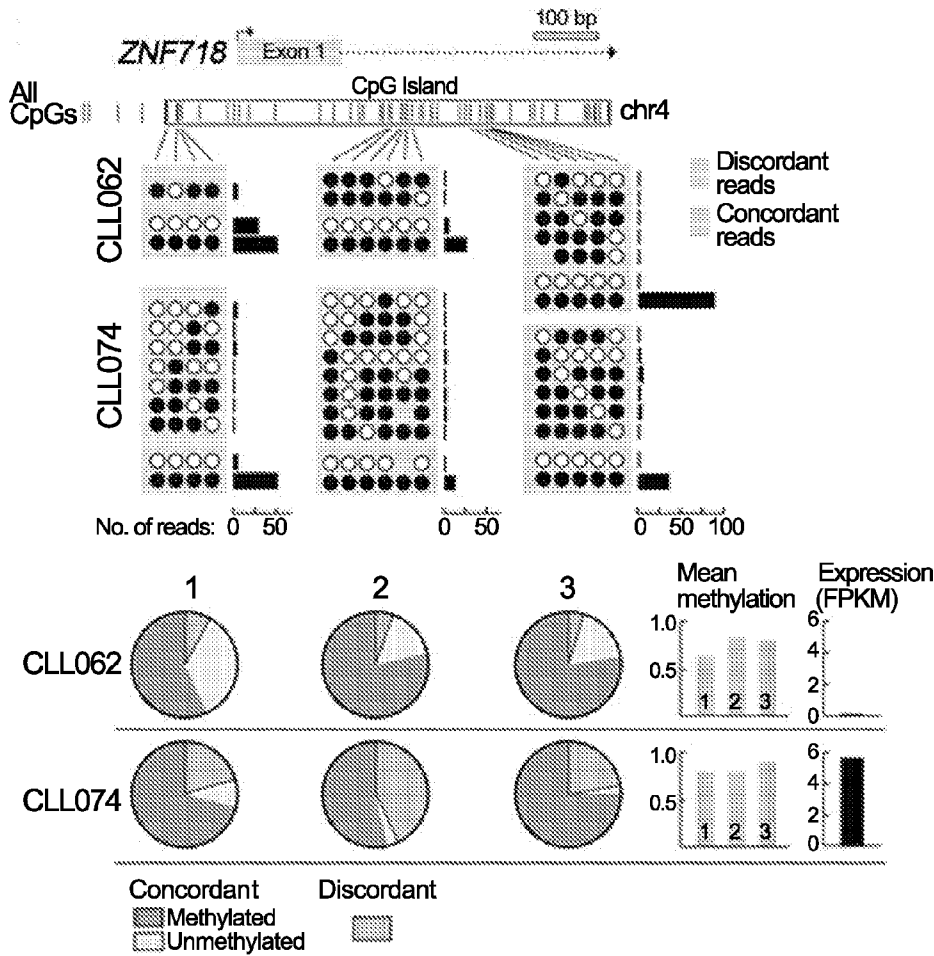


FIG. 15D

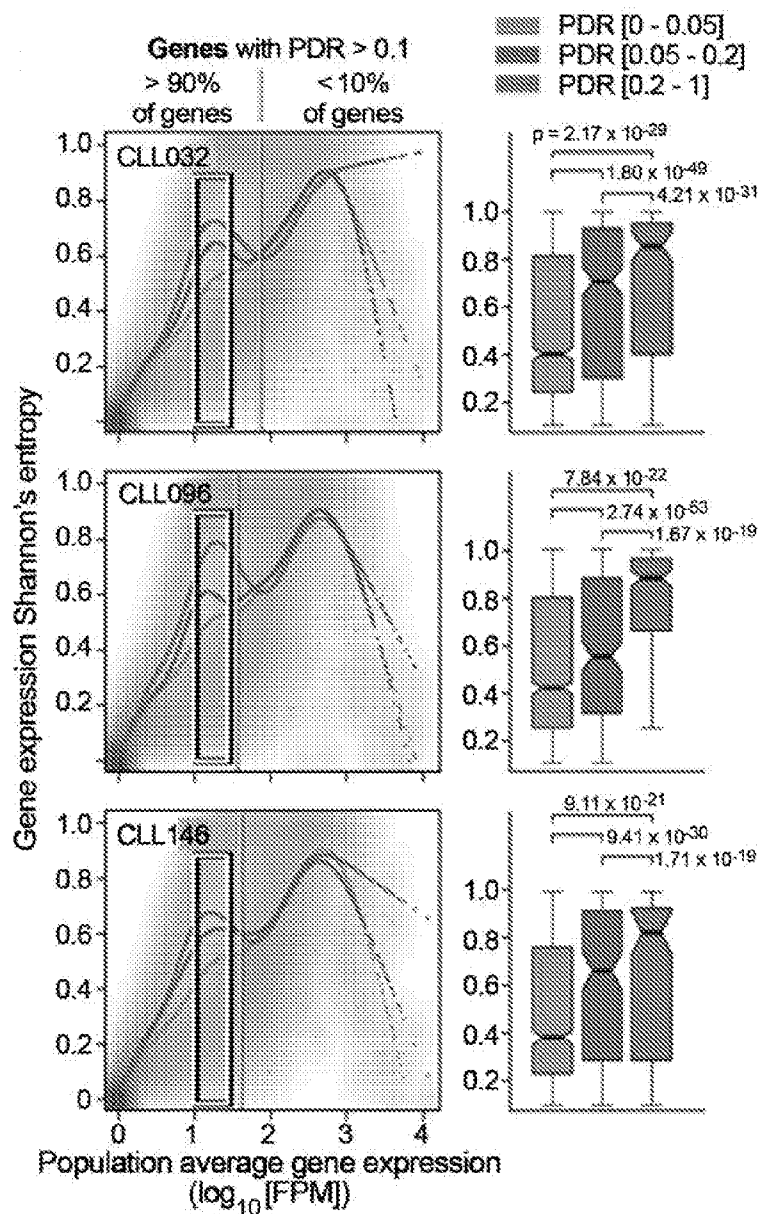


FIG. 15E

Sample	# single cells evaluated	p value for PDR in a model predicting gene expression entropy
CLL005	84	2.27×10^{-10}
CLL032	75	1.51×10^{-7}
CLL096	70	5.16×10^{-11}
CLL146	81	1.64×10^{-4}

A model for predicting single cell gene expression information entropy based on PDR, methylation, average gene expression and transcript length

FIG. 16A

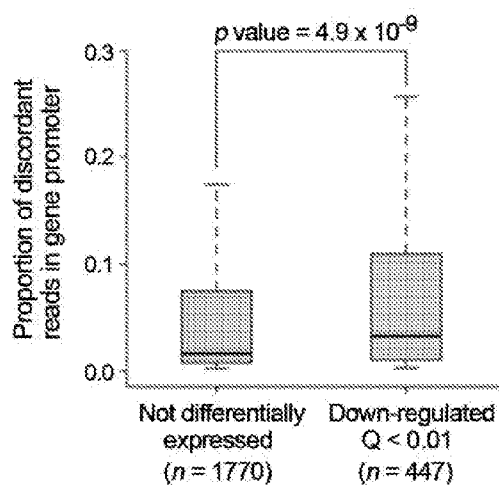


FIG. 16B

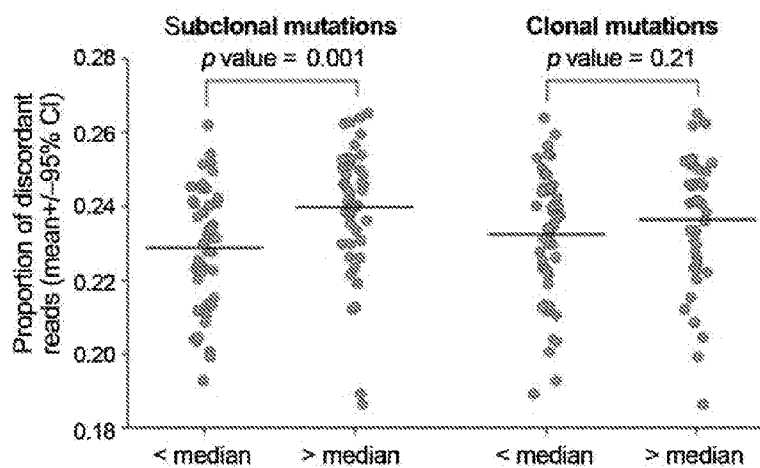
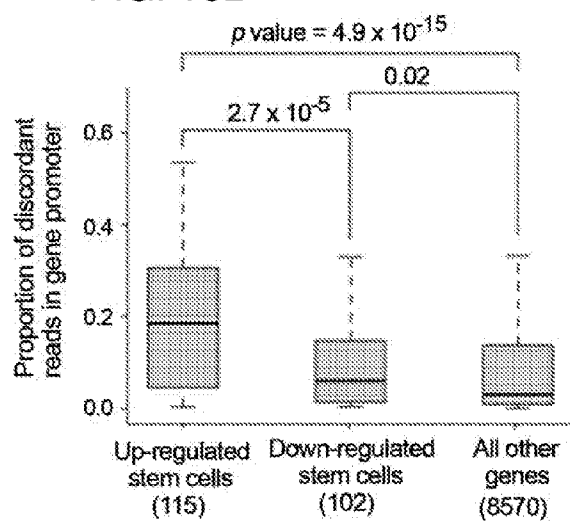


FIG. 16C

FIG. 17A

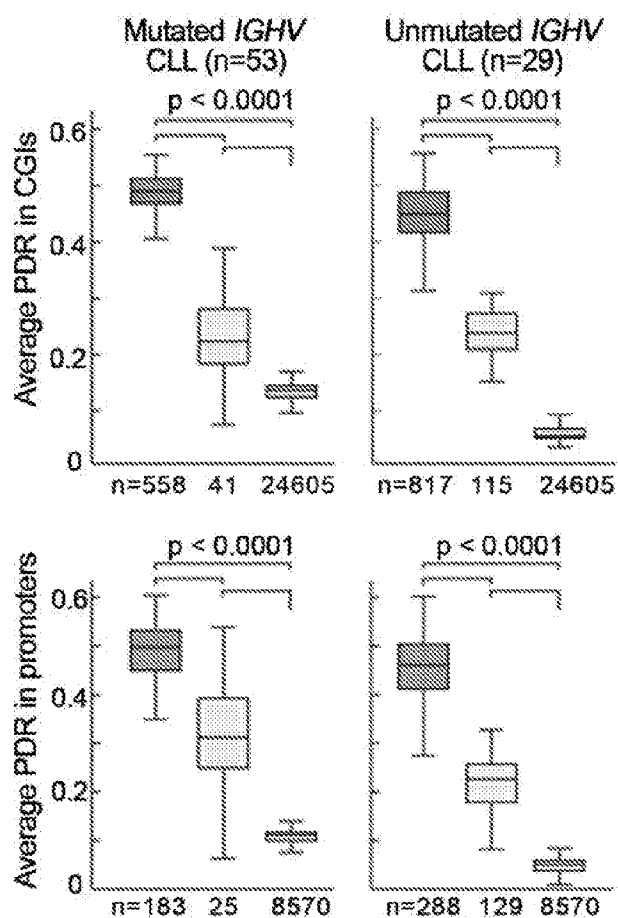


FIG. 17B

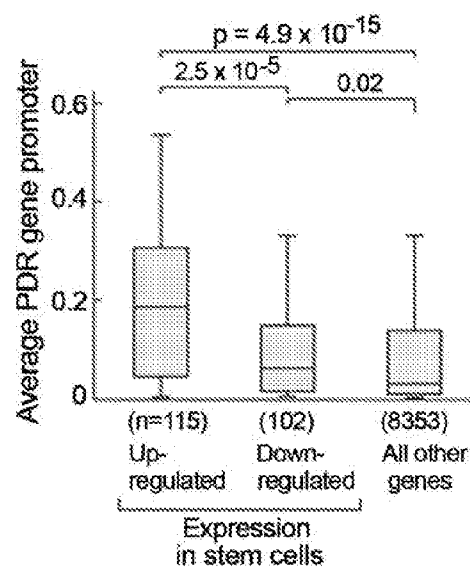


FIG. 18A

Gene sets enriched in genes with high promoter PDR	Q _{high}	Q _{high vs low}
TP63 targets (Perez)	4.83×10^{-25}	1.80×10^{-2}
TP53/TP63 targets (Perez)	4.06×10^{-18}	5.70×10^{-2}
Silenced by methylation in pancreatic cancer 1 (Sato)	1.98×10^{-28}	3.83×10^{-7}
Ewing sarcoma progenitor UP (Riggi)	3.68×10^{-23}	7.63×10^{-3}
Methylated in liver cancer DN (Acevedo)	4.91×10^{-22}	2.99×10^{-7}
Adult tissue stem module (Wong)	1.52×10^{-36}	1.12×10^{-2}
UVC response via TP53 group A (Bruins)	6.89×10^{-26}	4.22×10^{-2}
Mammary stem cell UP (Lim)	5.42×10^{-38}	2.93×10^{-7}
Methylated in cancer (Hoque)	3.78×10^{-10}	4.22×10^{-2}
Neuronal crest stem cell UP (Lim)	2.02×10^{-9}	4.18×10^{-2}

FIG. 18B

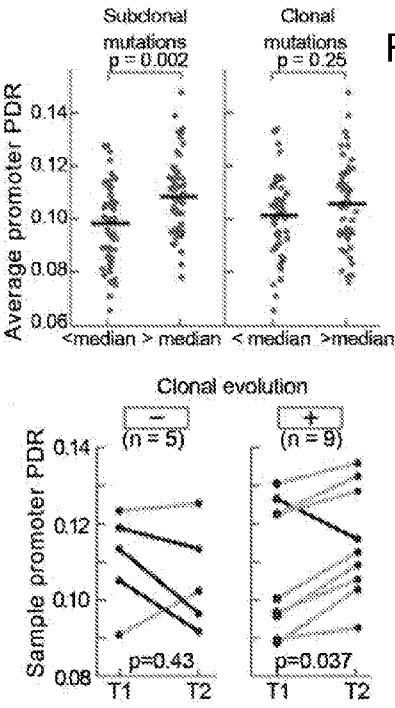
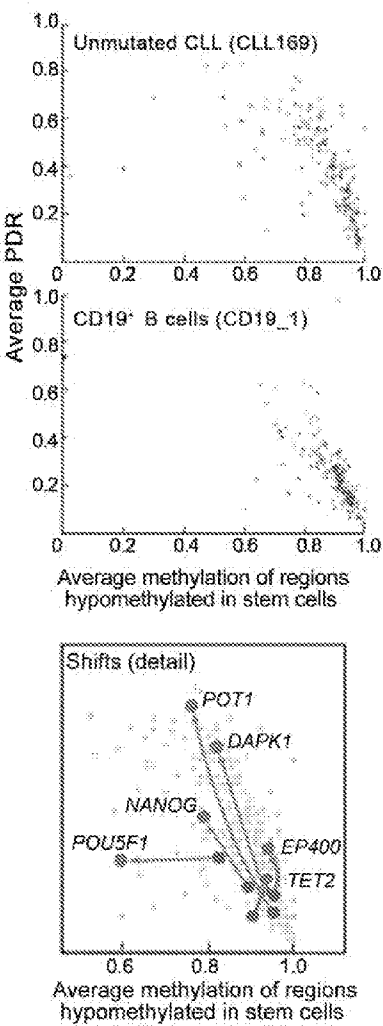


FIG. 18C

FIG. 18D

Gene sets enriched in Methylation change status T2 vs. T1	Q _{High}	Q _{High vs. Low}
↓ Hematopoietic stem cell UP (Jaatinen)	7.00×10^{-28}	2.83×10^{-13}
↓ Adult tissue stem module (Wong)	6.82×10^{-9}	2.89×10^{-13}
↓ Mammary stem cell UP (Lim)	5.58×10^{-8}	2.10×10^{-12}
↑ Methylated in lymphoma (Cawson)	9.59×10^{-7}	1.68×10^{-7}

FIG. 18E

FIG. 19A

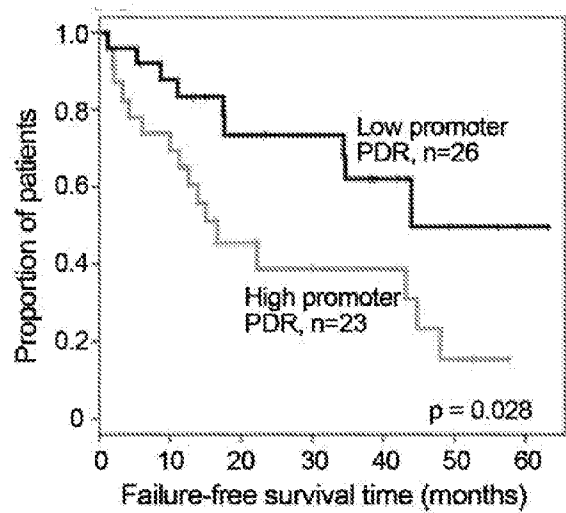


FIG. 19B

Variable	HR	95%CI	p
High promoter PDR	2.81	1.05 - 7.53	0.039
<i>IGHV</i> unmutated status	4.55	1.45 - 14.20	0.009
<i>del</i> (17p)	2.50	0.83 - 7.50	0.099
<i>del</i> (11q)	1.19	0.41 - 3.45	0.746
↓			
High promoter PDR	1.44	0.49 - 4.19	0.503
<i>IGHV</i> unmutated status	3.69	1.15 - 11.74	0.027
<i>del</i> (17p)	3.20	0.99 - 10.53	0.050
<i>del</i> (11q)	1.09	0.33 - 3.04	0.98
subclonal driver present	4.086	1.31 - 12.73	0.015

FIG. 20A

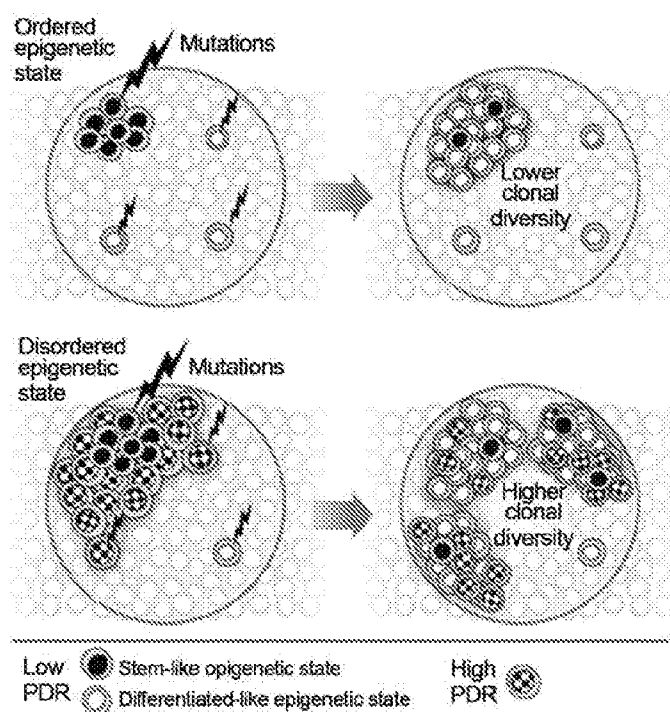
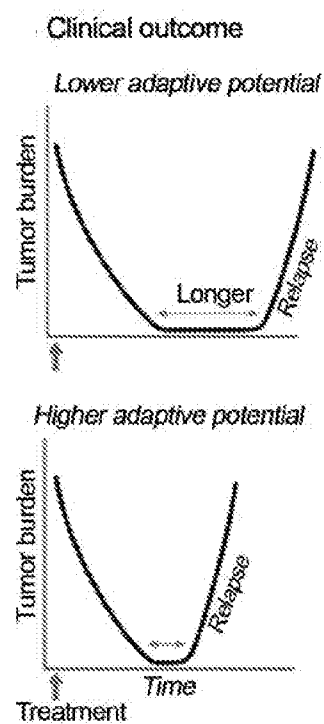


FIG. 20B



COMPOSITIONS AND METHODS FOR DIAGNOSING, EVALUATING AND TREATING CANCER

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] The present application is filed pursuant to 35. U.S.C. §371 as a U.S. National Phase Application of International Application Number PCT/US2014/067146, which was filed on Nov. 24, 2014, and published as WO2015/077717 on May 28, 2015, and claims benefit of and priority to U.S. provisional patent application Ser. No. 61/908,316, filed Nov. 25, 2013.

[0002] The foregoing applications, and all documents cited therein or during their prosecution (“appln cited documents”) and all documents cited or referenced in the appln cited documents, and all documents cited or referenced herein (“herein cited documents”), and all documents cited or referenced in herein cited documents, together with any manufacturer’s instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. More specifically, all referenced documents are incorporated by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

FIELD OF THE INVENTION

[0003] The present invention relates to methods for diagnosing, evaluating and treating cancer.

BACKGROUND OF THE INVENTION

[0004] In recent years, multiple effective treatment modalities for cancer have been introduced. Nevertheless, in many cases, despite striking initial responses, the malignant process evolves and adapts to the therapy, leading to disease recurrence. Therefore, an important barrier to curative cancer therapy is the plasticity of cancer, its ability to adapt to treatment. Chronic lymphocytic leukemia (CLL) constitutes an informative case study to dissect this phenomenon. Despite highly effective therapies, this common leukemia remains incurable. For example, chemoimmunotherapy with fludarabine, cyclophosphamide and rituximab (FCR) leads to a 44% complete response rate, yet the disease invariably recurs, often after evolving to a more aggressive and treatment-refractory form.

[0005] Cancer subpopulations compete and mold the malignant genetic landscape to yield adaptation to therapy. In CLL, the presence of co-existing cell subpopulations, distinguished on the basis of genetic differences, was first demonstrated using cytogenetic technologies and SNP arrays. Recently, massively parallel sequencing (MPS) has allowed genetic heterogeneity to be studied in CLL at an unprecedented resolution. Such studies demonstrated that evolution in response to therapy is the rule rather than the exception. Using this approach, a clear impact of pre-treatment heterogeneity on the rapidity of clonal evolution and the overall clinical outcome has been shown.

[0006] Genome-wide methylation assays, such as arrays and MPS with bisulfite conversion, have revealed that aberrant DNA methylation, in addition to dysregulated

genes and pathways, is involved in CLL pathogenesis. In CLL and other cancers it has been previously reported that there is a global decrease in DNA methylation and an increase in methylation specifically at CpG islands (CGI) (Baylin and Jones, 2011; Kulis et al., 2012). Specifically, it has been thought that in a normal cell the CpG islands are completely unmethylated at the CpG sites within a CpG island and when the cell becomes a tumor cell the CpG island becomes completely methylated at every CpG. Moreover, CpG islands, which are normally located near the promoters of genes and contain a higher than expected CG content, are normally kept hypomethylated. Presumably, this is to create an active euchromatin environment as well as preventing C to T mutations caused by deamination of methylated cytosine. In an unmethylated state cytosine is converted to uracil after deamination, which is recognized by the cell’s repair machinery and is removed, while in a methylated state deamination of cytosine results in the formation of thymine which is not recognized by the repair machinery. Therefore, the presence or absence of hypermethylation at these CpG islands can be used to detect tumor cells. As cancer cells are constantly evolving to avoid treatment regimens, there is a need for a method to not only detect a tumor cell, but to detect tumor cell plasticity. Determining plasticity of a tumor can allow a personalized treatment for a patient in need thereof.

[0007] Methylation profiles were also shown to have independent prognostic value in CLL. Like genetic alterations, DNA methylation modifications are heritable and therefore subject to natural selection in cancer. Furthermore, genetically uniform cell subpopulations can contain profound epigenetic differences leading to phenotypic differences in their survival capacity and proliferative potential. Together, these observations suggest that an integrative model of cancer evolution is warranted, which accounts for both epigenetic heterogeneity of genetically uniform subpopulations, and genetic heterogeneity of epigenetically uniform subpopulations.

[0008] Citation or identification of any document in this application is not an admission that such document is available as prior art to the present invention.

SUMMARY OF THE INVENTION

[0009] A first aspect of the present invention provides a method of assessing a subject’s tumor plasticity or the ability to acquire treatment resistance mutations. In another embodiment the method assesses a subject’s cancer treatment prognosis. In some embodiments, the method comprises detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells in a tumor sample from the subject; comparing the DNA methylation status of neighboring CpG sites along a sequence of CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells; and assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells. In one embodiment there are at least 2 neighboring CpG sites. In another embodiment there are at least 3, or 4, or 5, or 6, or 7, or 8, or 9, or 10 neighboring CpG sites, preferably greater than 4. Consistency can mean that all of the neighboring CpG sites along a sequence of CpG sites in DNA or CpG sites across multiple gene copies are methylated or all of the neighboring

CpG sites along a sequence of CpG sites in DNA are unmethylated. It may mean that greater than 50%, or 60% or 70%, or 75%, or 80%, or 85%, or 90%, or 95% of the CpG's are methylated or unmethylated. Inconsistency can mean that any sequence of neighboring CpG sites or CpG sites across multiple gene copies contain at least one methylated and at least one unmethylated CpG site. The presence or prevalence of inconsistent methylation status along the sequences or across the multiple gene copies may indicate that the subject is more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof. For example, the DNA methylation status of CpG sites along one or more sequencing reads, e.g., operatively linked to each other in a single polynucleotide molecule, may be detected. The neighboring CpG sites along the sequencing read may then be compared to each other or to corresponding positions of different sequencing reads (e.g., at the same genomic location) from the plurality of cells.

[0010] The DNA methylation may be detected by methylation-specific PCR, whole genome bisulfite sequence, the HELP assay, ChIP-on-chip assays, restriction landmark genomic scanning, methylated DNA immunoprecipitation, pyrosequencing of bisulfite treated DNA, molecular break light assay for DNA adenine methyltransferase activity, methyl sensitive Southern blotting, methylCpG binding proteins, and reduced representation bisulfite sequencing. In some embodiments, the DNA methylation is detected in a methylation assay utilizing next-generation sequencing. For example, DNA methylation may be detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing. Optionally, the DNA methylation is detected by microarray, such as a genome-wide microarray.

[0011] The methylation status of neighboring CpG sites may be compared by identifying and/or quantifying inconsistently methylated regions, such as by calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism, or calculating information entropy. In some embodiments, a proportion of discordant reads (PDR) is calculated. Optionally, each region of neighboring CpG sites (e.g., within a sequencing read) is assigned a consistent status or an inconsistent status before calculating the proportion of discordant reads, variance, epipolymorphism or information entropy. There may be multiple inconsistent statuses, each representing a distinct methylation pattern or class of similar methylation patterns.

[0012] The one or more regions of neighboring CpG sites having a locally disordered methylation status may be in a genomic location selected from a CpG island, a CpG shore, a CpG shelf, a promoter, an enhancer, an exon, an intron, a gene body, a stem cell associated region, a short interspersed element (SINE), a long interspersed element (LINE), and a long terminal repeat (LTR), preferably a CpG island or a promoter.

[0013] In some embodiments, the method further comprises detecting a subclonal genetic mutation. Optionally, the subclonal genetic mutation is within the one or more genomic regions having a locally disordered methylation status. A subclonal genetic mutation with the one or more genomic regions having a locally disordered methylation status may indicate that the subject is even more likely to (1) develop resistance to an antitumor agent; (2) relapse after

treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof.

[0014] The DNA methylation status at the one or more neighboring CpG sites may be detected before treatment with an antitumor agent. Optionally, the DNA methylation status at the one or more neighboring CpG sites is detected after treatment with an antitumor agent. The DNA methylation status at the one or more neighboring CpG sites may be detected both before and after treatment with an antitumor agent. In some embodiments, the DNA methylation status at the one or more neighboring CpG sites is detected throughout a time course of treatment with an antitumor agent. An increase in the number of regions of neighboring CpG sites having a locally disordered methylation status, or the level of inconsistent methylation status in these regions across sequences in a sample, may indicate that the subject is even more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof.

[0015] A second aspect of the present invention provides a method of calculating a proportion of discordant reads (PDR) in a first tumor sample from a subject. In some embodiments, the method comprises detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells in the tumor sample; comparing the DNA methylation status of sequencing reads in multiple regions of neighboring CpG sites along a sequence of CpG sites in DNA of the plurality of cells; and determining a relative number of cells in the tumor sample having inconsistent methylation status across the sequence of CpG sites as compared to the total number of cells in the tumor sample or a number of cells in the tumor sample having consistent methylation status across the sequence of CpG sites, or determining a level of inconsistent methylation status across the sequence of CpG sites in cells in the tumor sample.

[0016] The DNA methylation may be detected by methylation-specific PCR, whole genome bisulfite sequence, the HELP assay and other methods using methylation-sensitive restriction endonucleases, ChIP-on-chip assays, restriction landmark genomic scanning, COBRA, Ms-SNuPE, methylated DNA immunoprecipitation (MeDip), pyrosequencing of bisulfite treated DNA, molecular break light assay for DNA adenine methyltransferase activity, methyl sensitive Southern blotting, methylCpG binding proteins, mass spectrometry, HPLC, and reduced representation bisulfite sequencing. In some embodiments, the DNA methylation is detected in a methylation assay utilizing next-generation sequencing. For example, DNA methylation may be detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing. Optionally, the DNA methylation is detected by microarray, such as a genome-wide microarray.

[0017] The PDR may be calculated before treatment with an antitumor agent. Optionally, the PDR is calculated after treatment with an antitumor agent. The PDR may be calculated both before and after treatment with an antitumor agent. In some embodiments, the PDR is calculated throughout a time course of treatment with an antitumor agent.

[0018] In some embodiments, a PDR threshold, such as greater than 0.15, indicates that the patient is more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic

tumor; or (4) any combination thereof. A PDR threshold, such as less than 0.15, may indicate that the patient is more likely to respond to treatment with an antitumor agent. A change, such as an increase or in some instances a decrease, in PDR following treatment may indicate that the subject is likely to develop resistance to the treatment.

[0019] A third aspect of the present invention provides a method of treating a subject suffering from cancer. In some embodiments, the method comprises performing the method of identifying a subject's cancer treatment prognosis described herein and administering an antitumor agent to the subject if no or few inconsistencies in methylation status are identified or b) administering fewer antitumor agents to a subject having a low level of inconsistencies in methylation status and more antitumor agents to a subject having a high level of inconsistencies in methylation status. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if a high level of inconsistencies in methylation status is identified.

[0020] In another embodiment the treatment is based on the standard of care for a particular cancer. In one embodiment if the standard of care allows a physician to choose between two treatment options, such as surgery or chemotherapy, the addition of detecting plasticity based on DNA methylation discordance can determine the proper option. In another embodiment if the standard of care can be followed using different doses of an antitumor agent DNA methylation discordance may be used to select the proper dose.

[0021] In some embodiments, the method comprises, performing a prognostic method as described herein; administering an antitumor agent to the subject; and repeating the prognostic method, wherein the treatment is administered between the initial and subsequent prognostic methods. In some embodiments, the method comprises continuing to treat the subject with the antitumor agent if the level of inconsistent methylation status is substantially the same in the initial and subsequent prognostic methods or lower in the subsequent prognostic method than in the initial prognostic method. In some embodiments, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the level of inconsistent methylation status is different in the subsequent prognostic method compared to the initial prognostic method. The antitumor agent may be administered to the subject for at least 3 months, at least 6 months, at least 9 months, at least 12 months, at least 24 months, or at least 36 months before performance of the second prognostic method, preferably at least 12 months.

[0022] The methylation status of neighboring CpG sites may be compared by calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism, or calculating information entropy.

[0023] The method of treatment may further comprise detecting a genetic mutation. The genetic mutation may be a clonal mutation or a subclonal mutation, preferably a subclonal mutation. In some embodiments, the method comprises treating or continuing to treat the subject with the antitumor agent if no region of neighboring CpG sites having locally disordered methylation status also comprises a genetic mutation, such as a subclonal mutation. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if one or more regions of

neighboring CpG sites having locally disordered methylation status also comprises a genetic mutation, such as a subclonal mutation.

[0024] In some embodiments, the method comprises determining the genomic location of the one or more regions of neighboring CpG sites having locally disordered methylation status. The genomic location may be selected from a CpG island, a CpG shore, a CpG shelf, a promoter, an enhancer, an exon, an intron, a gene body, a stem cell associated region, a short interspersed element (SINE), a long interspersed element (LINE), and a long terminal repeat (LTR), preferably a CpG island or a promoter. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if one or more regions of neighboring CpG's having locally disordered methylation status is located in a CpG island, a promoter or an exon. The method may comprise treating or continuing to treat the subject with the antitumor agent if no or few regions of neighboring CpG sites having locally disordered methylation status are within a CpG island, a promoter, or an exon.

[0025] In some embodiments, the method of treatment comprises calculating a PDR from a tumor sample in a subject as described herein, and administering an antitumor agent to the subject if the PDR is less than a PDR threshold, such as 0.15. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the PDR is greater than a PDR threshold, such as 0.15.

[0026] In some embodiments, the method of treatment comprises calculating a PDR from a sample of cells shed from a tumor in a subject as described herein, and administering an antitumor agent to the subject if the PDR is less than a PDR threshold, such as 0.15. The shed cells can be collected from the colon, from the bladder, from the kidney's, from the prostate, or from the lungs. The shed cells could be in the urine, sputum, semen, or stool. In one embodiment, the PDR may determine that a lung, or bladder, or kidney, or colon, or prostate should be removed. In one aspect, removal is based on the plasticity of the tumor cells as determined by the PDR. In one aspect the plasticity indicates that the tumor may become invasive. In another aspect, determining the PDR allows the tumor to be removed before it becomes invasive. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the PDR is greater than a PDR threshold, such as 0.15.

[0027] In some embodiments, the method of treatment comprises calculating a first PDR from a first tumor sample obtained from the subject according to the methods described herein; treating the subject with an antitumor agent; and calculating a second PDR from a second tumor sample obtained from the subject according to the methods described herein. The method may comprise continuing to treat the subject with the antitumor agent if the second PDR is substantially the same as the first PDR. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the second PDR is different than the first PDR. The first and second tumor sample may be from the same tumor. The first and second tumor sample may be from a first tumor sample that was

treated such that no further tumor was detectable and the second tumor is from a relapse tumor or a tumor from a cancer that was in remission.

[0028] The subject may be treated with the antitumor agent for at least 3 months, at least 6 months, at least 9 months, at least 12 months, at least 24 months or at least 36 months before calculating the second PDR, preferably at least 12 months.

[0029] A fourth aspect of the present invention provides a method for identifying an antitumor agent that decreases the potential evolutionary capacity of cancer and, thus, the risk of relapse. In some embodiments the antitumor agent targets epigenetic proteins. In another embodiment the antitumor agent targets the DNA methylation machinery of the cell. In another embodiment the antitumor agent causes reprogramming of the DNA methylation within a cell. In another embodiment, the antitumor agent preferentially kills cells with a high potential evolutionary capacity as determined by calculating a PDR.

[0030] In some embodiments, the method comprises growing a first culture of hyperproliferative cells and a second culture of hyperproliferative cells, wherein the first culture is grown in the presence of an antitumor agent and the second culture is grown in the absence of the antitumor agent; detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells from the first culture and a plurality of cells from the second culture; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the first culture; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the second culture; and assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells.

[0031] In some embodiments, the method comprises treating an animal model of a cancer, wherein a first animal is treated with an antitumor agent and the second animal is treated with a placebo or no antitumor agent; detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells from the first animal and a plurality of cells from the second animal; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the first animal; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the second animal; and assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells. The antitumor agent decreases the potential evolutionary capacity of cancer if the level of inconsistent methylation status is less in the first animal than in the second animal. In another embodiment

the antitumor agent is administered to more than one cell or animal model using a range of doses.

[0032] The methylation status of neighboring CpG sites may be compared by calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism, or calculating information entropy. In some embodiments, the method comprises calculating a PDR, variance, epipolymorphism or information entropy in the first and second culture, wherein the antitumor agent decreases the potential evolutionary capacity of cancer if the PDR, variance, epipolymorphism or information entropy of the first culture is less than the PDR, variance, epipolymorphism or information entropy of the second culture.

[0033] The hyperproliferative cells in the first and second cultures may comprise cells from a cell line, e.g., a tumor cell line. Alternatively, the hyperproliferative cells in the first and second cultures may be cells from a tumor sample obtained from a subject, preferably a human, or cells cultured from such a sample. In some embodiments, the first and second cultures are the same culture, wherein the second culture is a sample of the hyperproliferative cells before addition of the antitumor agent and the first culture is a sample of the hyperproliferative cells after addition of the antitumor agent.

[0034] The animal model may be a model of any cancer. The animal model may be a mammal, more specifically a rodent, preferably a rat, and more preferably a mouse.

[0035] The first culture may be cultured in the presence of the antitumor agent for at least 6 hours, at least 12 hours, at least 18 hours, at least one day, at least two days, at least three days, at least four days, at least five days, at least six days or at least one week, preferably at least one day, prior to detecting methylation status.

[0036] The first animal may be treated with the antitumor agent for at least one day, one week, a month, 12 month's, 18 month's, 2 years, preferably at least one day, prior to detecting methylation status.

[0037] In some embodiments, the method comprises performing a first prognostic method on a first tumor sample from a subject, such as a laboratory animal, as described herein; administering an antitumor agent to the subject; and performing a second prognostic method on a second tumor sample from the subject as described herein, wherein the treatment is administered between the first and second prognostic methods. The antitumor agent decreases the potential evolutionary capacity of cancer if the level of inconsistent methylation status is less in the second tumor sample than in the first tumor sample.

[0038] The methylation status of neighboring CpG sites may be compared by calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism, or calculating information entropy. In some embodiments, the method comprises calculating a first PDR, variance, epipolymorphism or information entropy from a first tumor sample obtained from the subject, such as a laboratory animal, according to the methods described herein; treating the subject with an antitumor agent; and calculating a second PDR, variance, epipolymorphism or information entropy from a second tumor sample obtained from the subject according to the methods described herein. The antitumor agent decreases the potential evolutionary capacity of cancer if the second PDR, variance, epipolymorphism or information entropy is less than the first PDR, variance, epipolymorphism or information entropy.

[0039] The antitumor agent may be administered to the subject for at least one day, at least two days, at least three days, at least 4 days, at least five days, at least six days, at least one week, at least two weeks, at least three weeks, at least one month, preferably at least one week.

[0040] In any of the methods described herein, the tumor sample may be a solid tumor, such as carcinomas, sarcomas and lymphomas. In some embodiments, the solid tumor is selected from adrenocortical carcinoma, bone tumors, brain cancer, breast cancer, cervical cancer, colorectal carcinoma, desmoid tumors, desmoplastic small round cell tumors, endocrine tumors, esophageal cancer, Ewing sarcoma family tumors, gastric cancer, germ cell tumors, head or neck cancer, hepatoblastoma, hepatocellular carcinoma, lung cancer, melanoma, mesothelioma, nasopharyngeal carcinoma, neuroblastoma, non-rhabdomyosarcoma soft tissue sarcoma, osteosarcoma, ovarian cancer, pancreatic cancer, prostate cancer, retinoblastoma, rhabdomyosarcoma, skin carcinoma, testicular cancer, thyroid carcinoma, uterine cancer and Wilms tumors. The tumor sample may be a hematological cancer, such as leukemia, preferably CLL.

[0041] In any of the methods described herein, the antitumor agent may be selected from an angiogenesis inhibitor, such as angiostatin K1-3, DL- α -difluoromethyl-ornithine, endostatin, fumagillin, genistein, minocycline, staurosporine, and thalidomide; a DNA intercalator/cross-linker, such as Bleomycin, Carboplatin, Carmustine, Chlorambucil, Cyclophosphamide, cis-Diammineplatinum(II) dichloride (Cisplatin), Melphalan, Mitoxantrone, and Oxaliplatin; a DNA synthesis inhibitor, such as ($\pm$)-Amethopterin (Methotrexate), 3-Amino-1,2,4-benzotriazine 1,4-dioxide, Aminopterin, Cytosine β -D-arabino-furanoside, 5-Fluoro-5'-deoxyuridine, 5-Fluorouracil, Ganciclovir, Hydroxyurea, and Mitomycin C; a DNA-RNA transcription regulator, such as Actinomycin D, Daunorubicin, Doxorubicin, Homoharringtonine, and Idarubicin; an enzyme inhibitor, such as S(+)-Camptothecin, Curcumin, (-)-Deguelin, 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside, Etoposide, Formestane, Fostriecin, Hispidin, 2-Imino-1-imidazoli-dineacetic acid (Cyclocreatine), Mevinolin, Trichostatin A, Tyrphostin AG 34, and Tyrphostin AG 879; a gene regulator, such as 5-Aza-2'-deoxycytidine, 5-Azacytidine, Cholecalciferol (Vitamin D3), 4-Hydroxytamoxifen, Melatonin, Mifepristone, Raloxifene, all trans-Retinal (Vitamin A aldehyde), Retinoic acid, all trans (Vitamin A acid), 9-cis-Retinoic Acid, 13-cis-Retinoic acid, Retinol (Vitamin A), Tamoxifen, and Troglitazone; a microtubule inhibitor, such as Colchicine, docetaxel, Dolastatin 15, Nocodazole, Paclitaxel, Podophyllotoxin, Rhizoxin, Vinblastine, Vincristine, Vindesine, and Vinorelbine (Navelbine); and an unclassified antitumor agent, such as 17-(Allylamino)-17-demethoxygeldanamycin, 4-Amino-1,8-naphthalimide, Apigenin, Brefeldin A, Cimetidine, Dichloromethylene-diphosphonic acid, Leuprolide (Leuprolerin), Luteinizing Hormone-Releasing Hormone, Pifithrin- α , Rapamycin, Sex hormone-binding globulin, Thapsigargin, and Urinary trypsin inhibitor fragment (Bikunin). The antitumor agent may be a monoclonal antibody such as rituximab (Rituxan®), alemtuzumab (Campath®), Ipilimumab (Yervoy®), Bevacizumab (Avastin®), Cetuximab (Erbitux®), panitumumab (Vectibix®), and trastuzumab (Herceptin®), Vemurafenib (Zelboraf®) imatinib mesylate (Gleevec®), erlotinib (Tarceva®), gefitinib (Iressa®), Vismodegib (Erivedge™), 90Y-ibritumomab tiuxetan, 131I-tositumomab, ado-trastuzumab

emtansine, lapatinib (Tykerb®), pertuzumab (Perjeta™), ado-trastuzumab emtansine (Kadcyla™), regorafenib (Stivarga®), sunitinib (Sutent®), Denosumab (Xgeva®), sorafenib (Nexavar®), pazopanib (Votrient®), axitinib (Inlyta®), dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (Bosulif®), ofatumumab (Arzerra®), obinutuzumab (Gazyva™), ibrutinib (Imbruvica™), idelalisib (Zydelig®), crizotinib (Xalkori®), erlotinib (Tarceva®), afatinib dimaleate (Gilotrif®), ceritinib (LDK378/Zykadia), Tositumomab and 131I-tositumomab (Bexxar®), ibritumomab tiuxetan (Zevalin®), brentuximab vedotin (Adcetris®), bortezomib (Velcade®), siltuximab (Sylvant™), trametinib (Mekinist®), dabrafenib (Tafinlar®), pembrolizumab (Keytruda®), carfilzomib (Kyprolis®), Ramucirumab (Cyramza™), Cabozantinib (Cometriq™), vandetanib (Caprelsa®), Optionally, the antitumor agent is a neoantigen. The antitumor agent may be a neoantigen. Neoantigens are tumor-associated peptides that serve as active pharmaceutical ingredients of vaccine compositions which stimulate antitumor responses and are described in US 2011-0293637, which is incorporated by reference herein in its entirety. The antitumor agent may be a cytokine such as interferons (INFs), interleukins (ILs), or hematopoietic growth factors. The antitumor agent may be INF- α , IL-2, Aldesleukin IL-2, Erythropoietin, Granulocyte-macrophage colony-stimulating factor (GM-CSF) or granulocyte colony-stimulating factor. The antitumor agent may be a targeted therapy such as toremifene (Fareston®), fulvestrant (Faslodex®), anastrozole (Arimidex®), exemestane (Aromasin®), letrozole (Femara®), ziv-aflibercept (Zaltrap®), Alitretinoin (Panretin®), temsirolimus (Torisel®), Tretinoin (Vesanoid®), denileukin diftitox (Ontak®), vorinostat (Zolinza®), romidepsin (Istodax®), bexarotene (Targretin®), pralatrexate (Folotyn®), lenalidomide (Revlimid®), belinostat (Beleodaq™), lenalidomide (Revlimid®), pomalidomide (Pomalyst®), Cabazitaxel (Jevtana®), enzalutamide (Xtandi®), abiraterone acetate (Zytiga®), radium 223 chloride (Xofigo®), or everolimus (Afinitor®). The antitumor agent may be a checkpoint inhibitor such as an inhibitor of the programmed death-1 (PD-1) pathway, for example an anti-PD1 antibody (Nivolumab). The inhibitor may be an anti-cytotoxic T-lymphocyte-associated antigen (CTLA-4) antibody. The inhibitor may target another member of the CD28 CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. A checkpoint inhibitor may target a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3. Additionally, the antitumor agent may be an epigenetic targeted drug such as HDAC inhibitors, kinase inhibitors, DNA methyltransferase inhibitors, histone demethylase inhibitors, or histone methylation inhibitors. The epigenetic drugs may be Azacitidine (Vidaza), Decitabine (Dacogen), Vorinostat (Zolinza), Romidepsin (Istodax), or Ruxolitinib (Jakafi).

[0042] In any of the methods herein, the subject may be a mammal, preferably a human. In some embodiments, the subject may be a laboratory animal, such as a mouse, a rabbit, a rat, a guinea pig, and a hamster. In other embodiments the subject may be a primate or ungulate.

[0043] Accordingly, it is an object of the invention to not encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does

not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. §112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product.

[0044] It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as “comprises”, “comprising”, “comprised”, “comprising” and the like can have the meaning attributed to it in U.S. patent law; e.g., they can mean “includes”, “included”, “including”, and the like; and that terms such as “consisting essentially of” and “consists essentially of” have the meaning ascribed to them in U.S. patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

[0045] These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0046] The following detailed description, given by way of example, but not intended to limit the invention solely to the specific embodiments described, may best be understood in conjunction with the accompanying drawings.

[0047] FIG. 1A-B demonstrates the effect of clonal evolution on treatment of CLL. (A) Comparing cancer cell fractions (CCF) at two timepoints can distinguish CLL with and without clonal evolution. (B) Pre-treatment subclonal driver presence is associated with shorter time to retreatment or death.

[0048] FIG. 2A-D demonstrates differential methylation between normal B cells and CLL samples. (A) and (B) 450K methylation arrays showing different distribution in CLL vs. normal B cells, resulting in higher proportion of markers with intermediate methylation state. (C) Higher cumulative distribution of intermediate methylation seen in WGBS of a CLL sample vs. a normal B cell; data generated by Kulis et al. *Nat Genet* 2012; 44:1236-42. (D) Higher methylation heterogeneity in CLL demonstrated with RRBS.

[0049] FIG. 3A-F illustrates that higher DNA methylation intra-sample heterogeneity in CLL arises from locally disordered methylation. A. CLL Global and CGI methylation compared with normal B cells, measured with WGBS (top). Cumulative distribution analysis (bottom) enables the comparison of the proportion of intermediate methylation values in WGBS data of CLL and B cells from healthy adult volunteers. B. Mean intra-sample CpG variance measured with RRBS. C. Methylation patterns from RRBS data of a CLL sample (CLL007), show two patterns of methylation (black circles—methylated CpGs; white circles—unmethylated): (1) A pattern compatible with a mixture of cell populations with clear but distinct methylation states for a particular non-imprinted locus (left—SDHAP3 promoter [chr5:1594239-1594268]), and (2) a pattern compatible with an admixture of cells with locally disordered methylation (right—PIK3R5 promoter [chr17:8869616-8869640]). D. A comparison between the intra-sample CpG variance that arises from discordant compared with concordant reads across the 104 CLLs. E. CpG methylation and the proportion of discordant reads (PDR) were calculated as shown. F. Sample average PDR for CLL, cancer cell lines, normal B

cells and a collection of primary healthy human tissues. To enable an accurate comparison between samples, sample average PDR is calculated based on a consensus set of 63,443 CpGs that are covered with greater than 10 reads in >75% of all 202 RRBS samples.

[0050] FIG. 4A-O illustrates WGBS and RRBS data from CLLs and normal B cells shows higher Intratumoral DNA methylation heterogeneity that arises from locally disordered methylation. A-C (top). The genome was divided into 1 KB tiles. The analysis was limited to tiles that contained at least 5 CpGs covered with greater than 5 reads. The scatter plots enable the examination of the methylation patterns consistency of the two B cell samples (A), and the two CLL samples (B). Note that the somewhat decreased methylation in CLL007 perhaps results from the DNMT3A nonsense mutation affecting this tumor. A comparison between the average methylation values across the genome in CLL and normal B cells is also shown (C). A-C (bottom). The proportion of genomic 1 KB tiles with intermediate values is compared between CLL and normal B cells. D. The percentage of methylation values falling within each category (0-0.2, 0.2-0.6, 0.6-1.0) is shown for the 4 WGBS samples for different genomic features. Number of CpGs per sample per feature (mean [range]): Promoters—1,737,131 [1,728,620-1,747,890], CG islands—2,031,560 [2,025,376-2,044,203], enhancers—865,820 [860,997-870,134], exons—1,489,549 [1,483,138-1,493,987], introns—6,691,529 [6,599,956-6,739,995] and repeat elements—7,301,495 [7,163,887-7,368,831]. E. Reanalysis of WGBS data (Kulis et al., 2012) for the frequency of CpGs with intermediate methylation in CLL samples compared with B cells from healthy adult volunteers. Shown are cumulative distributions of CpG methylation values in unmutated IGHV CLL compared with naive B cells (left), as well as for mutated IGHV CLL vs. memory B cells (cs—class switched) (middle). The intratumoral DNA methylation heterogeneity in CLL from discordant reads (solid line) versus concordant reads (dashed) (right, analysis of WGBS data from CLL169). F. Histograms of individual normal B cell samples (blue) show bimodal distribution in methylation values as measured by DNA 450K methylation arrays (Kulis et al., 2012), while CLL samples (red) show more CpGs with intermediate methylation values, diverging from a pure bimodal distribution. G-H. Violin plots comparing the proportion of intermediate methylation values from 450K array data (Kulis et al. (Kulis et al., 2012)) from 127 CLL samples and 26 normal B cell samples (beta methylation values between 0.35 and 0.65, average±SEM, 13.7±0.002% vs. 10.1±0.01%, respectively, $p=5\times 10^{-8}$, Wilcoxon rank sum test). I. While overall purity of the CLL samples was consistently high (median of 90.2%), contaminating non-malignant cells in samples may contribute to the PDR, Applicants therefore compared the PDR in CLL samples with high vs. low tumor purity (above and below the overall average; 86.6%). J. Stochastic disorder in methylation patterns is expected to yield discordant reads that involve both parental alleles in a given locus (in contrast to an allele-specific methylation (ASM) phenomenon). We therefore measured the proportion of germline SNPs for which a discordant read is found to involve both parental alleles (Y axis). As expected, with an increasing number of discordant reads in the studied locus (X axis), the proportion of SNPs with a discordant read involving both parental alleles increases and converges towards 1. K. Even within a given genotype, different

methylation patterns were seen. For example, in the left most panel, 3 distinct methylation patterns are seen to affect both the A genotype parental allele and the G genotype parental allele. L. Applicants measured the number of distinct discordant methylation patterns found in each locus (similar to a previous analysis (Landan et al., 2012)). Presence of 1 or 2 patterns of discordancy across all reads covered for a particular locus would be expected of ASM. The plot shows the distribution of the number of methylation patterns in loci with 10-20 discordant reads across 10 randomly selected CLL and normal B cell samples. The distribution shows that there are generally more than 2 discordant methylation patterns per locus for both normal (blue) and CLL (red) samples. In addition, the high number of distinct methylation profiles per locus excludes also the possibility that PDR arises from reads that cover an ordered transition point from one methylation state to another. The shaded distribution (grey) shows the number of distinct patterns if the state of CpG methylation was purely random (with equal frequencies of the number of reads as in the experimental data). The finding that the measured distribution demonstrates less distinct patterns than purely random is consistent with inheritance of discordant patterns to progeny cells. M. To assess for possible amplification biases, the allelic frequencies of germline SNP not involving CpGs was measured and shows a tight distribution around 0.5 compatible with limited amplification biases. N. To assess for possible amplification biases, the methylation of imprinted control regions was measured and shows a tight distribution around 0.5 compatible with limited amplification biases. O. Similar PDR values are seen in regions of somatic copy number variations (sCNV) in the two CLLs that underwent WGBS (CLL169 and CLL007), both for promoter CpGs (top) and for all CpGs (bottom).

[0051] FIG. 5A-C demonstrates that increased locally disordered DNA methylation is the predominant source of the epigenetic heterogeneity in CLL. (A) Bisulfite MPS distinguishes between two sources of methylation heterogeneity. (B) Most of the overall variance in DNA methylation results from discordant reads. (C) The proportion of discordant reads (PDR) is higher in CLL.

[0052] FIG. 6A-D demonstrates the genomic location of locally disordered methylation. (A) Higher PDR affects most genomic features, with unique patterns in CpG islands and promoters, as seen in (B). (C) Similarly increased PDR is seen in cancer cell lines compared with normal human tissues. (D) Increased PDR in repetitive regions shown with WGBS.

[0053] FIG. 7A-C illustrates that locally disordered methylation affects all genomic regions in CLL, including CpG islands (CGIs) and repeat regions. Comparison of mean PDR (A) and mean CpG methylation (B) per genomic region between CLLs and normal B cells using RRBS data (Table 2 provides the average number of CpGs analyzed for each genomic region). Error bars represent upper 95% CI of the mean. C. Top—The distribution of PDR and methylation across all promoters covered by RRBS for randomly selected 6 CLL and 6 normal B cell samples. The distribution was derived by dividing each promoter into 100 bins, and then averaging methylation and PDR for CpGs falling into each bin across all promoters in the sample. The PDR and methylation values in the adjacent 2 KB upstream and downstream are also shown. Bottom—An analogous analysis of CGIs and adjacent shore regions.

[0054] FIG. 8A-C illustrates genomic characterization of locally disordered methylation including analysis of CpG island subtypes and repeat elements based on WGBS. A. Increase in PDR with concomitant increase in methylation in 104 CLLs compared to 26 B cell samples affects all 3 major categories of CpG islands (CGIs: promoters, intragenic, other). B. WGBS based analysis of the 2 CLL samples (CLL007 and CLL169), compared with 2 normal B cell samples (Normal_CD19_1 and Normal_IGD_3) showing increased PDR in repeat elements concomitant with decreased methylation. Comparison between PDR and methylation values individually between each CLL sample and each normal B cell sample yielded a statistically significant difference ($p < 1 \times 10^{-32}$). C. The three of 104 CLL samples with nonsilent mutations in methylation modulators (DNMT3A-Q153*, TET1-N789I, IDH1-S210N) revealed high average PDR by RRBS compared to samples with wildtype alleles for these genes.

[0055] FIG. 9A-H illustrates that locally disordered methylation in CLL is consistent with a stochastic process. A. As an additional measure of methylation disorder in individual reads, Applicants calculated Shannon's information entropy (Shannon, 1948) for intra-sample methylation variation. Information entropy was calculated for each read and then averaged across all reads for each CpG as shown. B. Increased average Shannon's information entropy was observed in CLL and cancer cell line samples compared with normal B cells and primary healthy diverse human tissue samples, demonstrating an increase in stochastic methylation variation. C. An increase in information entropy is seen across all measured regions in RRBS data from CLL samples (red) compared with B cells from healthy adult volunteers (blue). Error bars indicate upper 95% CI. Relative increase in average entropy from B cells to CLL samples and p value for Wilcoxon rank sum test are shown. D. Analysis of outlier genes falling outside of the expected distribution of PDR in relation to methylation level. Left panels—Outlier genes (black) were identified by the Tukey method in which promoter CGI PDR was lower than expected given the methylation level. Right panel—the comparative location of selected gene promoters in CLL (red) compared with normal B cells (blue). This plot highlights the considerable CLL hypermethylation without a significant concomitant change in PDR in tumor suppressor genes (WIF1, DCC, DUSP22; solid circles). In contrast, imprinted genes (empty circles, e.g., GNAS) show relative little difference between CLL and normal B cells. E. Scatter plots for methylation and PDR values were generated for a CLL sample (CLL169) and a normal B cell sample (Normal_CD19_1). Values were calculated for each element (enhancers, promoter CGIs, exons, introns, LINE family repeat elements and LTR family repeat elements) as long as at least 20 evaluable CpGs were contained in the specific element, with at least 4 CpG per read and read depth > 10 ("unfiltered"). The same data were analyzed with filtering such that only CpGs covered by reads with 4-6 CpGs per read (similar to RRBS data) were examined (second row), or such that a more stringent criteria on the number of evaluable CpGs (> 100) per evaluated element was used. Together the plots follow the same distribution of PDR to methylation values suggestive of a stochastic change in methylation (FIG. 3A). F. A scatter plot for methylation and PDR values for promoter CGIs utilizing RRBS data (CLL007 compared with Normal_IGD_2). G. Similar distribution can be seen

for the methylation and PDR values of promoter regions of the key tumor suppressor genes DAPK1 and WT1 across CLL samples. H. The strong correlation between average promoter CGI PDR and methylation across 104 CLL samples is shown separately for 3 groups of genes, arranged according to their average methylation values across 104 CLLs (0-0.1, left; 0.1-0.5, center; 0.5-1.0, right).

[0056] FIG. 10A-G illustrates that locally disordered methylation in CLL is consistent with a stochastic process. A. Applicants developed a model to determine the probability of observing any PDR value in a random CpG methylation state model [given: (1) the total number of reads that cover the locus, (2) the number of neighboring CpGs contained in individual reads, and (3) the locus methylation level]. The plot demonstrates the case in which a locus is covered at a read depth of 30 and each read contains 4 neighboring CpGs. The expected PDR value is shown by the dashed line, and the shaded region represents methylation-PDR tuples with a probability greater than 0.01 under the random model. B. The CLL methylation data are consistent with the stochastic pattern shown in (A). Average promoter CGI methylation and PDR were calculated for 13,943 CGIs covered by WGBS (>10 CpGs per island) in both the CLL and the normal B cell samples. Outliers represent 1.4% of events. C. Average LINE element methylation and PDR were calculated for 1,894 elements covered by WGBS (>20 CpGs per element) in the same samples as in (B). D. The correlation in CLL between sample average of CGI methylation and PDR is shown ($8,740.2 (\pm 3,102.8)$ promoter CGIs per sample were evaluated). E. Similarly, the correlation in CLL between sample average LINE element methylation and PDR are also shown. The RRBS based results of CLL169 are highlighted with a purple square. F. To study the correlation between difference in PDR (Δ PDR) and difference in methylation (Δ Meth), Applicants paired representative CLL and normal B cell samples. For each promoter (>20 CpGs per promoter, $n=2119$), Δ Meth and Δ PDR were plotted (red). An identical procedure was performed with a pairing of the same normal B cell sample to an adult lung sample (Lung_normal_BioSam_235, blue). These data enable the comparison between the Pearson's coefficient for the correlation between Δ PDR and Δ Meth in cancer related changes vs. normal physiological state changes. G. To confirm this finding across the entire dataset, random pairings were performed in each category listed on the X-axis, avoiding repeated use of any individual sample within a category. This procedure was repeated 100 times, and the means of the correlation coefficients for each iteration are plotted and compared.

[0057] FIG. 11A-E illustrates that locally disordered methylation affects preferentially gene-poor regions and can be traced back to non-expressed genes in normal B cells. A. Promoter PDR (orange, error bars represent 95% CI of means) in relation to gene density (genes/MB, left) and CTCF binding site density (right) regions. As reference, the CpG content is also provided (black). B. PDR and methylation in hypomethylated blocks (Hansen et al., 2011) is plotted for CLL and normal B cells (shown are blocks with >1,000 CpGs in WGBS). C. Replication time and PDR are correlated; PDR was averaged for each promoter covered in >70% of 104 CLLs, and these values were grouped in replication time bins. D. To assess the relationship between somatic mutations and PDR, sSNVs were identified with whole genome sequencing of matched tumor and germline

DNA (CLL169). Average PDR (left) and methylation (right) were measured in 1,000 bp increments from each somatic mutation. Values of CpGs in each 1,000 bp bin were averaged over 4,973 sSNVs, and plotted as a function of the distance from the somatic mutation. Orange lines—the LOWESS (locally weighted scatterplot smoothing). E. Left—promoter CGI PDR is correlated between CLL and normal B cells samples (Pearson, evaluated with 5,811 consistently covered CGIs). Right—Promoter CGI PDR in B cells and CLLs is shown for genes expressed and non-expressed in normal B cells (FPKM<1, $n=1,002$ from RNAseq data of 7 healthy donor B cell samples).

[0058] FIG. 12A-C illustrates the association between PDR and distance from somatic mutation is similar for clonal and subclonal mutations. A. To study the specificity of the PDR increase in the previously defined hypomethylated blocks (Hansen et al., 2011), Applicants identified size, GC and repeat content matched regions at random from the genome. Of these regions, Applicants retained only those that harbored more than 1000 CpGs each (covered with greater than 10 reads and 4 or more CpGs per read in the CLL169 and Normal_CD19_1 WGBS data). Compared to the control genomic regions, the hypomethylated blocks exhibit higher PDR in both CLL and normal B cells, as well as a greater increase in CLL compared to the normal B cells. To assess for a relationship between somatic mutations and PDR, somatic single nucleotide variants (sSNVs) were identified with WGS of CLL169 and matched germline DNA. Subsequently, sSNVs with sufficient read depth (>40) were classified as clonal ($n=866$) or subclonal ($n=602$) based on the allelic frequency (above or below 0.2, respectively, analysis limited to sSNVs with greater than 40 reads and that do not involve sCNVs to enhance the confidence in the clonal vs. subclonal classification). Average PDR (B) and methylation (C) were measured in 1000 bp increments from each somatic mutation. Values for each 1000 bp bin were averaged over sSNVs, and plotted as a function of the distance from the somatic mutation. Red lines—the LOWESS (locally weighted scatterplot smoothing).

[0059] FIG. 13A-B demonstrates that locally disordered methylation contributes to the increase transcriptional variation found to be associated with epigenetic heterogeneity. (A) The likelihood of methylation associated gene silencing is lower in genes with high PDR. (B) Allele-specific expression is associated with higher promoter PDR.

[0060] FIG. 14A-F illustrates that locally disordered methylation is associated with transcriptional variation. A. Mean promoter PDR and gene expression are correlated (evaluated with 8,570 genes that had promoter RRBS coverage in >70% of 33 samples with matched RRBS and RNAseq). B. PDR and expression variability as measured with coefficient of variation (CV) of 5,874 transcribed genes (FPKM>1). Black circles (brackets)—mean CV (95% CI) for genes within PDR bins (number of genes per bin in blue). Red line—cubic smoothing spline of CV and PDR values (unbinned). Note that the analysis was limited to transcribed genes to avoid an artificial enhancement of CV that occurs with very low mean expression values. As >97.5% of transcribed genes had PDR<0.3, Applicants limited the X axis to PDR <0.3. C. Left—Odds ratio (bars—95% CI) for gene expression (FPKM>1) with a methylated promoter (average methylation>0.8) versus unmethylated promoter (average methylation<0.2) is calculated for genes with high (orange, $27.5 \pm 2.6\%$ of genes) or low promoter PDR (black).

Right—Linear models that combine information from all 33 CLLs as continuous variables to predict expression. D. PDR and intra-sample gene expression heterogeneity (assessed by Shannon's information entropy) across the range of population average expression (FPM—fragments per million), by single cell RNA sequencing of 84 cells from CLL005. Local regression lines for genes with low PDR (0-0.05, blue), intermediate PDR (0.05-0.2, purple) and high PDR (0.2-1.0, red) are shown. E. Results of generalized additive regression tests that model single cell gene expression Shannon's information entropy based on PDR, population average expression, and transcript length across the 4 CLL samples. F. Single cell gene expression patterns for genes within a narrow population average expression range of 1.0-1.2 (black rectangle in panel D). Consistent with the higher gene expression Shannon's information entropy observed in genes with higher PDR (top), genes with low PDR (bottom left) tend to be expressed at high magnitude (larger dot size) in fewer cells, while genes with high PDR (bottom right) are frequently expressed at low expression magnitudes across many cells.

[0061] FIG. 15A-E illustrates that locally disordered methylation is linked to transcriptional variation. A. Genes were divided into 9 bins according to their mean expression over 33 samples (starting from 0, and then in increments of 1 until 9; $\log 2[\text{FPKM}+1]$). PDR is shown for each bin in boxplots, demonstrating that PDR is highest in genes with low expression values (bottom). The number of genes in each expression bin is shown (top). B. Density scatter plot of mean promoter methylation in relation to mean expression ($\log 2[\text{FPKM}+1]$), showing that these features are negatively correlated. 8,570 genes were evaluated that had promoter RRBS coverage in at least 70% of 33 samples with matched RRBS and RNAseq. C. An example is shown of the promoter region of ZNF718 from two samples (CLL062 and CLL074) with similar promoter methylation values but different PDR and different expression as measured by RNAseq (bottom right). ZNF718 promoter RRBS reads for CLL062 and CLL074 are shown (top). The number of concordantly methylated (grey background) or discordantly methylated (orange background) sequencing reads for each distinct methylation pattern is indicated to the right of each read pattern. D. Gene expression Shannon's information entropy (y-axis) in relation to the population average gene expression (x-axis, $\log 10[\text{FPM}]$) for each gene covered in single cells of CLL032, CLL096 and CLL146, evaluated by single cell transcriptome sequencing. Colored lines—local regression curves for genes with low PDR (0-0.05, blue), intermediate PDR (0.05-0.2, purple), and high PDR (0.2-1.0, red). 90% of genes with higher promoter PDR (PDR >0.1) have lower population average expression (bounded by the yellow highlighted line). Right panels—Boxplots of the gene expression Shannon's information entropy for each of the three PDR bins for genes with population average gene expression of 1.0-1.5 (to control for differences in this variable). E. Generalized additive regression tests that model gene expression Shannon's information entropy based on: PDR, population average gene expression (locally smoothed), transcript length and promoter methylation across the 4 CLL samples that underwent single-cell transcriptome sequencing.

[0062] FIG. 16A-C demonstrates that locally disordered methylation creates a rich substrate for CLL evolution by stochastic variation amenable to positive selection. (A)

Higher PDR in genes down-regulated in CLL. (B) Higher PDR in genes known to be up-regulated in stem cells. (C) Higher PDR in samples with a higher number of subclonal mutations, but not clonal mutations.

[0063] FIG. 17A-B illustrates that increased locally disordered methylation involves differentially methylated regions and affects stem cell related genes. A. Two sets of differentially methylated CpG islands and promoter regions were identified by comparing methylation across: i) unmutated IGHV CLL vs. normal naive B cell samples, and ii) across mutated IGHV CLL vs. normal memory B cell samples. Significantly differentially methylated regions were defined as having a >10% average methylation change with a t-test p value <0.01. Average PDR was then calculated for each one of these regions. Higher PDR was measured in differentially methylated (both increased and decreased methylation) promoters and CpG islands compared with regions that are not differentially methylated between CLL and normal B cells (Wilcoxon rank sum test). B. Average promoter PDR is highest in promoters of 115 genes up-regulated in stromal stem cells compared with 102 genes down-regulated in stromal stem cells (Boquest et al., 2005) as well as the average for 8,353 genes without a differential expression in stem cells (all comparisons by Wilcoxon rank sum test). Boxes represent median and interquartile range (IQR). Whiskers represent 1.5 times IQR.

[0064] FIG. 18A-E illustrates that locally disordered methylation may interact with evolution through drift towards a stem-like state. A. Gene set enrichment analysis comparing 1,668 genes with consistently high promoter PDR (>0.1 in >75% of samples) to 5,392 genes with consistently low promoter PDR (<0.1 in >75% of samples, selected 10 gene sets displayed). Enrichment in genes with consistently high PDR was calculated for hypergeometric distribution followed by BH-FDR ('Q(high)'). In addition, enrichment in high PDR genes vs. low PDR genes was calculated using Fisher's exact test followed by BH-FDR ('Q(high vs low)'). B. PDR and methylation in regions hypomethylated in embryonic stem cells (Ziller et al., 2013), in CLL compared with normal B cells (WGBS data). Regions include 91 enhancers (e.g., POU5F1, NANOG), 41 enhancer CGIs (e.g., TET2, EP400), 6 CGIs (e.g., DAPK1), 6 promoters and 84 other putative regulatory elements (e.g., DEC1 and POT1) (Ziller et al., 2013). The inset shows individual changes of selected regions. C. PDR in CLLs with high vs. low number of subclonal (median=7.5 sSNVs) and clonal mutations (median=10 sSNVs). D. Fourteen CLLs were sampled longitudinally at two time points (T1, T2; median interval time—3.5 years), and change in PDR over time was compared between CLLs that underwent genetic clonal evolution (n=9) and those without genetic evolution (n=5, paired t test). E. Gene set enrichment of the 899 genes from the 14 cases with significant promoter methylation change between timepoints T1 and T2 (absolute change >10%, FDR BH Q<0.1) in genes with promoter demethylation over time (456 genes), and in genes with promoter methylation over time (443 genes) see Table 6 for top 30 enrichments).

[0065] FIG. 19A-B illustrates that locally disordered methylation is associated with adverse clinical outcome. A. Kaplan-Meier plot showing failure free survival time (failure defined as retreatment or death from the time of first therapy after RRBS analysis) in CLLs with higher versus lower than average promoter PDR. Note that the analysis

could only be performed for the 49 patients who received therapy after RRBS sampling. B. Multivariable analysis for this association with the addition of well-established poor outcome predictors in CLL (IGHV unmutated status, del(17p) and del(11q)), as well as with the addition of the presence of a subclonal driver (including somatic copy number changes, sSNVs and indels), as previously described (Landau et al., 2013)) to the model.

[0066] FIG. 20A-B illustrates a model of the interaction between methylation disorder and clonal evolution. A novel somatic mutation (depicted with lightning bolts) would have to coincide with an epigenetic state that will be permissive to the propagation of the new genotype to a progeny population. In a cellular population with limited stochastic methylation changes (top panel), the proportion of cells that are therefore able to actively participate in the evolutionary process is small. However, in a more malleable epigenetic landscape resulting from high level of locally disordered methylation, a greater proportion of cells can give birth to new subclones, increasing the diversity and the adaptive capacity of the cancer population, resulting in adverse clinical outcome with therapy.

DETAILED DESCRIPTION OF THE INVENTION

[0067] In order that the invention described herein may be fully understood, the following detailed description is set forth.

[0068] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as those commonly understood by one of skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. The materials, methods and examples are illustrative only, and are not intended to be limiting. All publications, patents and other documents mentioned herein are incorporated by reference in their entirety.

[0069] Throughout this specification, the word “comprise” or variations such as “comprises” or “comprising” will be understood to imply the inclusion of a stated integer or groups of integers but not the exclusion of any other integer or group of integers.

[0070] The terms “antitumor agent” and “chemotherapeutic agent” are used interchangeably herein and refer to an agent for the treatment of cancer. Typically, an antitumor agent is a cytotoxic anti-neoplastic drug, which is administered as part of a standardized regimen. Without being bound by theory, antitumor agents act by killing cells that divide rapidly, one of the main properties of most cancer cells. Preferably, the antitumor agent is not indiscriminately cytotoxic, but rather targets proteins that are abnormally expressed in cancer cells and that are essential for their growth. Non-limiting examples of antitumor agents include: angiogenesis inhibitors, such as angiostatin K1-3, DL- α -difluoromethyl-ornithine, endostatin, fumagillin, genistein, minocycline, staurosporine, and ($\pm$)-thalidomide; DNA intercalator/cross-linkers, such as Bleomycin, Carboplatin, Carmustine, Chlorambucil, Cyclophosphamide, cis-Diammineplatinum(II) dichloride (Cisplatin), Melphalan, Mitoxantrone, and Oxaliplatin; DNA synthesis inhibitors, such as ($\pm$)-Amethopterin (Methotrexate), 3-Amino-1,2,4-benzotriazine 1,4-dioxide, Aminopterin, Cytosine 3-D-arabino-fura-

noside, 5-Fluoro-5'-deoxyuridine, 5-Fluorouracil, Ganciclovir, Hydroxyurea, and Mitomycin C; DNA-RNA transcription regulators, such as Actinomycin D, Daunorubicin, Doxorubicin, Homoharringtonine, and Idarubicin; enzyme inhibitors, such as S(+)-Camptothecin, Curcumin, (–)-Deguelin, 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside, Etoposide, Formestane, Fostriecin, Hispidin, 2-Imino-1-imidazoli-dineacetic acid (Cyclocreatine), Mevinolin, Trichostatin A, Tyrphostin AG 34, and Tyrphostin AG 879; gene regulators, such as 5-Aza-2'-deoxycytidine, 5-Azacytidine, Cholecalciferol (Vitamin D3), 4-Hydroxytamoxifen, Melatonin, Mifepristone, Raloxifene, all trans-Retinal (Vitamin A aldehyde), Retinoic acid, all trans (Vitamin A acid), 9-cis-Retinoic Acid, 13-cis-Retinoic acid, Retinol (Vitamin A), Tamoxifen, and Troglitazone; microtubule inhibitors, such as Colchicine, Dolastatin 15, Nocodazole, Paclitaxel, Podophyllotoxin, Rhizoxin, Vinblastine, Vincristine, Vindesine, and Vinorelbine (Navelbine); and unclassified antitumor agents, such as 17-(Allylamino)-17-demethoxygeldanamycin, 4-Amino-1,8-naphthalimide, Apigenin, Brefeldin A, Cimetidine, Dichloromethylene-diphosphonic acid, Leuprolide (Leuprorelin), Luteinizing Hormone-Releasing Hormone, Pifithrin- α , Rapamycin, Sex hormone-binding globulin, Thapsigargin, and Urinary trypsin inhibitor fragment (Bikunin). The antitumor agent may be a neoantigen. Neoantigens are tumor-associated peptides that serve as active pharmaceutical ingredients of vaccine compositions which stimulate antitumor responses and are described herein in its entirety. The antitumor agent may be a monoclonal antibody such as rituximab (Rituxan®), alemtuzumab (Campath®), Ipilimumab (Yervoy®), Bevacizumab (Avastin®), Cetuximab (Erbix®), panitumumab (Vectibix®), and trastuzumab (Herceptin®), Vemurafenib (Zelboraf®) imatinib mesylate (Gleevec®), erlotinib (Tarceva®), gefitinib (Iressa®), Vismodegib (Erivedge™), 90Y-ibritumomab tiuxetan, 131I-tositumomab, ado-trastuzumab emtansine, lapatinib (Tykerb®), pertuzumab (Perjeta™), ado-trastuzumab emtansine (Kadcyla™), regorafenib (Stivarga®), sunitinib (Sutent®), Denosumab (Xgeva®), sorafenib (Nexavar®), pazopanib (Votrient®), axitinib (Inlyta®), dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (Bosulif®), ofatumumab (Arzerra®), obinutuzumab (Gazyva™), ibrutinib (Imbruvica™), idelalisib (Zydelig®), crizotinib (Xalkori®), erlotinib (Tarceva®), afatinib dimaleate (Gilotrif®), ceritinib (LDK378/Zykadia), Tositumomab and 131I-tositumomab (Bexxar®), ibritumomab tiuxetan (Zevalin®), brentuximab vedotin (Adcetris®), bortezomib (Velcade®), siltuximab (Sylvant™), trametinib (Mekinist®), dabrafenib (Tafinlar®), pembrolizumab (Keytruda®), carfilzomib (Kyprolis®), Ramucirumab (Cyramza™), Cabozantinib (Cometriq™), vandetanib (Caprelsa®). The antitumor agent may be a cytokine such as interferons (INFs), interleukins (ILs), or hematopoietic growth factors. The antitumor agent may be INF- α , IL-2, Aldesleukin, IL-2, Erythropoietin, Granulocyte-macrophage colony-stimulating factor (GM-CSF) or granulocyte colony-stimulating factor. The antitumor agent may be a targeted therapy such as toremifene (Fareston®), fulvestrant (Faslodex®), anastrozole (Arimidex®), exemestane (Aromasin®), letrozole (Femara®), ziv-aflibercept (Zaltrap®), Alitretinoin (Panretin®), temsirolimus (Torisel®), Tretinoin (Vesanoid®), denileukin diftitox (Ontak®), vorinostat (Zolinza®), romidepsin (Istodax®),

bexarotene (Targretin®), pralatrexate (Folotyn®), lenalio-
mide (Revlimid®), belinostat (Beleodaq™), lenalio-
mide (Revlimid®), pomalidomide (Pomalyst®), Cabazitaxel
(Jevtana®), enzalutamide (Xtandi®), abiraterone acetate
(Zytiga®), radium 223 chloride (Xofigo®), or everolimus
(Afinitor®). The antitumor agent may be a checkpoint
inhibitor such as an inhibitor of the programmed death-1
(PD-1) pathway, for example an anti-PD1 antibody
(Nivolumab). The inhibitor may be an anti-cytotoxic T-lym-
phocyte-associated antigen (CTLA-4) antibody. The inhibi-
tor may target another member of the CD28 CTLA4 Ig
superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. A
checkpoint inhibitor may target a member of the TNFR
superfamily such as CD40, OX40, CD137, GITR, CD27 or
TIM-3. Additionally, the antitumor agent may be an epigen-
etic targeted drug such as HDAC inhibitors, kinase inhibi-
tors, DNA methyltransferase inhibitors, histone demethylase
inhibitors, or histone methylation inhibitors. The epigenetic
drugs may be Azacitidine (Vidaza), Decitabine (Dacogen),
Vorinostat (Zolinza), Romidepsin (Istodax), or Ruxolitinib
(Jakafi).

[0071] The term “chemotherapy” refers to the treatment of
cancer with an antitumor or chemotherapeutic agent as part
of a standardized regimen. Chemotherapy may be given
with a curative intent or it may aim to prolong life or to
palliate symptoms. It may be used in conjunction with other
cancer treatments, such as radiation therapy or surgery.

[0072] The term “clonal genetic mutation” refers an altera-
tion of genetic sequence of one or more cells to create a
clone (i.e., a progenitor cell) from which a population of
identical cells is derived. For example, a clonal genetic
mutation can be the genetic change that changes a healthy
cell into a cancerous cell in a subject, giving rise to a tumor
in the subject through clonal expansion. In such cases, the
clonal genetic mutation may change the nucleotide sequence
of an oncogene or a tumor suppressor gene.

[0073] The term “CpG” refers to a dinucleotide sequence,
wherein a cytosine nucleotide occurs next to a guanine
nucleotide in the linear sequence of bases along its length.
In a CpG sequence, the cytosine nucleotide is 5' to the
guanine nucleotide, and the two nucleotides are connected
by a phosphate molecule. Cytosines in CpG dinucleotides
can be methylated to form 5-methylcytosine. In mammals,
methylation of the cytosine within a gene or promoter can
affect transcriptional regulation of the gene. Enzymes that
add a methyl group are called DNA methyltransferases.

[0074] The term “CpG” island refers to a genomic region
that contains a high frequency of CpG sites. A CpG island
is characterized by CpG dinucleotide content of at least 60%
of that which would be statistically expected (about. 4-6%),
whereas the rest of the genome has a much lower CpG
frequency (about. 1%).

[0075] The term “epipolymorphism” refers to the proba-
bility that two randomly sampled DNA molecules differ in
their methylation pattern. Epipolymorphism can be deter-
mined by calculating the probability that two reads, selected
at random from a collection of overlapping reads will be not
be methylated identically. This probability will increase with
higher locally disordered methylation. See, e.g., Landan et
al., *Nat. Genetics*, 2012, vol. 44: 1207-1216, incorporated by
reference herein in its entirety.

[0076] The terms “information entropy” and “methylation
entropy” are used interchangeably herein and refer to a
measure of the randomness of DNA methylation patterns in

a cell population. For example, information entropy can be
calculated by computing Shannon’s entropy for the meth-
ylation state of neighboring CpGs. Entropies may be com-
bined in a variety of ways. See, e.g., Xie et al., (*Nucleic
Acids Research*, 2011, vol. 39, 4099-4108), incorporated by
reference herein in its entirety.

[0077] The term “likely to respond” to a therapy refers to
the plasticity of a tumor. A tumor with greater heterogeneity
is less likely to respond to an antitumor agent because there
is a greater possibility of a resistant subclone being present
or spontaneously arising within the tumor. Similarly, a tumor
that is likely to undergo subclonal evolution is less likely to
respond to an antitumor agent because the tumor may
develop resistance to the treatment.

[0078] The terms “locally disordered methylation” and
“discordant methylation” are used interchangeably and refer
alterations of CpG methylation patterns over a short genetic
distance or within a genomic feature. Typically, short-range
concordance is expected to be very high in non-disease
states, as DNA methylation generally changes by feature
(e.g., a specific gene promoter, or a CG island) rather than
by individual CpG. These terms may also refer, in a stricter
sense, to the concordance status of CpGs on the same
sequencing read. If all CpGs contained within one sequenc-
ing read are uniformly methylated or uniformly unmethyl-
ated, the read is classified as concordantly methylated.
Otherwise the read is classified as discordantly methylated.
A sequencing read may be over a short genetic distance,
such as 25 basepairs (bp); 30 bp; 35 bp; 40 bp; 45 bp; 50 bp;
60 bp; 70 bp; 80 bp; 90 bp; 100 bp; 250 bp; 500 bp; 750 bp
or 1000 bp, preferably less than 50 bp, or even less than 40
or less than 30 bp. FIG. 3 provides a graphical depiction of
discordant and concordant methylation patterns.

[0079] The term “methylation” refers to the addition of a
methyl group to the 5' carbon of the cytosine base in a
deoxyribonucleic acid sequence of CpG within a genome.

[0080] The term “methylation status” refers to the pres-
ence or absence of a methylated cytosine base at a CpG site.

[0081] The term “neighboring CpG site” refers to the
collection of CpG sites within a genomic feature or over a
short genetic distance. The genomic feature may be a
promoter, an enhancer, an exon, an intron, a 5'-untranslated
region (UTR), a 3'-UTR, a gene body, a stem cell associated
region, a CpG island, a CpG shelf, a CpG shore, a LINE, a
SINE, or an LTR. The short genetic distance may be 10 bp,
11 bp, 12 bp, 13 bp, 14 bp, 15 bp, 16 bp, 17 bp, 18 bp, 19
bp, 20 bp, 21 bp, 22 bp, 23 bp, 24 bp, 25 bp, 26 bp, 27 bp,
28 bp, 29 bp, 30 bp, 31 bp, 32 bp, 33 bp, 34 bp, 35 bp, 36
bp, 37 bp, 38 bp, 39 bp, 40 bp, 41 bp, 42 bp, 43 bp, 44 bp,
45 bp, 46 bp, 47 bp, 48 bp, 49 bp, 50 bp, 51 bp, 52 bp, 53
bp, 54 bp, 55 bp, 56 bp, 57 bp, 58 bp, 59 bp, 60 bp, 61 bp,
62 bp, 63 bp, 64 bp, 65 bp, 66 bp, 67 bp, 68 bp, 69 bp, 70
bp, 71 bp, 72 bp, 73 bp, 74 bp, 75 bp, 76 bp, 77 bp, 78 bp,
79 bp, 80 bp, 81 bp, 82 bp, 83 bp, 84 bp, 85 bp, 86 bp, 87
bp, 88 bp, 89 bp, 90 bp, 91 bp, 92 bp, 93 bp, 94 bp, 95 bp,
96 bp, 97 bp, 98 bp, 99 bp, 100 bp, 250 bp, 500 bp, 750 bp
or 1,000 bp, preferably 29 bp. Optionally, neighboring CpG
sites occur within a sequencing read.

[0082] The terms “proportion of discordant reads” and
“PDR” are used interchangeably and refer to the ratio of
discordant reads of the total number of overlapping reads for
a specific genomic location. FIG. 3 provides a sample
calculation of the proportion of discordant reads in a tumor
sample.

[0083] The term “sodium bisulfite” refers to sodium hydrogen sulfite having the chemical formula of NaHSO_3 . Sodium bisulfite functions to deaminate cytosine into uracil; but does not affect 5-methylcytosine (a methylated form of cytosine with a methyl group attached to carbon 5). When the bisulfite-treated DNA is amplified via polymerase chain reaction, the uracil is amplified as thymine and the methylated cytosine is amplified as cytosine.

[0084] The term “subclonal genetic mutation” refers to an alteration in a genetic sequence of one or more cells of a clonal population. Accordingly, a subclonal genetic mutation occurs subsequently to a clonal genetic mutation. Typically, individual cancer samples are genetically heterogeneous and contain subclonal populations. In cancer, subclonal genetic alterations may have an impact on clinical course. For example, a subclonal genetic mutation may arise in response to a selective pressure, such as treatment with an antitumor agent, to confer resistance. Similarly, in a heterogeneous tumor, a subclonal population, comprising a subclonal genetic mutation, may thrive and emerge as dominant, while other subclonal populations will decline, in response to selective pressure. Subclonal genetic mutations that permit a subclonal population to overcome a selective pressure are known as “subclonal driver mutations” or “subclonal drivers.”

[0085] The term “subject” refers to a vertebrate or invertebrate animal. In some embodiments, the subject is a vertebrate animal, e.g., a mammal, preferably a human. In some embodiments, a subject is a domestic or laboratory animal, including but not limited to, household pets, such as dogs, cats, pigs, rabbits, rats, mice, gerbils, hamsters, guinea pigs, and ferrets. In some embodiments, a subject is a livestock animal. Non-limiting examples of livestock animals include: alpaca, bison, camel, cattle, deer, pigs, horses, llamas, mules, donkeys, sheep, goats, rabbits, reindeer, and yak.

[0086] The term “variance” refers to a statistical measurement of how far a set of numbers is spread out. A variance of zero indicates that all the values are identical. A non-zero variance is always positive. A small variance indicates that the data points tend to be very close to the mean (expected value, e.g., concordance) and hence to each other, while a high variance indicates that the data points are very spread out from the mean and from each other. Variance may be calculated as a sum of variance that stems from discordant reads and the variance that stems from concordant reads. One can then estimate which contributes more to CpG variance.

[0087] As noted herein, genetically uniform cell subpopulations can contain profound epigenetic differences leading to phenotypic differences in their survival capacity and proliferative potential. The proportion of cells that ultimately participate in the evolutionary process may be limited by the fact that in order to form a new subclone, a novel somatic mutation would need to coincide with an epigenetic state permissive to its propagation. Applicants data shows that CLL cells have substantially increased epigenetic stochasticity, which results in a more malleable epigenetic landscape and likely increases the pool of cells that serve as substrate for the evolutionary process. These results demonstrate the mechanistic effects of epigenetic stochasticity on transcriptional regulation and chromatin modification. In addition, these results indicate the need to determine how genetic and epigenetic characteristics cooperate in CLL

clonal evolution in a large clinical trial cohort. Indeed, these results define locally disordered methylation as a key evolution-enabling feature of cancer and a predictive biomarker. Importantly, they may pave the way for the future development of therapeutic modalities to address the cancer’s evolutionary adaptive capacity.

[0088] This phenomenon is defined as locally disordered methylation. For example, the degree of methylation disorder as measured in DNA fragments that are sequenced in shotgun sequencing (up to 100 bases long) is higher in leukemia and cancer cell line samples than in normal samples. In certain embodiments, locally disordered methylation may be measured by identifying the overlapping reads (corresponding to DNA fragments originating from individual cells) for each genomic location covered by massive parallel sequencing; identifying the CpGs and their methylation status within each sequencing read (DNA fragment); if all CpGs contained within one sequencing read are uniformly methylated or uniformly unmethylated, the read may be classified as concordantly methylated, otherwise the read is classified as discordantly methylated.

Prognostic Methods

[0089] A first aspect of the present invention provides a method of assessing a subject’s cancer treatment prognosis. In some embodiments, the method comprises detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells in a tumor sample from the subject; comparing the DNA methylation status of neighboring CpG sites along a sequence of CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells; and assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells. The presence or prevalence of inconsistent methylation status along the sequences or across the multiple gene copies may indicate that the subject is more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof. For example, the DNA methylation status of CpG sites along one or more sequencing reads, e.g., operatively linked to each other in a single polynucleotide molecule, may be detected. The neighboring CpG sites along the sequencing read may then be compared to each other or to corresponding positions of different sequencing reads (e.g., at the same genomic location) from the plurality of cells.

[0090] DNA methylation may be detected by any method known in the art, including methylation-specific PCR, whole genome bisulfite sequence, the HELP assay and other methods using methylation-sensitive restriction endonucleases, ChIP-on-chip assays, restriction landmark genomic scanning, COBRA, Ms-SNuPE, methylated DNA immunoprecipitation (MeDip), pyrosequencing of bisulfite treated DNA, molecular break light assay for DNA adenine methyltransferase activity, methyl sensitive Southern blotting, methylCpG binding proteins, mass spectrometry, HPLC, and reduced representation bisulfite sequencing.

[0091] In some embodiments methylation is detected at specific sites of DNA methylation using pyrosequencing after bisulfite treatment and optionally after amplification of the methylation sites. Pyrosequencing technology is a method of sequencing-by-synthesis in real time. It is based

on an indirect bioluminometric assay of the pyrophosphate (PPi) that is released from each deoxynucleotide (dNTP) upon DNA-chain elongation. This method presents a DNA template-primer complex with a dNTP in the presence of an exonuclease-deficient Klenow DNA polymerase. The four nucleotides are sequentially added to the reaction mix in a predetermined order. If the nucleotide is complementary to the template base and thus incorporated, PPi is released. The PPi and other reagents are used as a substrate in a luciferase reaction producing visible light that is detected by either a luminometer or a charge-coupled device. The light produced is proportional to the number of nucleotides added to the DNA primer and results in a peak indicating the number and type of nucleotide present in the form of a pyrogram. Pyrosequencing can exploit the sequence differences that arise following sodium bisulfite-conversion of DNA.

[0092] In some embodiments, the DNA methylation is detected in a methylation assay utilizing next-generation sequencing. For example, DNA methylation may be detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing. Optionally, the DNA methylation is detected by microarray, such as a genome-wide microarray. Microarrays, and massively parallel sequencing, have enabled the interrogation of cytosine methylation on a genome-wide scale (Zilberman D, Henikoff S. 2007. Genome-wide analysis of DNA methylation patterns. *Development* 134(22): 3959-3965.). Genome wide methods have been described previously (Deng, et al. 2009. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol* 27(4): 353-360; Meissner, et al. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33(18): 5868-5877; Down, et al. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* 26(7): 779-785; Gu et al. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* 6(4): 468-481).

[0093] The most comprehensive, highest resolution method for detecting DNA methylation is whole genome bisulfite sequencing (WGBS) (Cokus, et al. 2008. Shotgun bisulfite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452(7184): 215-219; Lister, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271): 315-322; Harris, et al. 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* 28(10): 1097-1105).

[0094] To detect DNA methylation, a preferred embodiment provides for first converting the DNA to be analyzed so that the unmethylated cytosine is converted to uracil. In one embodiment, a chemical reagent that selectively modifies either the methylated or non-methylated form of CpG dinucleotide motifs may be used. Suitable chemical reagents include hydrazine and bisulfite ions and the like. Preferably, isolated DNA is treated with sodium bisulfite (NaHSO₃) which converts unmethylated cytosine to uracil, while methylated cytosines are maintained. Without wishing to be bound by a theory, it is understood that sodium bisulfite reacts readily with the 5,6-double bond of cytosine, but poorly with methylated cytosine. Cytosine reacts with the

bisulfite ion to form a sulfonated cytosine reaction intermediate that is susceptible to deamination, giving rise to a sulfonated uracil. The sulfonated group can be removed under alkaline conditions, resulting in the formation of uracil. The nucleotide conversion results in a change in the sequence of the original DNA. It is general knowledge that the resulting uracil has the base pairing behavior of thymine, which differs from cytosine base pairing behavior. To that end, uracil is recognized as a thymine by DNA polymerase. Therefore after PCR or sequencing, the resultant product contains cytosine only at the position where 5-methylcytosine occurs in the starting template DNA. This makes the discrimination between unmethylated and methylated cytosine possible.

[0095] The methylation status of neighboring CpG sites may be compared by calculating the proportion of discordant reads, calculating variance, or calculating information entropy identifying differentially methylated regions, by quantifying methylation difference, or by gene-set analysis (i.e., pathway analysis), preferably by calculating the proportion of discordant reads, calculating variance, or calculating information entropy. Optionally, information entropy is calculated by adapting Shannon entropy. In some embodiments, gene-set analysis is performed by tools such as DAVID, GoSeq or GSEA. In some embodiments, a proportion of discordant reads (PDR) is calculated. Optionally, each region of neighboring CpG sites (e.g., within a sequencing read) is assigned a consistent status or an inconsistent status before calculating the proportion of discordant reads, variance, epipolymorphism or information entropy. There may be multiple inconsistent statuses, each representing a distinct methylation pattern or class of similar methylation patterns.

[0096] The one or more regions of neighboring CpG sites may be a short genetic sequence or a genomic feature. The short genetic sequence may consist of 10 bp, 11 bp, 12 bp, 13 bp, 14 bp, 15 bp, 16 bp, 17 bp, 18 bp, 19 bp, 20 bp, 21 bp, 22 bp, 23 bp, 24 bp, 25 bp, 26 bp, 27 bp, 28 bp, 29 bp, 30 bp, 31 bp, 32 bp, 33 bp, 34 bp, 35 bp, 36 bp, 37 bp, 38 bp, 39 bp, 40 bp, 41 bp, 42 bp, 43 bp, 44 bp, 45 bp, 46 bp, 47 bp, 48 bp, 49 bp, 50 bp, 51 bp, 52 bp, 53 bp, 54 bp, 55 bp, 56 bp, 57 bp, 58 bp, 59 bp, 60 bp, 61 bp, 62 bp, 63 bp, 64 bp, 65 bp, 66 bp, 67 bp, 68 bp, 69 bp, 70 bp, 71 bp, 72 bp, 73 bp, 74 bp, 75 bp, 76 bp, 77 bp, 78 bp, 79 bp, 80 bp, 81 bp, 82 bp, 83 bp, 84 bp, 85 bp, 86 bp, 87 bp, 88 bp, 89 bp, 90 bp, 91 bp, 92 bp, 93 bp, 94 bp, 95 bp, 96 bp, 97 bp, 98 bp, 99 bp, 100 bp, 250 bp, 500 bp, or 1,000 bp, preferably 29 bp.

[0097] Preferably, an optimal amplicon length when one or more regions of neighboring CpG sites is amplified by PCR is between about 80 base pairs and about 150 base pairs. There is an inverse relationship between the amplicon length and PCR efficiency. The underlying rationale is related to the fact that sodium bisulfite treatment causes degradation of DNA and therefore PCR efficiency decreases as amplicon size gets larger.

[0098] Optionally, the region of neighboring CpG sites is a genomic feature selected from a CpG island, a CpG shore, a CpG shelf, a promoter, an enhancer, an exon, an intron, a gene body, a stem cell associated region, a short interspersed element (SINE), a long interspersed element (LINE), and a long terminal repeat (LTR), preferably a CpG island or a promoter.

[0099] The one or more regions of neighboring CpG sites having a locally disordered methylation status may be located within a genomic location selected from a CpG island, a CpG shore, a CpG shelf, a promoter, an enhancer, an exon, an intron, a gene body, a stem cell associated region, a short interspersed element (SINE), a long interspersed element (LINE), and a long terminal repeat (LTR), preferably a CpG island or a promoter.

[0100] In some embodiments, the method further comprises detecting a subclonal genetic mutation. Optionally, the subclonal genetic mutation is within the one or more genomic regions having a locally disordered methylation status. A subclonal genetic mutation with the one or more genomic regions having a locally disordered methylation status may indicate that the subject is even more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof. Optionally, the presence of a subclonal mutation in a subject with a prevalence of inconsistent methylation status along the sequences or across the multiple gene copies may indicate that the subject is even more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof.

[0101] The subclonal genetic mutation may be detected by any method known in the art. For example, the subclonal genetic mutation may be detected by Comparative Genomic Hybridization Array, Multiple Ligation-dependent Probe Amplification, Multiplex Amplifiable Probe Hybridization, Single Condition Amplification/Internal Primer, Multiplex PCR, Southern Blot, Sanger gene sequence, Resequencing Array, mRNA analysis, cDNA sequencing, microarray analysis, whole-genome sequence, massively parallel signature sequencing, Polony sequencing, 454 pyrosequencing, SOLiD sequencing, Illumina dye sequencing, ion semiconductor sequence, DNA nanoball sequencing, heliscope single molecule sequencing, single molecule real time sequencing, nanopore DNA sequencing, tunneling currents DNA sequencing, sequencing by hybridization, sequencing with mass spectrometry, transmission electron microscopy sequencing, and RNA polymerase sequencing, preferably massively parallel signature sequencing.

[0102] The DNA methylation status at the one or more neighboring CpG sites may be detected before treatment with an antitumor agent. Optionally, the DNA methylation status at the one or more neighboring CpG sites is detected after treatment with an antitumor agent. The DNA methylation status at the one or more neighboring CpG sites may be detected both before and after treatment with an antitumor agent. In some embodiments, the DNA methylation status at the one or more neighboring CpG sites is detected throughout a time course of treatment with an antitumor agent. An increase in the number of genomic regions having a locally disordered methylation status may indicate that the subject is less likely to respond to an antitumor agent.

[0103] The tumor sample may be a solid tumor, such as carcinomas, sarcomas and lymphomas. In some embodiments, the solid tumor is selected from adrenocortical carcinoma, bone tumors, brain cancer, breast cancer, cervical cancer, colorectal carcinoma, desmoid tumors, desmoplastic small round cell tumors, endocrine tumors, esophageal cancer, Ewing sarcoma family tumors, gastric cancer, germ cell tumors, head or neck cancer, hepatoblastoma, hepatocellular

carcinoma, lung cancer, melanoma, mesothelioma, nasopharyngeal carcinoma, neuroblastoma, non-rhabdomyosarcoma soft tissue sarcoma, osteosarcoma, ovarian cancer, pancreatic cancer, prostate cancer, retinoblastoma, rhabdomyosarcoma, skin carcinoma, testicular cancer, thyroid carcinoma, uterine cancer and Wilms tumors. The tumor sample may be a hematological cancer, such as leukemia, preferably CLL.

[0104] The antitumor agent is selected from an angiogenesis inhibitor, such as angiostatin K1-3, DL- α -Difluoromethyl-ornithine, endostatin, fumagillin, genistein, minocycline, staurosporine, and ($\pm$)-thalidomide; a DNA intercalator/cross-linker, such as Bleomycin, Carboplatin, Carmustine, Chlorambucil, Cyclophosphamide, cis-Diammineplatinum (II) dichloride (Cisplatin), Melphalan, Mitoxantrone, and Oxaliplatin; a DNA synthesis inhibitor, such as ($\pm$)-Ame-thopterin (Methotrexate), 3-Amino-1,2,4-benzotriazine 1,4-dioxide, Aminopterin, Cytosine β -D-arabinofuranoside, 5-Fluoro-5'-deoxyuridine, 5-Fluorouracil, Ganciclovir, Hydroxyurea, and Mitomycin C; a DNA-RNA transcription regulator, such as Actinomycin D, Daunorubicin, Doxorubicin, Homoharringtonine, and Idarubicin; an enzyme inhibitor, such as S(+)-Camptothecin, Curcumin, (-)-Deguelin, 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside, Etoposide, Formestane, Fostriecin, Hispidin, 2-Imino-1-imidazoli-dineacetic acid (Cyclocreatine), Mevinolin, Trichostatin A, Tyrphostin AG 34, and Tyrphostin AG 879; a gene regulator, such as 5-Aza-2'-deoxycytidine, 5-Azacytidine, Cholecalciferol (Vitamin D3), 4-Hydroxytamoxifen, Melatonin, Mifepristone, Raloxifene, all trans-Retinal (Vitamin A aldehyde), Retinoic acid, all trans (Vitamin A acid), 9-cis-Retinoic Acid, 13-cis-Retinoic acid, Retinol (Vitamin A), Tamoxifen, and Troglitazone; a microtubule inhibitor, such as Colchicine, Dolastatin 15, Nocodazole, Paclitaxel, Podophyllotoxin, Rhizoxin, Vinblastine, Vincristine, Vindesine, and Vinorelbine (Navelbine); and an unclassified antitumor agent, such as 17-(Allylamino)-17-demethoxygeldanamycin, 4-Amino-1,8-naphthalimide, Apigenin, Brefeldin A, Cimetidine, Dichloromethylene-diphosphonic acid, Leupro-lide (Leuprorelin), Luteinizing Hormone-Releasing Hormone, Pifithrin- α , Rapamycin, Sex hormone-binding globulin, Thapsigargin, and Urinary trypsin inhibitor fragment (Bikunin). The antitumor agent may be a monoclonal antibody such as rituximab (Rituxan®), alemtuzumab (Campath®), Ipilimumab (Yervoy®), Bevacizumab (Avastin®), Cetuximab (Erbix®), panitumumab (Vectibix®), and trastuzumab (Herceptin®), Vemurafenib (Zelboraf®) imatinib mesylate (Gleevec®), erlotinib (Tarceva®), gefitinib (Iressa®), Vismodegib (Erivedge™), 90Y-ibritumomab tiuxetan, 131I-tositumomab, ado-trastuzumab emtansine, lapatinib (Tykerb®), pertuzumab (Perjeta™), ado-trastuzumab emtansine (Kadcyla™), regorafenib (Stivarga®), sunitinib (Sutent®), Denosumab (Xgeva®), sorafenib (Nexavar®), pazopanib (Votrient®), axitinib (Inlyta®), dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (Bosulif®), ofatumumab (Arzerra®), obinituzumab (Gazyva™), ibrutinib (Imbruvica™), idelalisib (Zydelig®), crizotinib (Xalkori®), erlotinib (Tarceva®), afatinib dimaleate (Gilotrif®), ceritinib (LDK378/Zykadia), Tositumomab and 131I-tositumomab (Bexxar®), ibritumomab tiuxetan (Zevalin®), brentuximab vedotin (Adcetris®), bortezomib (Velcade®), siltuximab (Sylvant™), trametinib (Mekinist®), dabrafenib (Tafinlar®), pembrolizumab (Keytruda®), carfilzomib (Kyprolis®), Ramucirumab (Cyramza™), Cabozantinib (Cometriq™), vandetanib (Caprelsa®), Optionally, the

antitumor agent is a neoantigen. The antitumor agent may be a cytokine such as interferons (INFs), interleukins (ILs), or hematopoietic growth factors. The antitumor agent may be INF- α , IL-2, Aldesleukin, IL-2, Erythropoietin, Granulocyte-macrophage colony-stimulating factor (GM-CSF) or granulocyte colony-stimulating factor. The antitumor agent may be a targeted therapy such as toremifene (Fareston®), fulvestrant (Faslodex®), anastrozole (Arimidex®), exemestane (Aromasin®), letrozole (Femara®), ziv-aflibercept (Zaltrap®), Alitretinoin (Panretin®), temsirolimus (Torisel®), Tretinoin (Vesanoid®), denileukin diftitox (Ontak®), vorinostat (Zolinza®), romidepsin (Istodax®), bexarotene (Targretin®), pralatrexate (Folotyn®), lenalidomide (Revlimid®), belinostat (Beleodaq™), lenalidomide (Revlimid®), pomalidomide (Pomalyst®), Cabazitaxel (Jevtana®), enzalutamide (Xtandi®), abiraterone acetate (Zytiga®), radium 223 chloride (Xofigo®), or everolimus (Afinitor®). The antitumor agent may be a checkpoint inhibitor such as an inhibitor of the programmed death-1 (PD-1) pathway, for example an anti-PD1 antibody (Nivolumab). The inhibitor may be an anti-cytotoxic T-lymphocyte-associated antigen (CTLA-4) antibody. The inhibitor may target another member of the CD28 CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. A checkpoint inhibitor may target a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3. Additionally, the antitumor agent may be an epigenetic targeted drug such as HDAC inhibitors, kinase inhibitors, DNA methyltransferase inhibitors, histone demethylase inhibitors, or histone methylation inhibitors. The epigenetic drugs may be Azacitidine (Vidaza), Decitabine (Dacogen), Vorinostat (Zolinza), Romidepsin (Istodax), or Ruxolitinib (Jakafi).

[0105] The subject may be a mammal, preferably a human. In some embodiments, the subject may be a laboratory animal, such as a mouse, a rabbit, a rat, a guinea pig, a hamster, and a primate.

Methods of Calculating a Proportion of Discordant Reads

[0106] A second aspect of the present invention provides a method of calculating a proportion of discordant reads (PDR) in a first tumor sample from a subject. In some embodiments, the method comprises detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells in the tumor sample; comparing the DNA methylation status of sequencing reads in one or more regions of neighboring CpG sites along a sequence of CpG sites in DNA of the plurality of cells; assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells; and determining a relative number of cells in the tumor sample having variable methylation status across the sequence of CpG sites as compared to the total number of cells in the tumor sample or a number of cells in the tumor sample having consistent methylation status across the sequence of CpG sites.

[0107] The DNA methylation may be detected methylation-specific PCR, whole genome bisulfite sequence, the HELP assay and other methods using methylation-sensitive restriction endonucleases, ChIP-on-chip assays, restriction landmark genomic scanning, COBRA, Ms-SNuPE, methylated DNA immunoprecipitation (MeDip), pyrosequencing of bisulfite treated DNA, molecular break light assay for DNA adenine methyltransferase activity, methyl sensitive Southern blotting, methylCpG binding proteins, mass spec-

trometry, HPLC, and reduced representation bisulfite sequencing. In some embodiments, the DNA methylation is detected in a methylation assay utilizing next-generation sequencing. For example, DNA methylation may be detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing. Optionally, the DNA methylation is detected by microarray, such as a genome-wide microarray.

[0108] The one or more neighboring CpG sites may be a short genetic sequence or a genomic feature. The short genetic sequence may consist of 10 bp, 11 bp, 12 bp, 13 bp, 14 bp, 15 bp, 16 bp, 17 bp, 18 bp, 19 bp, 20 bp, 21 bp, 22 bp, 23 bp, 24 bp, 25 bp, 26 bp, 27 bp, 28 bp, 29 bp, 30 bp, 31 bp, 32 bp, 33 bp, 34 bp, 35 bp, 36 bp, 37 bp, 38 bp, 39 bp, 40 bp, 41 bp, 42 bp, 43 bp, 44 bp, 45 bp, 46 bp, 47 bp, 48 bp, 49 bp, 50 bp, 51 bp, 52 bp, 53 bp, 54 bp, 55 bp, 56 bp, 57 bp, 58 bp, 59 bp, 60 bp, 61 bp, 62 bp, 63 bp, 64 bp, 65 bp, 66 bp, 67 bp, 68 bp, 69 bp, 70 bp, 71 bp, 72 bp, 73 bp, 74 bp, 75 bp, 76 bp, 77 bp, 78 bp, 79 bp, 80 bp, 81 bp, 82 bp, 83 bp, 84 bp, 85 bp, 86 bp, 87 bp, 88 bp, 89 bp, 90 bp, 91 bp, 92 bp, 93 bp, 94 bp, 95 bp, 96 bp, 97 bp, 98 bp, 99 bp, 100 bp, 250 bp, 500 bp, or 1,000 bp, preferably 29 bp. Optionally, the neighboring CpG sites are located within a genomic feature selected from a CpG island, a CpG shore, a CpG shelf, a promoter, an enhancer, an exon, an intron, a gene body, a stem cell associated region, a short interspersed element (SINE), a long interspersed element (LINE), and a long terminal repeat (LTR), preferably a CpG island or a promoter.

[0109] The PDR may be calculated for a genomic location or for a genomic feature, such as a CpG island, a CpG shore, a CpG shelf, a promoter, an enhancer, an exon, an intron, a gene body, a stem cell associated region, a short interspersed element (SINE), a long interspersed element (LINE), and a long terminal repeat (LTR), preferably a CpG island or a promoter.

[0110] The PDR may be calculated before treatment with an antitumor agent. Optionally, the PDR is calculated after treatment with an antitumor agent. The PDR may be calculated both before and after treatment with an antitumor agent. In some embodiments, the PDR is calculated throughout a time course of treatment with an antitumor agent.

[0111] In some embodiments, a PDR threshold, such as greater than 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.38, 0.37, 0.38, 0.39, or 0.40, preferably 0.15, indicates that the patient is less likely to respond to treatment with an antitumor agent. A PDR threshold, such as less than 0.10, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.38, 0.37, 0.38, 0.39, or 0.40, preferably 0.15, may indicate that the patient is more likely to respond to treatment with an antitumor agent. A change in PDR following treatment may indicate that the subject is likely to relapse despite treatment. Without being bound by theory, an increase in PDR may suggest clonal evolution; a decrease in PDR may signal selection of a dominant subclone; and a constant PDR may suggest that an antitumor agent is equally effective across subclones.

[0112] The tumor sample may be a solid tumor, such as carcinomas, sarcomas and lymphomas. In some embodiments, the solid tumor is selected from adrenocortical carcinoma, bone tumors, brain cancer, breast cancer, cervical

cancer, colorectal carcinoma, desmoid tumors, desmoplastic small round cell tumors, endocrine tumors, esophageal cancer, Ewing sarcoma family tumors, gastric cancer, germ cell tumors, head or neck cancer, hepatoblastoma, hepatocellular carcinoma, lung cancer, melanoma, mesothelioma, nasopharyngeal carcinoma, neuroblastoma, non-rhabdomyosarcoma soft tissue sarcoma, osteosarcoma, ovarian cancer, pancreatic cancer, prostate cancer, retinoblastoma, rhabdomyosarcoma, skin carcinoma, testicular cancer, thyroid carcinoma, uterine cancer and Wilms tumors. The tumor sample may be a hematological cancer, such as leukemia, preferably CLL.

[0113] The antitumor agent is selected from an angiogenesis inhibitor, such as angiostatin K1-3, DL- α -difluoro-methyl-ornithine, endostatin, fumagillin, genistein, minocycline, staurosporine, and ($\pm$)-thalidomide; a DNA intercalator/cross-linker, such as Bleomycin, Carboplatin, Carmustine, Chlorambucil, Cyclophosphamide, cis-Diammineplatinum (II) dichloride (Cisplatin), Melphalan, Mitoxantrone, and Oxaliplatin; a DNA synthesis inhibitor, such as ($\pm$)-Ame-thopterin (Methotrexate), 3-Amino-1,2,4-benzotriazine 1,4-dioxide, Aminopterin, Cytosine 3-D-arabinofuranoside, 5-Fluoro-5'-deoxyuridine, 5-Fluorouracil, Ganciclovir, Hydroxyurea, and Mitomycin C; a DNA-RNA transcription regulator, such as Actinomycin D, Daunorubicin, Doxorubicin, Homoharringtonine, and Idarubicin; an enzyme inhibitor, such as S(+)-Camptothecin, Curcumin, (-)-Deguelin, 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside, Etoposide, Formestane, Fostriecin, Hispidin, 2-Imino-1-imidazoli-dineacetic acid (Cyclocreatine), Mevinolin, Trichostatin A, Tyrphostin AG 34, and Tyrphostin AG 879; a gene regulator, such as 5-Aza-2'-deoxycytidine, 5-Azacytidine, Cholecalciferol (Vitamin D3), 4-Hydroxytamoxifen, Melatonin, Mifepristone, Raloxifene, all trans-Retinal (Vitamin A aldehyde), Retinoic acid, all trans (Vitamin A acid), 9-cis-Retinoic Acid, 13-cis-Retinoic acid, Retinol (Vitamin A), Tamoxifen, and Troglitazone; a microtubule inhibitor, such as Colchicine, Dolastatin 15, Nocodazole, Paclitaxel, Podophyllotoxin, Rhizoxin, Vinblastine, Vincristine, Vindesine, and Vinorelbine (Navelbine); and an unclassified antitumor agent, such as 17-(Allylamino)-17-demethoxygeldanamycin, 4-Amino-1,8-naphthalimide, Apigenin, Brefeldin A, Cimetidine, Dichloromethylene-diphosphonic acid, Leuprolide (Leuporelin), Luteinizing Hormone-Releasing Hormone, Pifithrin- α , Rapamycin, Sex hormone-binding globulin, Thapsigargin, and Urinary trypsin inhibitor fragment (Bikunin). The antitumor agent may be a monoclonal antibody such as rituximab (Rituxan®), alemtuzumab (Campath®), Ipilimumab (Yervoy®), Bevacizumab (Avastin®), Cetuximab (Erbix®), panitumumab (Vectibix®), and trastuzumab (Herceptin®), Vemurafenib (Zelboraf®) imatinib mesylate (Gleevec®), erlotinib (Tarceva®), gefitinib (Iressa®), Vismodegib (Erivedge™), 90Y-ibritumomab tiuxetan, 131I-tositumomab, ado-trastuzumab emtansine, lapatinib (Tykerb®), pertuzumab (Perjeta™), ado-trastuzumab emtansine (Kadcyla™), regorafenib (Stivarga®), sunitinib (Sutent®), Denosumab (Xgeva®), sorafenib (Nexavar®), pazopanib (Votrient®), axitinib (Inlyta®), dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (Bosulif®), ofatumumab (Arzerra®), obinutuzumab (Gazyva™), ibrutinib (Imbruvica™), idelalisib (Zydelig®), crizotinib (Xalkori®), erlotinib (Tarceva®), afatinib dimaleate (Gilotrif®), ceritinib (LDK378/Zykadia), Tositumomab and 131I-tositumomab (Bexxar®), ibritumomab tiuxetan (Zevalin®), brentuximab vedotin (Adcetris®), bortezomib (Vel-

cade®), siltuximab (Sylvant™), trametinib (Mekinist®), dabrafenib (Tafinlar®), pembrolizumab (Keytruda®), carfilzomib (Kyprolis®), Ramucirumab (Cyramza™), Cabozantinib (Cometriq™), vandetanib (Caprelsa®), Optionally, the antitumor agent is a neoantigen. The antitumor agent may be a cytokine such as interferons (INFs), interleukins (ILs), or hematopoietic growth factors. The antitumor agent may be INF- α , IL-2, Aldesleukin, IL-2, Erythropoietin, Granulocyte-macrophage colony-stimulating factor (GM-CSF) or granulocyte colony-stimulating factor. The antitumor agent may be a targeted therapy such as toremifene (Fareston®), fulvestrant (Faslodex®), anastrozole (Arimidex®), exemestane (Aromasin®), letrozole (Femara®), ziv-aflibercept (Zaltrap®), Alitretinoin (Panretin®), temsirolimus (Torisel®), Tretinoin (Vesanoid®), denileukin difitox (Ontak®), vorinostat (Zolinza®), romidepsin (Istodax®), bexarotene (Targretin®), pralatrexate (Foloty®), lenalidomide (Revlimid®), belinostat (Beleodaq™), lenalidomide (Revlimid®), pomalidomide (Pomalyst®), Cabazitaxel (Jevtana®), enzalutamide (Xtandi®), abiraterone acetate (Zytiga®), radium 223 chloride (Xofigo®), or everolimus (Afinitor®). The antitumor agent may be a checkpoint inhibitor such as an inhibitor of the programmed death-1 (PD-1) pathway, for example an anti-PD1 antibody (Nivolumab). The inhibitor may be an anti-cytotoxic T-lymphocyte-associated antigen (CTLA-4) antibody. The inhibitor may target another member of the CD28 CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. A checkpoint inhibitor may target a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3. Additionally, the antitumor agent may be an epigenetic targeted drug such as HDAC inhibitors, kinase inhibitors, DNA methyltransferase inhibitors, histone demethylase inhibitors, or histone methylation inhibitors. The epigenetic drugs may be Azacitidine (Vidaza), Decitabine (Dacogen), Vorinostat (Zolinza), Romidepsin (Istodax), or Ruxolitinib (Jakafi).

[0114] The subject may be a mammal, preferably a human. In some embodiments, the subject may be a laboratory animal, such as a mouse, a rabbit, a rat, a guinea pig, a hamster, and a primate.

Methods of Treatment

[0115] A third aspect of the present invention provides a method of treating a subject suffering from cancer. In some embodiments, the method comprises performing a prognostic method as described herein and administering an antitumor agent to the subject if no or few regions of neighboring CpG sites having a locally disordered methylation status is identified. Optionally, if the presence or prevalence of regions of neighboring CpG sites having a locally disordered methylation status is identified, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation).

[0116] In some embodiments, the method comprises performing a first prognostic method as described herein; administering an antitumor agent to the subject; and performing a second prognostic method as described herein, wherein the treatment is administered between the first and second prognostic methods. In some embodiments, the method comprises continuing to treat the subject with the antitumor agent if the number of regions of neighboring CpG sites having locally disordered methylation is substantially the same (or, in some instances, is lower than) in the

second prognostic method compared to the first prognostic method. In some embodiments, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the number of regions of neighboring CpG sites having locally disordered methylation is different (e.g., greater than, or in some instances less than) in the second prognostic method compared to the first prognostic method. The antitumor agent may be administered to the subject for at least 3 months, at least 6 months, at least 9 months, at least 12 months, at least 24 months, or at least 36 months before performance of the second prognostic method, preferably at least 12 months.

[0117] The methylation status of neighboring CpG sites may be compared by calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism, or calculating information entropy. In some embodiments, a proportion of discordant reads (PDR) is calculated.

[0118] The method of treatment may further comprise detecting a genetic mutation. The genetic mutation may be a clonal mutation or a subclonal mutation, preferably a subclonal mutation. In some embodiments, the method comprises treating or continuing to treat the subject with the antitumor agent if no region of neighboring CpG sites having locally disordered methylation status also comprises a genetic mutation or is proximal to a genetic mutation, such as a subclonal mutation. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if one or more regions of neighboring CpG sites having locally disordered methylation status also comprises a genetic mutation or is proximal to a genetic mutation, such as a subclonal mutation.

[0119] The subclonal genetic mutation may be detected by any method known in the art. For example, the subclonal genetic mutation may be detected by Comparative Genomic Hybridization Array, Multiple Ligation-dependent Probe Amplification, Multiplex Amplifiable Probe Hybridization, Single Condition Amplification/Internal Primer, Multiplex PCR, Southern Blot, Sanger gene sequence, Resequencing Array, mRNA analysis, cDNA sequencing, microarray analysis, whole-genome sequence, massively parallel signature sequencing, Polony sequencing, 454 pyrosequencing, SOLiD sequencing, Illumina dye sequencing, ion semiconductor sequence, DNA nanoball sequencing, heliscope single molecule sequencing, single molecule real time sequencing, nanopore DNA sequencing, tunneling currents DNA sequencing, sequencing by hybridization, sequencing with mass spectrometry, transmission electron microscopy sequencing, and RNA polymerase sequencing, preferably massively parallel signature sequencing.

[0120] In some embodiments, the method comprises determining the genomic location of the one or more regions of neighboring CpG sites having locally disordered methylation status. The genomic location may be selected from a CpG island, a CpG shore, a CpG shelf, a promoter, an enhancer, an exon, an intron, a gene body, a stem cell associated region, a short interspersed element (SINE), a long interspersed element (LINE), and a long terminal repeat (LTR), preferably a CpG island or a promoter. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if one or more regions of neighboring CpG sites having locally disordered methyl-

ation status is located in a CpG island, a promoter or an exon. The method may comprise treating or continuing to treat the subject with the antitumor agent if no region of neighboring CpG sites having locally disordered methylation status is within a CpG island, a promoter, or an exon.

[0121] In some embodiments, the method of treatment comprises calculating a PDR from a tumor sample in a subject as described herein, and administering an antitumor agent to the subject if the PDR is less than a PDR threshold, such as 0.10, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.38, 0.37, 0.38, 0.39, or 0.40, preferably 0.15. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the PDR is greater than a PDR threshold, such as 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.38, 0.37, 0.38, 0.39, or 0.40, preferably 0.15.

[0122] In some embodiments, the method of treatment comprises calculating a first PDR, variance, epipolymorphism or information entropy from a first tumor sample obtained from the subject according to the methods described herein; treating the subject with an antitumor agent; and calculating a second PDR, variance, epipolymorphism or information entropy from a second tumor sample obtained from the subject according to the methods described herein, wherein the antitumor agent is administered between obtaining the first and second tumor samples. The method may comprise continuing to treat the subject with the antitumor agent if the second PDR, variance, epipolymorphism or information entropy has not changed compared to the first PDR. Optionally, the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the second PDR, variance, epipolymorphism or information entropy has changed compared to the first PDR. For example, an increase in PDR may suggest clonal evolution. A decrease in PDR, however, may signal selection of a dominant subclone. A constant PDR may suggest that an antitumor agent is equally effective across subclones. The first and second tumor sample may be from the same tumor.

[0123] The subject may be treated with the antitumor agent for at least 3 months, at least 6 months, at least 9 months, at least 12 months, at least 24 months or at least 36 months before calculating the second PDR, variance, epipolymorphism or information entropy, preferably at least 12 months.

[0124] The tumor sample may be a solid tumor, such as carcinomas, sarcomas and lymphomas. In some embodiments, the solid tumor is selected from adrenocortical carcinoma, bone tumors, brain cancer, breast cancer, cervical cancer, colorectal carcinoma, desmoid tumors, desmoplastic small round cell tumors, endocrine tumors, esophageal cancer, Ewing sarcoma family tumors, gastric cancer, germ cell tumors, head or neck cancer, hepatoblastoma, hepatocellular carcinoma, lung cancer, melanoma, mesothelioma, nasopharyngeal carcinoma, neuroblastoma, non-rhabdomyosarcoma soft tissue sarcoma, osteosarcoma, ovarian cancer, pancreatic cancer, prostate cancer, retinoblastoma, rhabdomyosarcoma, skin carcinoma, testicular cancer, thyroid carcinoma, uterine cancer and Wilms tumors. The tumor sample may be a hematological cancer, such as leukemia, preferably CLL.

[0125] The antitumor agent is selected from an angiogenesis inhibitor, such as angiostatin K1-3, DL- α -difluoromethyl-ornithine, endostatin, fumagillin, genistein, minocycline, staurosporine, and ($\pm$)-thalidomide; a DNA intercalator/cross-linker, such as Actinomycin D, Carboplatin, Carmustine, Chlorambucil, Cyclophosphamide, cis-Diammineplatinum (II) dichloride (Cisplatin), Melfalan, Mitoxantrone, and Oxaliplatin; a DNA synthesis inhibitor, such as ($\pm$)-Ame-thopterin (Methotrexate), 3-Amino-1,2,4-benzotriazine 1,4-dioxide, Aminopterin, Cytosine β -D-arabinofuranoside, 5-Fluoro-5'-deoxyuridine, 5-Fluorouracil, Ganciclovir, Hydroxyurea, and Mitomycin C; a DNA-RNA transcription regulator, such as Actinomycin D, Daunorubicin, Doxorubicin, Homoharringtonine, and Idarubicin; an enzyme inhibitor, such as S(+)-Camptothecin, Curcumin, (-)-Deguelin, 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside, Etoposide, Formestane, Fostriecin, Hispidin, 2-Imino-1-imidazoli-dineacetic acid (Cyclocreatine), Mevinolin, Trichostatin A, Tyrophostin AG 34, and Tyrophostin AG 879; a gene regulator, such as 5-Aza-2'-deoxycytidine, 5-Azacytidine, Cholecalciferol (Vitamin D3), 4-Hydroxytamoxifen, Melatonin, Mifepristone, Raloxifene, all trans-Retinal (Vitamin A aldehyde), Retinoic acid, all trans (Vitamin A acid), 9-cis-Retinoic Acid, 13-cis-Retinoic acid, Retinol (Vitamin A), Tamoxifen, and Troglitazone; a microtubule inhibitor, such as Colchicine, Dolastatin 15, Nocodazole, Paclitaxel, Podophyllotoxin, Rhizoxin, Vinblastine, Vincristine, Vindesine, and Vinorelbine (Navelbine); and an unclassified antitumor agent, such as 17-(Allylamino)-17-demethoxygeldanamycin, 4-Amino-1,8-naphthalimide, Apigenin, Brefeldin A, Cimetidine, Dichloromethylene-diphosphonic acid, Leuprolide (Leuporelin), Luteinizing Hormone-Releasing Hormone, Pifithrin- α , Rapamycin, Sex hormone-binding globulin, Thapsigargin, and Urinary trypsin inhibitor fragment (Bikunin). The antitumor agent may be a monoclonal antibody such as rituximab (Rituxan®), alemtuzumab (Campath®), Ipilimumab (Yervoy®), Bevacizumab (Avastin®), Cetuximab (Erbix®), panitumumab (Vectibix®), and trastuzumab (Herceptin®), Vemurafenib (Zelboraf®) imatinib mesylate (Gleevec®), erlotinib (Tarceva®), gefitinib (Iressa®), Vismodegib (Erivedge™), 90Y-ibritumomab tiuxetan, 131I-tositumomab, ado-trastuzumab emtansine, lapatinib (Tykerb®), pertuzumab (Perjeta™), ado-trastuzumab emtansine (Kadcyla™), regorafenib (Stivarga®), sunitinib (Sutent®), Denosumab (Xgeva®), sorafenib (Nexavar®), pazopanib (Votrient®), axitinib (Inlyta®), dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (Bosulif®), ofatumumab (Arzerra®), obinutuzumab (Gazyva™), ibritinib (Imbruvica™), idelalisib (Zydelig®), crizotinib (Xalkori®), erlotinib (Tarceva®), afatinib dimaleate (Gilotrif®), ceritinib (LDK378/Zykadia), Tositumomab and 131I-tositumomab (Bexxar®), ibritumomab tiuxetan (Zevalin®), brentuximab vedotin (Adcetris®), bortezomib (Velcade®), siltuximab (Sylvant™), trametinib (Mekinist™), dabrafenib (Tafinlar®), pembrolizumab (Keytruda®), carfilzomib (Kyprolis®), Ramucirumab (Cyramza™), Cabozantinib (Cometriq™), vandetanib (Caprelsa®). Optionally, the antitumor agent is a neoantigen. The antitumor agent may be a cytokine such as interferons (INFs), interleukins (ILs), or hematopoietic growth factors. The antitumor agent may be INF- α , IL-2, Aldesleukin, IL-2, Erythropoietin, Granulocyte-macrophage colony-stimulating factor (GM-CSF) or granulocyte colony-stimulating factor. The antitumor agent may be a targeted therapy such as toremifene (Fareston®),

fulvestrant (Faslodex®), anastrozole (Arimidex®), exemestane (Aromasin®), letrozole (Femara®), ziv-aflibercept (Zaltrap®), Alitretinoin (Panretin®), temsirolimus (Torisel®), Tretinoin (Vesanoid®), denileukin diftitox (Ontak®), vorinostat (Zolinza®), romidepsin (Istodax®), bexarotene (Targretin®), pralatrexate (Foloty®), lenalidomide (Revlimid®), belinostat (Beleodaq™), lenalidomide (Revlimid®), pomalidomide (Pomalyst®), Cabazitaxel (Jevtana®), enzalutamide (Xtandi®), abiraterone acetate (Zytiga®), radium 223 chloride (Xofigo®), or everolimus (Afinitor®). The antitumor agent may be a checkpoint inhibitor such as an inhibitor of the programmed death-1 (PD-1) pathway, for example an anti-PD1 antibody (Nivolumab). The inhibitor may be an anti-cytotoxic T-lymphocyte-associated antigen (CTLA-4) antibody. The inhibitor may target another member of the CD28 CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. A checkpoint inhibitor may target a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3. Additionally, the antitumor agent may be an epigenetic targeted drug such as HDAC inhibitors, kinase inhibitors, DNA methyltransferase inhibitors, histone demethylase inhibitors, or histone methylation inhibitors. The epigenetic drugs may be Azacitidine (Vidaza), Decitabine (Dacogen), Vorinostat (Zolinza), Romidepsin (Istodax), or Ruxolitinib (Jakafi).

[0126] The subject may be a mammal, preferably a human. In some embodiments, the subject may be a laboratory animal, such as a mouse, a rabbit, a rat, a guinea pig, a hamster, and a primate.

[0127] In another embodiment treatment is consistent with the standard of care for a patient in need thereof. In one embodiment the prognostic methods are used to determine the proper standard of care for a patient in need thereof. The standards of care for the most common cancers can be found on the website of National Cancer Institute (<http://www.cancer.gov/cancertopics>). The standard of care is the current treatment that is accepted by medical experts as a proper treatment for a certain type of disease and that is widely used by healthcare professionals. Standard of care is also called best practice, standard medical care, and standard therapy. The prognostic methods of the present invention can be incorporated into a treatment plan by deciding the proper standard of care. The prognostic methods may also be used in treatment plans where the standard of care has changed due to advances in medicine.

[0128] In one embodiment the prognostic methods described herein are used to determine the proper treatment in a cancer where the standard of care is primarily surgery followed by treatment to remove possible micro-metastases, such as breast cancer. Breast cancer is commonly treated by various combinations of surgery, radiation therapy, chemotherapy, and hormone therapy based on the stage and grade of the cancer.

[0129] In one embodiment the prognostic methods are used to determine the proper treatment consistent with the standard of care in Ductal carcinoma in situ (DCIS). The standard of care for this breast cancer type are:

[0130] 1. Breast-conserving surgery and radiation therapy with or without tamoxifen.

[0131] 2. Total mastectomy with or without tamoxifen.

[0132] 3. Breast-conserving surgery without radiation therapy.

[0133] The prognostic methods may be applied to determine whether or not breast conserving surgery or total

mastectomy should be performed. In the case where the PDR is below a threshold a treatment plan that includes breast conserving surgery may be chosen. In this case the tumor would be less likely to gain resistance mutations to tamoxifen or radiation. On the contrary, if the tumor has a PDR above the threshold total mastectomy may be chosen.

[0134] In another embodiment patients diagnosed with stage I, II, IIIA, and Operable IIIC breast cancer are tested with the prognostic methods as described herein. The standard of care for this breast cancer type are:

[0135] 1. Local-regional treatment:

[0136] Breast-conserving therapy (lumpectomy, breast radiation, and surgical staging of the axilla).

[0137] Modified radical mastectomy (removal of the entire breast with level I-II axillary dissection) with or without breast reconstruction.

[0138] Sentinel node biopsy.

[0139] 2. Adjuvant radiation therapy postmastectomy in axillary node-positive tumors:

[0140] For one to three nodes: unclear role for regional radiation (infra/supraclavicular nodes, internal mammary nodes, axillary nodes, and chest wall).

[0141] For more than four nodes or extranodal involvement: regional radiation is advised.

[0142] 3. Adjuvant systemic therapy

[0143] In one embodiment the prognostic methods are used to determine the correct surgery type to use. A PDR above the threshold may suggest that the treatment should include a radical mastectomy and a PDR below the threshold may indicate that breast conserving therapy is chosen. In another embodiment the prognostic methods are used to determine the proper adjuvant therapy.

[0144] In another embodiment patients diagnosed with inoperable stage IIIB or IIIC or inflammatory breast cancer are tested with the prognostic methods as described herein. The standards of care for this breast cancer type are:

[0145] 1. Multimodality therapy delivered with curative intent is the standard of care for patients with clinical stage IIIB disease.

[0146] 2. Initial surgery is generally limited to biopsy to permit the determination of histology, estrogen-receptor (ER) and progesterone-receptor (PR) levels, and human epidermal growth factor receptor 2 (HER2/neu) overexpression. Initial treatment with anthracycline-based chemotherapy and/or taxane-based therapy is standard. For patients who respond to neoadjuvant chemotherapy, local therapy may consist of total mastectomy with axillary lymph node dissection followed by postoperative radiation therapy to the chest wall and regional lymphatics. Breast-conserving therapy can be considered in patients with a good partial or complete response to neoadjuvant chemotherapy. Subsequent systemic therapy may consist of further chemotherapy. Hormone therapy should be administered to patients whose tumors are ER-positive or unknown. All patients should be considered candidates for clinical trials to evaluate the most appropriate fashion in which to administer the various components of multimodality regimens.

[0147] In one embodiment the prognostic methods are used to determine the most appropriate fashion in which to administer the various components of multimodality regimens.

[0148] In another embodiment the prognostic methods described herein are used to determine the proper treatment

in a cancer where the standard of care is primarily not surgery and is primarily based on systemic treatments, such as CLL.

[0149] In another embodiment patients diagnosed with stage 0 Chronic Lymphocytic Leukemia are tested with the prognostic methods as described herein. The standard of care for this cancer type is:

[0150] 1. Because of the indolent nature of stage 0 chronic lymphocytic leukemia (CLL), treatment is not indicated.

[0151] In one embodiment the prognostic methods are used to monitor a patient's cancer. In another embodiment a change in PDR may indicate that treatment should be initiated.

[0152] In another embodiment patients diagnosed with stage I, II, III, and IV Chronic Lymphocytic Leukemia are tested with the prognostic methods as described herein. The standard of care for this cancer type is:

[0153] 1. Observation in asymptomatic or minimally affected patients.

[0154] 2. Rituximab

[0155] 3. Ofatumomab

[0156] 4. Oral alkylating agents with or without corticosteroids.

[0157] 5. Fludarabine, 2-chlorodeoxyadenosine, or pentostatin

[0158] 6. Bendamustine.

[0159] 7. Lenalidomide.

[0160] 8. Combination chemotherapy.

[0161] combination chemotherapy regimens include the following: Fludarabine plus cyclophosphamide plus rituximab.

[0162] Fludarabine plus rituximab as seen in the CLB-9712 and CLB-9011 trials.

[0163] Fludarabine plus cyclophosphamide versus fludarabine plus cyclophosphamide plus rituximab.

[0164] Pentostatin plus cyclophosphamide plus rituximab as seen in the MAYO-MC0183 trial, for example.

[0165] Ofatumumab plus fludarabine plus cyclophosphamide.

[0166] CVP: cyclophosphamide plus vincristine plus prednisone.

[0167] CHOP: cyclophosphamide plus doxorubicin plus vincristine plus prednisone.

[0168] Fludarabine plus cyclophosphamide versus fludarabine as seen in the E2997 trial [NCT00003764] and the LRF-CLL4 trial, for example.

[0169] Fludarabine plus chlorambucil as seen in the CLB-9011 trial, for example.

[0170] 9. Involved-field radiation therapy.

[0171] 10. Alemtuzumab

[0172] 11. Bone marrow and peripheral stem cell transplantations are under clinical evaluation.

[0173] 12. Ibrutinib

[0174] In one embodiment the prognostic methods are used as a tool to further evaluate the best treatment combination for CLL. In another embodiment the prognostic methods are used to evaluate the best treatment combination for an individual patient in need thereof. In one embodiment a more aggressive treatment strategy is employed when the PDR is above the threshold and in another embodiment a less aggressive treatment strategy is employed.

Methods of Identifying an Antitumor Agent

[0175] A fourth aspect of the present invention provides a method for identifying an antitumor agent that decreases the potential evolutionary capacity of cancer (i.e., plasticity) and, thus, the risk of relapse. In some embodiments, the method comprises growing a first culture of hyperproliferative cells and a second culture of hyperproliferative cells, wherein the first culture is grown in the presence of an antitumor agent and the second culture is grown in the absence of the antitumor agent; detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells from the first culture and a plurality of cells from the second culture; comparing the DNA methylation status of one or more regions of neighboring CpG sites along a sequence of CpG sites in the plurality of cells of the first culture and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the first culture; comparing the DNA methylation status of one or more regions of neighboring CpG sites along a sequence of CpG sites in the plurality of cells of the second culture and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells in the second culture; and assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells. For example, the DNA methylation status of CpG sites along one or more sequencing reads, e.g., operatively linked to each other in a single polynucleotide molecule, may be detected. The neighboring CpG sites along the sequencing read may then be compared to each other or to corresponding positions of different sequencing reads (e.g., at the same genomic location) from the plurality of cells. The antitumor agent decreases the potential evolutionary capacity of cancer if the number of regions of neighboring CpG sites having locally disordered methylation status is less in the first culture than in the second culture. In some embodiments, the method comprises calculating a PDR, a variance, an epipolymorphism, or an information entropy in the first and second culture, wherein the antitumor agent decreases the potential evolutionary capacity of cancer if the PDR of the first culture is less than the PDR, the variance, the epipolymorphism, or the information entropy of the second culture. Optionally, each region of neighboring CpG sites (e.g., within a sequencing read) is assigned a consistent status or an inconsistent status before calculating the proportion of discordant reads, variance, epipolymorphism or information entropy. There may be multiple inconsistent statuses, each representing a distinct methylation pattern or class of similar methylation patterns.

[0176] In some embodiments, the method comprises treating an animal model of a cancer, wherein a first animal is treated with an antitumor agent and the second animal is treated with a placebo or no antitumor agent; detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells from the first animal and a plurality of cells from the second animal; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the first animal; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of

the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the second animal; and assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells. The antitumor agent decreases the potential evolutionary capacity of cancer if the level of inconsistent methylation status is less in the first animal than in the second animal. In another embodiment the antitumor agent is administered to more than one cell or animal model using a range of doses. In some embodiments the cancer animal models are selected from the National Cancer Institutes Cancer Model Database (<https://cancer-models.nci.nih.gov/camod/>).

[0177] The hyperproliferative cells in the first and second cultures may comprise cells from a cell line, e.g., a tumor cell line. Alternatively, the hyperproliferative cells in the first and second cultures may be cells from a tumor sample obtained from a subject, preferably a human, or cells cultured from such a sample. In some embodiments, the first and second cultures are the same culture, wherein the second culture is a sample of the hyperproliferative cells before addition of the antitumor agent and the first culture is a sample of the hyperproliferative cells after addition of the antitumor agent.

[0178] The first culture may be cultured in the presence of the antitumor agent for at least 6 hours, at least 12 hours, at least 18 hours, at least one day, at least two days, at least three days, at least four days, at least five days, at least six days or at least one week, preferably at least one day, prior to detecting methylation status.

[0179] In some embodiments, the method comprises performing a first prognostic method on a first tumor sample from a subject, such as a laboratory animal, as described herein; administering an antitumor agent to the subject; and performing a second prognostic method on a second tumor sample from the subject as described herein. The antitumor agent decreases the potential evolutionary capacity of cancer if the number of genomic regions having locally disordered methylation status is less in the second tumor sample than in the first tumor sample.

[0180] The methylation status of neighboring CpG sites may be compared by calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism, or calculating information entropy. In some embodiments, a proportion of discordant reads (PDR) is calculated. Optionally, each region of neighboring CpG sites (e.g., within a sequencing read) is assigned a consistent status or an inconsistent status before calculating the proportion of discordant reads, variance, epipolymorphism or information entropy. There may be multiple inconsistent statuses, each representing a distinct methylation pattern or class of similar methylation patterns.

[0181] In some embodiments, the method comprises calculating a first PDR, variance, epipolymorphism, or information entropy from a first tumor sample obtained from the subject, such as a laboratory animal, according to the methods described herein; treating the subject with an antitumor agent; and calculating a second PDR, variance, epipolymorphism, or information entropy from a second tumor sample obtained from the subject according to the methods described herein, wherein the antitumor agent is administered between obtaining the first and second tumor samples. The antitumor agent decreases the potential evo-

lutionary capacity of cancer if the second PDR, variance, epipolymorphism, or information entropy is less than the first PDR, variance, epipolymorphism, or information entropy.

[0182] The antitumor agent may be administered to the subject for at one day, at least two days, at least three days, at least 4 days, at least five days, at least six days, at least one week, at least two weeks, at least three weeks, at least one month, preferably at least one week, e.g., prior to performing the second prognostic method or calculating the second PDR, variance, epipolymorphism, or information entropy.

[0183] The tumor sample may be a solid tumor, such as carcinomas, sarcomas and lymphomas. In some embodiments, the solid tumor is selected from adrenocortical carcinoma, bone tumors, brain cancer, breast cancer, cervical cancer, colorectal carcinoma, desmoid tumors, desmoplastic small round cell tumors, endocrine tumors, esophageal cancer, Ewing sarcoma family tumors, gastric cancer, germ cell tumors, head or neck cancer, hepatoblastoma, hepatocellular carcinoma, lung cancer, melanoma, mesothelioma, nasopharyngeal carcinoma, neuroblastoma, non-rhabdomyosarcoma soft tissue sarcoma, osteosarcoma, ovarian cancer, pancreatic cancer, prostate cancer, retinoblastoma, rhabdomyosarcoma, skin carcinoma, testicular cancer, thyroid carcinoma, uterine cancer and Wilms tumors. The tumor sample may be a hematological cancer, such as leukemia, preferably CLL.

[0184] The antitumor agent is selected from an angiogenesis inhibitor, such as angiostatin K1-3, DL- α -Difluoromethyl-ornithine, endostatin, fumagillin, genistein, minocycline, staurosporine, and ($\pm$)-thalidomide; a DNA intercalator/cross-linker, such as Bleomycin, Carboplatin, Carmustine, Chlorambucil, Cyclophosphamide, cis-Diammineplatinum (II) dichloride (Cisplatin), Melphalan, Mitoxantrone, and Oxaliplatin; a DNA synthesis inhibitor, such as ($\pm$)-Ame-thopterin (Methotrexate), 3-Amino-1,2,4-benzotriazine 1,4-dioxide, Aminopterin, Cytosine β -D-arabinofuranoside, 5-Fluoro-5'-deoxyuridine, 5-Fluorouracil, Ganciclovir, Hydroxyurea, and Mitomycin C; a DNA-RNA transcription regulator, such as Actinomycin D, Daunorubicin, Doxorubicin, Homoharringtonine, and Idarubicin; an enzyme inhibitor, such as S(+)-Camptothecin, Curcumin, (-)-Deguelin, 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside, Etoposide, Formestane, Fostriecin, Hispidin, 2-Imino-1-imidazoli-dineacetic acid (Cyclocreatine), Mevinolin, Trichostatin A, Tyrphostin AG 34, and Tyrphostin AG 879; a gene regulator, such as 5-Aza-2'-deoxycytidine, 5-Azacytidine, Cholecalciferol (Vitamin D3), 4-Hydroxytamoxifen, Melatonin, Mifepristone, Raloxifene, all trans-Retinal (Vitamin A aldehyde), Retinoic acid, all trans (Vitamin A acid), 9-cis-Retinoic Acid, 13-cis-Retinoic acid, Retinol (Vitamin A), Tamoxifen, and Troglitazone; a microtubule inhibitor, such as Colchicine, Dolastatin 15, Nocodazole, Paclitaxel, Podophyllotoxin, Rhizoxin, Vinblastine, Vincristine, Vindesine, and Vinorelbine (Navelbine); and an unclassified antitumor agent, such as 17-(Allylamino)-17-demethoxygeldanamycin, 4-Amino-1,8-naphthalimide, Apigenin, Brefeldin A, Cimetidine, Dichloromethylene-diphosphonic acid, Leuprolide (Leuprorelin), Luteinizing Hormone-Releasing Hormone, Pifithrin- α , Rapamycin, Sex hormone-binding globulin, Thapsigargin, and Urinary trypsin inhibitor fragment (Bikunin). The antitumor agent may be a monoclonal antibody such as rituximab (Rituxan®), alemtuzumab (Campath®), Ipilimumab (Yervoy®), Bevacizumab (Avastin®), Cetuximab (Erbix®), panitumumab (Vectibix®), and tras-

tuzumab (Herceptin®), Vemurafenib (Zelboraf®) imatinib mesylate (Gleevec®), erlotinib (Tarceva®), gefitinib (Iressa®), Vismodegib (Erivedge™), 90Y-ibritumomab tiuxetan, 131I-tositumomab, ado-trastuzumab emtansine, lapatinib (Tykerb®), pertuzumab (Perjeta™), ado-trastuzumab emtansine (Kadcyla™), regorafenib (Stivarga®), sunitinib (Sutent®), Denosumab (Xgeva®), sorafenib (Nexavar®), pazopanib (Votrient®), axitinib (Inlyta®), dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (Bosulif®), ofatumumab (Arzerra®), obinutuzumab (Gazyva™), ibrutinib (Imbruvica™), idelalisib (Zydelig®), crizotinib (Xalkori®), erlotinib (Tarceva®), afatinib dimaleate (Gilotrif®), ceritinib (LDK378/Zykadia), Tositumomab and 131I-tositumomab (Bexxar®), ibritumomab tiuxetan (Zevalin®), brentuximab vedotin (Adcetris®), bortezomib (Velcade®), siltuximab (Sylvant™), trametinib (Mekinist®), dabrafenib (Tafinlar®), pembrolizumab (Keytruda®), carfilzomib (Kyprolis®), Ramucirumab (Cyramza™), Cabozantinib (Cometriq™), vandetanib (Caprelsa®). Optionally, the antitumor agent is a neoantigen. The antitumor agent may be a cytokine such as interferons (INFs), interleukins (ILs), or hematopoietic growth factors. The antitumor agent may be INF- α , IL-2, Aldesleukin, IL-2, Erythropoietin, Granulocyte-macrophage colony-stimulating factor (GM-CSF) or granulocyte colony-stimulating factor. The antitumor agent may be a targeted therapy such as toremifene (Fareston®), fulvestrant (Faslodex®), anastrozole (Arimidex®), exemestane (Aromasin®), letrozole (Femara®), ziv-aflibercept (Zaltrap®), Alitretinoin (Panretin®), temsirolimus (Torisel®), Tretinoin (Vesanoid®), denileukin difitox (Ontak®), vorinostat (Zolinza®), romidepsin (Istodax®), bexarotene (Targretin®), pralatrexate (Folotyn®), lenaliomide (Revlimid®), belinostat (Beleodaq™), lenaliomide (Revlimid®), pomalidomide (Pomalyst®), Cabazitaxel (Jevtana®), enzalutamide (Xtandi®), abiraterone acetate (Zytiga®), radium 223 chloride (Xofigo®), or everolimus (Afinitor®). The antitumor agent may be a checkpoint inhibitor such as an inhibitor of the programmed death-1 (PD-1) pathway, for example an anti-PD1 antibody (Nivolumab). The inhibitor may be an anti-cytotoxic T-lymphocyte-associated antigen (CTLA-4) antibody. The inhibitor may target another member of the CD28 CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. A checkpoint inhibitor may target a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3. Additionally, the antitumor agent may be an epigenetic targeted drug such as HDAC inhibitors, kinase inhibitors, DNA methyltransferase inhibitors, histone demethylase inhibitors, or histone methylation inhibitors. The epigenetic drugs may be Azacitidine (Vidaza), Decitabine (Dacogen), Vorinostat (Zolinza), Romidepsin (Istodax), or Ruxolitinib (Jakafi).

[0185] In some embodiments, the laboratory animal is a mouse, a rabbit, a rat, a guinea pig, a hamster, or a primate.

[0186] An unexpected advantage of the present invention is that the treatment of a patient in need thereof is greatly improved and personalized based on the analysis of DNA methylation discordance. The analysis is based on an unexpected fundamental difference between cancer and normal methylomes: locally disordered methylation arising from a stochastic process, that leads to a high degree of intra-sample methylation heterogeneity.

[0187] Another advantage is that the methods of the present invention allow a patient's tumor to be evaluated for stochastic methylation changes that enhance epigenetic plas-

ticity and likewise enable tumor cells to better explore the evolutionary space in search of superior fitness trajectories. Treatment regimens that include this analysis can be made more or less aggressive or use different modalities.

[0188] The present invention provides improved methods to identify fitness-enhancing differentially methylated regions.

[0189] Although the present invention and its advantages have been described in detail, it should be understood that various changes, substitutions and alterations can be made herein without departing from the spirit and scope of the invention as defined in the appended claims.

[0190] The present invention will be further illustrated in the following Examples which are given for illustration purposes only and are not intended to limit the invention in any way.

EXAMPLES

Example 1

Genetic Heterogeneity is Associated with Clonal Evolution

[0191] Applicants used an analytic approach in which whole-exome sequencing (WES) was used to infer the fraction of cancer cells (CCF) that harbors each somatic mutation in 149 CLLs by correcting the allelic fraction measured by WES for sample purity and local copy number at the mutated sites. To directly assess clonal evolution (FIG. 1A), Applicants compared CCF for each mutation at 2 time points in 18 CLL cases (median interval between time points 3.5 yrs.). Applicants observed a higher rate of clonal evolution in 12 cases that received treatment between time points compared with 6 continuously untreated cases ($P=0.012$). Clonal evolution involved the expansion of subclones that harbored mutations highly enriched in driver mutations. These subclonal driver mutations were often already detectable in the pretreatment sample (red clusters, FIG. 1A right). Applicants therefore hypothesized that the pretreatment detection of subclonal drivers would be associated with worse clinical outcome with treatment. Indeed, in a cohort of 67 CLL patients who received CLL treatment, the detection of a subclonal driver mutation was associated with earlier retreatment or death, indicative of a more aggressive disease course (independently of known risk factors, $P=0.006$, FIG. 1B). Thus, the pre-treatment identification of subclonal mutations can provide information regarding the rapidity and genetic composition of the relapsing leukemia.

Example 2

DNA Methylation Heterogeneity in CLL

[0192] While investigations of evolutionary dynamics in cancer have focused primarily on the role of genetic alterations, epigenetic modifications are likely also responsible for the phenotypic differences that ultimately affect fitness.

[0193] Applicants therefore assessed the degree of DNA methylation heterogeneity in CLL. Applicants performed a re-analysis of a published set of DNA methylation arrays (Kulis M, et al. Nat Genet, 2012; 44:1236-42) and found that the 127 CLL samples showed a higher degree of DNA methylation heterogeneity compared with 25 normal B cell samples (FIG. 2A-B). Applicants confirmed the higher heterogeneity in CLL by reanalyzing published whole genome

bisulfite sequencing (WGBS) data (FIG. 2C). Applicants also studied this question with reduced representation bisulfite sequencing (RRBS) which allowed the use of MPS technology to study DNA methylation in larger cohorts. Applicants found higher methylation heterogeneity in CLL by comparing the mean weighted variance of ~2.5 million CpGs in 101 CLL samples vs. 22 normal B cell samples (FIG. 2D). Similar to other lymphoid malignancies and solid tumors, these data robustly demonstrate that CLL contains increased genome-wide methylation heterogeneity compared with normal B cells.

Example 3

Increased Intra-Sample DNA Methylation Heterogeneity in CLL Arises from Locally Disordered Methylation

[0194] To measure intra-sample CLL DNA methylation heterogeneity, Applicants compared WGBS data generated from two CLL cases and two healthy donor B cell samples (FIG. 3A). Applicants observed globally decreased methylation in CLL compared to normal B cells, with focally increased methylation of CpG islands (CGI) (FIG. 3A—top, FIG. 4A-C), as previously reported in CLL and other cancers (Baylin and Jones, 2011; Kulis et al., 2012), but also a markedly increased frequency of intermediate methylation values in CLL (FIG. 3A-bottom, FIG. 4A-D), pointing to a large proportion of CpGs that are methylated in some cells in the sample and unmethylated in others. Applicants reanalyzed published WGBS and 450K methylation array data (Kulis et al., 2012) and confirmed the increased cell-to-cell variability in CpG methylation in CLL compared to normal B cells (FIG. 4E-H).

[0195] Applicants next applied RRBS to 104 primary CLL samples that had been previously characterized by WES (Landau et al., 2013) (Table 1), and examined mean CpG variance. Consistent with the WGBS data, a greater than 50% increase in intra-sample methylation heterogeneity was detected in CLL cells compared to 26 normal B cell samples ($p=5.13 \times 10^{-14}$; FIG. 3B). Applicants considered two possible sources for intra-sample heterogeneity: variability between concordantly methylated fragments (i.e., whereby CpGs in an individual fragment are consistently methylated or unmethylated; FIG. 3C, left); or variability within DNA fragments (i.e., discordant methylation by which CpGs in an individual fragment are variably methylated; FIG. 3C, right).

[0196] Based on established observations that short-range methylation is highly correlated in normal physiological states (Eckhardt et al., 2006; Jones, 2012), Applicants initially hypothesized that intra-sample heterogeneity in CLL stems from variability between concordantly methylated fragments, reflecting a mixture of subpopulations with distinct but uniform methylation patterns. To test this, Applicants focused on CpGs covered by reads containing 4 or more neighboring CpGs, as previously suggested (Landan et al., 2012), and with sufficient read depth (greater than 10 reads per CpG, with ~6.5 million CpGs/sample covered by 100mer WGBS reads, and an average of 307,041 [range 278,105-335,977] CpGs/sample covered by 29mer RRBS reads). Contrary to the expected hypothesis, Applicants found that $67.6 \pm 3.2\%$ (average $\pm$ SD) of the intra-tumoral methylation variance resulted from discordantly methylated reads across the 104 CLL samples (FIG. 3D; $p=3.24 \times 10^{-35}$).

Similarly, the CLL WGBS confirmed a higher proportion of heterogeneously methylated CpGs in the discordant reads compared with the concordant reads (FIG. 4E, right). These results demonstrate that methylation heterogeneity in CLL primarily arises from variability within DNA fragments, which Applicants have therefore termed ‘locally disordered methylation’.

[0197] Applicants performed several analyses to exclude potential alternative explanations to these findings, including the impact of contaminating non-malignant cells (FIG. 4I), allele-specific methylation (FIG. 4J-L), the contribution of reads that cover an ordered transition point from one methylation state to another (FIG. 9L) and technical biases (see Methods). The sex chromosomes were excluded from this analysis to avoid possible confounding sex-chromosome specific effects. In addition, CLL genomes are near diploid (Brown et al., 2012), and therefore the analysis was not significantly impacted by somatic copy number variations (see methods, FIG. 4O).

[0198] To quantify the magnitude of this phenomenon across large collections of normal and malignant human tissues, Applicants analyzed RRBS data not only from the 104 CLL and 26 B cells samples, but also from 45 solid and blood cancer cell lines and from 27 primary human tissue samples. Applicants then calculated the proportion of discordant reads (PDR) as the number of discordant over the total number of reads for each CpG in the consensus set (FIG. 3E). Applicants found that the average PDR was higher in CLL compared to normal B cells ($p=5.60 \times 10^{-14}$). Similarly, Applicants found higher PDR in cancer cell lines compared to a diverse collection of healthy human tissue samples ($p=4.35 \times 10^{-12}$, FIG. 1F). These results support the idea that locally disordered methylation is a general property of the malignant process.

TABLE 1

Characteristics and mean promoter PDR of the 104 CLL patients whose DNA were analyzed by WES and RRBS.			
	N (%)	Mean Promoter PDR (SD)	p value [†]
N	104		
Age (median = 54 yrs)			
<54 yrs.	46 (44)	0.101 (0.016)	0.15
≥54 yrs.	58 (56)	0.105 (0.016)	
Sex			
Female	38 (37)	0.106 (0.016)	0.30
Male	66 (63)	0.102 (0.016)	
Rai Stage at Sample			
0-1	78 (75)	0.102 (0.016)	0.049
2-4	26 (25)	0.109 (0.015)	
Treatment Status at time of Sample			
Chemotherapy naïve	82 (79)	0.103 (0.017)	0.59
Prior Treatment	22 (21)	0.105 (0.014)	
IGHV status			
Mutated	57 (55)	0.107 (0.017)	0.035
Not Mutated	34 (33)	0.0996 (0.014)	
Unknown	13 (13)	0.0977 (0.014)	
FISH Cytogenetics ^{††}			
del(13q) present	67 (67)	0.105 (0.016)	0.059
absent	33 (33)	0.099 (0.016)	

TABLE 1-continued

Characteristics and mean promoter PDR of the 104 CLL patients whose DNA were analyzed by WES and RRBS.			
	N (%)	Mean Promoter PDR (SD)	p value [†]
Trisomy 12 present	18 (18)	0.099 (0.015)	0.21
absent	82 (82)	0.104 (0.016)	
del(11q) present	18 (18)	0.095 (0.013)	0.019
absent	82 (82)	0.105 (0.016)	
del(17p) present	14 (14)	0.105 (0.016)	0.62
absent	86 (86)	0.103 (0.016)	
Mutational Status			
Subclonal Mutation Present	49 (47)	0.105 (0.016)	0.25
Absent	55 (53)	0.102 (0.016)	
TP53 Present	15 (14)	0.110 (0.016)	0.091
Absent	89 (86)	0.102 (0.016)	
NOTCH1 Present	11 (11)	0.096 (0.016)	0.097
Absent	93 (89)	0.104 (0.016)	
SF3B1 Present	9 (9)	0.108 (0.015)	0.36
Absent	95 (90)	0.103 (0.016)	
MYD88 Present	8 (8)	0.111 (0.009)	0.19
Absent	96 (92)	0.103 (0.016)	
ATM Present	6 (6)	0.112 (0.017)	0.18
Absent	98 (94)	0.103 (0.016)	

[†]Testing excludes unknown categories; Welch t-test (variances were not significantly different);

^{††}N = 100

Example 4

Epigenetic Variation in CLL is Predominantly Locally Disordered DNA Methylation

[0199] Heterogeneity could arise from two possible methylation patterns: 1) mixing of two subpopulations with ordered but distinct methylation states for a particular locus (FIG. 5A left), or 2) admixture of cells with “noisy” disordered methylation at the particular CpG locus (FIG. 5A right). The ability to distinguish between the two relies on the fact that short-range methylation state concordance between neighboring CpGs is expected to be very high in non-disease states, as DNA methylation typically changes by feature (e.g., a specific gene promoter, or a CpG island) rather than by individual CpG. Applicants therefore studied the concordance of methylation status across multiple CpGs contained within an individual 29-bp long sequencing read, and compared the relative contribution of concordant vs. discordant reads to the overall variance. Applicants observed a significantly larger contribution from locally disordered methylation seen in CLL in the RRBS data (FIG. 5B). Applicants also defined the proportion of discordant reads (PDR) for each CpG as a simple measure of locally disordered DNA methylation (FIG. 5C).

[0200] Applicants next identified the genomic elements affected by locally disordered methylation. The increase in the PDR in CLL compared to normal samples was prominent in most genomic elements (FIG. 6A), with a more marked degree of change in CpG islands (CGI) compared with their shelves or shores (FIG. 6B). In addition, Applicants studied RRBS data of 50 cancer cell lines (from solid and hematological malignancies included in the CCLE collection) and 32 samples from various normal human tissues, and showed that while locally disordered DNA methylation exists in normal differentiated cells, it is significantly higher

in both CLLs and cancer cell lines (FIG. 6C). Performing a similar analysis in WGBS allowed Applicants to probe genomic areas that are not well covered by RRBS. Repetitive regions, showed a marked increase in PDR in CLL, as did the large hypo-methylated blocks (FIG. 6D).

[0201] While CLLs tended to have uniformly high PDR, the mean-weighted PDR varied between 0.186 and 0.265. RRBS covers, on average, ~2.5 million CpGs, which are enriched in regions important for transcriptional regulation. Therefore, even small differences may signify changes affecting thousands of CpGs genome-wide. For example, when comparing within CLL samples, Applicants found that older age at diagnosis (age>median) was associated with an increase in PDR (average $\pm$ 95% CI, 0.24 $\pm$ 0.005 vs. 0.23 $\pm$ 0.005, $P=0.028$), consistent with data showing that aging in itself is associated with epigenetic drift, and suggesting that some of the local methylation disorder may originate prior to the malignant transformation in the leukemia initiating cell. Other clinical factors were not associated with significant differences (e.g., the IGHV mutation status, an important predictor of poor outcome in CLL showed no association with PDR [$P=0.175$]). Samples with somatic mutations in methylation modulators (DNMT3A, TET1, and IDH1, each $n=1$) had increased PDR compared with 98 wild type samples (0.251 $\pm$ 0.01 vs. 0.228 $\pm$ 0.004, $P=0.027$). These data suggest that a genotype enhancing the potential for epigenetic drift may be selected for, as was seen in other malignancies. When compared to normal B cells samples ($n=32$), several methylation regulators were significantly differentially expressed in CLL ($n=247$) by expression array analysis ($Q<0.1$, FDR). In particular, although DNMT1 was up-regulated by 2.25-fold in CLL (consistent with reports across cancer37), DNMT3B and TET1 were down-regulated by 3- and 7-fold, respectively. This latter finding suggests that several enzymes that regulate DNA methylation are down-regulated compared to normal B cells, and may contribute to the increase in locally disordered methylation. These results are consistent with previous reports in lymphoid malignancies that suggest that DNMT3A and DNMT3B inhibition may promote the lympho-proliferative process. Collectively, we show that much of the heterogeneity of DNA methylation in CLL results from locally disordered methylation. This form of epigenetic drift may cause some of the epigenetic alterations seen in CLL and cancer.

Example 5

Locally Disordered Methylation Broadly Affects the CLL Genome

[0202] To determine whether specific elements in the genome harbor higher levels of locally disordered methylation in CLL compared to normal B cells, Applicants calculated the average PDR across the 104 CLL samples and 26 healthy donor B cell samples (Table 2).

[0203] In normal B cells, PDR levels were lowest in regions with a major role in gene regulation (promoters, CGI, exons, enhancers), and higher in regions with presumably less of a regulatory role (CGI shelves and shores, intergenic regions). In CLL, PDR was higher across all measured regions (FIG. 7A), regardless of whether they were relatively hypermethylated (e.g. CGI) or hypomethylated (e.g. intergenic regions) compared to normal B cells (FIG. 7B). This phenomenon appeared to be neither specific

to a sub-region of CGIs or promoters (for example, CGI borders, FIG. 7C), nor restricted to a subtype of CGI (FIG. 8A). Increased PDR in CLL was also observed in highly repetitive DNA sequences (e.g., long interspersed elements [LINE] and long terminal repeat [LTR] retrotransposons; FIG. 7A-RRBS data and FIG. 8B—WGBS data), which largely account for the global DNA hypomethylation observed in cancer (Ehrlich, 2009).

[0204] Alterations in the DNA methylation regulatory machinery could impact PDR. Unlike other hematological malignancies (Ley et al., 2010), somatic mutations affecting direct DNA methylation modulators in CLL are rare (Landau et al., 2013). Nonetheless, three CLL samples with such somatic mutations (DNMT3A-Q153*, TET1-N789I, IDH1-S210N) showed increased PDR compared to the 101 CLL samples wildtype for these genes (FIG. 8C).

TABLE 2

Average number of CpGs covered by RRBS with 4 or more CpGs per read, and read depth greater than 10, given by genomic feature.

Genomic feature	CLL samples		Normal B cell samples		Total # of CpGs in the human genome
	Mean	Standard Dev.	Mean	Standard Dev.	
Promoters	129212.20	81086.22	163485.80	107845.00	1954610
CpG islands	171342.80	108393.30	215941.40	143384.80	2124041
Exons	71536.45	45162.34	83029.35	54823.08	1954610
Enhancers	92322.87	62354.34	91093.69	62145.70	1176256
Introns	155736.40	102363.70	164236.30	110325.50	14479789
Genes	29397.06	18483.37	34193.31	22639.84	26917396
LTR	14222.88	9679.65	9102.58	6533.52	2133049
LINE	5256.03	3754.65	3055.23	2309.95	3516060
Shores	2488.20	1755.74	2349.08	1801.29	3886809
Shelves	3094.95	2403.47	4682.19	3636.86	1259327
Intergenic	35881.12	26900.36	18867.27	15753.56	5087650

Example 6

Locally Disordered Methylation Appears to be a Largely Stochastic Process

[0205] Two observations suggest that PDR measures a process that stochastically increases variation in methylation, a notion which was recently conceptualized as a feature of the cancer epigenome (Pujadas and Feinberg, 2012). First, the pervasiveness of locally disordered methylation across every region evaluated in CLL compared to B cells suggested a stochastic genome-wide process. Second, consistent with a stochastic process, wherein the expected rate of increase in PDR would be related to the starting level of disorder, Applicants observed a larger relative PDR increase in CLL in regions with lower PDR in normal B cells. To formally measure the level of disorder, Applicants undertook a parallel analysis to calculate Shannon's information entropy of intra-sample methylation variation (FIG. 9A). Applicants determined this entropy to be higher in CLL than in normal B cells (as well as higher in cancer cell lines compared to normal tissues), consistent with an increase in stochastic 'noise' (FIG. 9B-C).

[0206] To model the relationship between methylation and PDR under completely stochastic conditions, Applicants plotted the expected distribution of PDR for any level of methylation assuming a purely random assignment of methylation states at each individual CpG (FIG. 10A, see meth-

ods). Strikingly, the distribution of measured PDR and methylation values of ~14,000 individual promoter CGIs from CLL WGBS data closely followed the pattern of the modeled stochastic process (FIG. 10B). In outlier genes (i.e. those with less promoter PDR than expected based on the promoter methylation level; $n=195$ (1.4%), FIG. 9D), imprinted genes were enriched (Morison et al., 2005) as expected, since these are hemimethylated under normal physiological conditions ($n=10$, Fishers exact test $p=1.94 \times 10^{-6}$). In addition, the outlier genes contained at least three tumor suppressor genes (WIF1, DUSP22 and DCC) that have an established role in hematopoietic malignancies (Chim et al., 2008; Inokuchi et al., 1996; Jantus Lewintre et al., 2009), and also had >10% higher methylation in the CLL169 sample compared with the normal CD19⁺ B cell sample.

[0207] Similar to promoters, methylation of ~1900 LINE repeat elements also displayed a similar relationship between methylation and PDR (FIG. 10C). A comparable distribution was observed for other genomic features (FIG. 9E), and with RRBS data (FIG. 9F). This pattern was also found in promoter CpGs of tumor suppressor genes implicated in lymphoproliferation, such as WT1 (Menke et al., 2002) and DAPK1 (Raval et al., 2007) (FIG. 9G).

[0208] Altogether, these data support the hypothesis that the most commonly described cancer-related methylation alterations (Baylin and Jones, 2011), increased methylation of CGIs and decreased methylation in repeat regions, are largely generated through a seemingly stochastic process. Indeed, across the 104 CLLs, sample average promoter CGI PDR was highly correlated with an increase in sample average promoter CGI methylation (Pearson correlation coefficient $r=0.90$, $p=1.01 \times 10^{-38}$, FIG. 10D). When this analysis was repeated with genes grouped based on their average methylation level across the samples, this strong correlation was positive for genes with methylation <0.5 and negative for genes with methylation >0.5 as expected from the afore-described distribution in FIG. 10B (FIG. 9H). Overall, a key implication of this analysis is that a change in CGI methylation in CLL does not arise from alteration in a relatively small proportion of cells with uniformly methylated alleles but rather from a larger proportion of cells with randomly scattered methylation. Applicants likewise observed sample average LINE repeat elements PDR to be correlated with a decrease in methylation ($r=-0.32$, $p=6.99 \times 10^{-4}$, FIG. 10E).

[0209] These data reveal that DNA methylation changes in this cancer predominately arise from a disordered change in methylation, resulting in a strong correlation between difference in PDR (Δ PDR) and difference in methylation (Δ Meth). Since previous reports have indicated that a large degree of methylation disorder occurs during normal differentiation (Landan et al., 2012), Applicants sought to compare the correlation between Δ PDR and Δ Meth amongst pairs of cancer and normal samples, to the correlation between pairs of healthy human tissues. Indeed, the correlation coefficient between Δ PDR and Δ Meth was significantly higher when CLL samples were paired to either normal B cells or to other healthy primary tissue samples, compared to the pairing of healthy primary tissues against either normal B cells or other healthy tissue samples (FIG. 10F-G). Thus, methylation changes associated with the malignant process differ substantially from those that occur

during changes in physiological cellular states, and show a significantly higher degree of methylation disorder.

Example 7

Increased Susceptibility to Locally Disordered Methylation in Gene-Poor Regions and Silent Genes

[0210] Some regions of the genome may be more prone to stochastic variation in methylation (Pujadas and Feinberg, 2012). Applicants found three-fold higher promoter PDR in regions with the lowest gene density compared to those with highest gene density (with similar correlations to CTCF density, FIG. 11A). In addition, previously described hypomethylated blocks are regions notable for their association with the nuclear lamina, and furthermore, are enriched with genes that have high expression variability in cancer and impact critical cellular processes such as mitosis and cell cycle control (Hansen et al., 2011; Timp and Feinberg, 2013). In these regions as well, Applicants observed a significant PDR increase in CLL (FIG. 11B, FIG. 12A). Finally, in concert with these findings Applicants observed higher promoter PDR in genes with later replication time across the 104 CLL samples (FIG. 11C, $r=0.35$, $P=1.3 \times 10^{-153}$), in agreement with other recent reports (Berman et al., 2012; Shipony et al., 2014). Notably, late replication time is closely associated with increased somatic mutation rate (Lawrence et al., 2013). Thus, similar genomic regions may share lower genetic and epigenetic fidelity, as we observed in a joint analysis of somatic single nucleotide variants (sSNVs) and locally disordered methylation (FIG. 11D, FIG. 12B-C).

[0211] As many features of chromatin and spatial organization may be shared between the CLL and normal B cell genomes, Applicants hypothesized that some degree of locally disordered methylation might exist in normal B cells in regions with high PDR in CLL. In fact, average PDR of individual CGI in CLL and B cell samples was highly correlated (FIG. 11E-left, $r=0.83$, $p<2 \times 10^{-16}$). Thus, the promoters with highest PDR in CLL already have increased PDR in normal B cells. Consistent with the notion that non-expressed genes are the most vulnerable to aberrant methylation (Meissner et al., 2008), promoter CGIs with a high PDR in both CLL and normal B cells were often found in genes not expressed in normal B cells (FIG. 11E-right).

Example 8

Disordered DNA Methylation Alters the Relationship Between Promoter Methylation and Transcription

[0212] To test the impact of locally disordered methylation on CLL gene transcription, Applicants generated matching RNAseq data for 40 CLLs. Applicants then computed the odds ratio of a gene with a methylated promoter (promoter methylation >0.8 vs. promoter methylation <0.2) to be transcribed across ~8000 genes in 33 samples where sequencing coverage was sufficient to perform this analysis. Applicants found that the relationship between promoter methylation and gene transcription was markedly weakened in promoters with PDR >0.1 (mean of promoter PDR means=0.1001, FIG. 13A). These results may explain the observation that promoter methylation in CLL is not very predictive of effects on gene transcription. Indeed, the

addition of the PDR information to a model that utilizes promoter methylation to predict gene transcription (evaluated for 363,846 matched values of expression and methylation from 40 patient samples) resulted in a striking increase of the model R2 from 0.11 to 0.285. In addition, promoters with significantly skewed allele specific expression had significantly higher PDR in 26 of 40 samples evaluable (at $p < 0.05$, FIG. 13B). These data suggest that locally disordered methylation contributes to the increase transcriptional variation associated with epigenetic heterogeneity.

Example 9

Locally Disordered Methylation and Gene Expression

[0213] To examine the relationship between locally disordered DNA methylation and gene expression in more detail, Applicants analysed matched RRBS and RNA-seq profiles of 33 CLL samples (PDR and methylation calculated based on an average ($\pm$ SD) of 12.1 ($\pm$ 4.8) CpGs per promoter). As in normal B cells, in the 33 CLL samples, PDR was inversely correlated with gene expression ($r = -0.51$, $p < 2 \times 10^{-16}$, FIG. 14A, FIG. 15A-B). Notably, while promoter PDR was negatively correlated with mean transcript levels, it was positively correlated with inter-sample variation in transcript levels (FIG. 14B). While it may be difficult to definitively deconvolute the positive correlation between PDR and expression variation from the strong negative correlation of mean expression and expression variation, both low gene expression and high promoter PDR levels were predictive of higher coefficient of variation of gene expression in a linear model ($p < 2 \times 10^{-16}$ for both).

[0214] To further examine the impact of locally disordered methylation in CLL on expression levels, Applicants calculated the odds ratio of gene expression (defined as fragments per kilobase of exon per million fragments mapped (FPKM) > 1) with a methylated promoter (defined as methylation > 0.8 , unmethylated defined as < 0.2). Promoters with low PDR (i.e., lower than the mean PDR [mean ($\pm$ SD) promoter PDR was 0.10 ($\pm$ 0.01)]) tended to preserve the expected relationship between promoter methylation and expression, and rarely generated transcripts in the presence of a methylated promoter. Across 33 CLL samples, the average odds ratio (OR) was 0.043 (range 0.036-0.050). In contrast, genes with high PDR promoters ($>$ mean PDR) had a greater likelihood of undergoing transcription (OR 0.396 [range 0.259-0.698], Wilcoxon $p = 6.5 \times 10^{-11}$, FIG. 14C), despite comparable promoter methylation levels. As a representative example, Applicants show ZNF718 in two samples with comparable levels of promoter methylation (0.82 in CLL062, 0.87 in CLL074) but low promoter PDR (0.04) in the former, and high promoter PDR (0.24) in the latter. Consistent with the odds-ratio analysis above, Applicants observed undetectable expression in CLL062 (FPKM of 0.03) and measurable RNA expression in CLL074 (FPKM of 5.6) (FIG. 15C).

[0215] These observations demonstrate how locally disordered methylation and epigenetic heterogeneity may contribute to increased transcriptional variation. To assess the relationship between PDR and gene expression as continuous variables, Applicants utilized linear models to predict expression based on methylation information. Across the 33 samples, a univariate model that predicts expression based on average promoter methylation yielded an adjusted R2 of

0.092 while one utilizing promoter PDR yielded an average adjusted R2 of 0.202. Inclusion of additional features such as CpG and repeat content only modestly improved the predictive power of the model (average adjusted R2=0.214, Table 3). Indeed, the addition of PDR information to a model that utilizes promoter methylation to predict gene expression as a continuous variable (evaluated for 320,574 matched values of expression and methylation from 33 CLL) resulted in a significant improvement with more than doubling of the model's explanatory power (increase in adjusted R2 value from 0.0915 to 0.1992, likelihood ratio test $p < 1 \times 10^{-16}$). This held true when the model included only genes with lowly methylated or only genes with highly methylated promoters ($p < 1 \times 10^{-16}$). Even after adding additional variables such as repeat element content, the presence of a CGI in the promoter and CpG content, PDR remained the strongest predictor of expression (FIG. 14C-right).

TABLE 3

Results of models of prediction of gene expression for 33 CLL samples with matched RNAseq and RRBS: Values represent the adjusted R squared for the model.					
Sample	PDR + Meth + CpG_content + Repeat_content	Meth	PDR	Meth + PDR	Genes measured per sample
CLL146	0.223	0.087	0.210	0.210	6110
CLL124	0.170	0.068	0.160	0.161	7016
CLL131	0.168	0.062	0.155	0.156	7264
CLL170	0.191	0.080	0.179	0.179	7636
CLL097	0.213	0.073	0.192	0.193	8511
CLL141	0.222	0.089	0.217	0.217	8818
CLL117	0.285	0.128	0.276	0.276	8926
CLL140	0.196	0.088	0.182	0.182	9005
CLL096	0.222	0.093	0.209	0.209	9016
CLL003	0.237	0.118	0.226	0.226	9044
CLL041	0.231	0.104	0.220	0.220	9045
CLL074	0.217	0.086	0.208	0.208	9566
CLL120	0.215	0.099	0.203	0.203	9871
CLL138	0.268	0.133	0.255	0.256	9914
CLL129	0.221	0.093	0.206	0.206	9976
CLL068	0.248	0.112	0.235	0.235	10029
CLL038	0.117	0.038	0.100	0.101	10058
CLL062	0.163	0.066	0.143	0.143	10141
CLL105	0.226	0.074	0.211	0.211	10311
CLL119	0.239	0.128	0.232	0.232	10351
CLL100	0.230	0.115	0.221	0.221	10387
CLL153	0.256	0.132	0.247	0.248	10426
CLL069	0.205	0.082	0.196	0.196	10600
CLL123	0.189	0.069	0.182	0.182	10655
CLL067	0.242	0.116	0.230	0.230	10684
CLL057	0.237	0.110	0.227	0.228	10745
CLL128	0.171	0.068	0.158	0.158	10750
CLL054	0.208	0.083	0.198	0.198	10755
CLL152	0.219	0.096	0.210	0.210	10828
CLL007	0.210	0.101	0.199	0.199	10883
CLL126	0.203	0.075	0.181	0.181	11051
CLL005	0.218	0.111	0.202	0.203	11064
CLL049	0.193	0.086	0.180	0.180	11138

Example 10

Single Cell Gene Expression Patterns of Genes with Disordered Promoter Methylation

[0216] Applicants next isolated 96 individual cells from four CD19+CD5+ purified CLL samples and generated single-cell full-length transcriptomes using SMART-seq (75-84 cells analyzed per sample after excluding cells with $< 1 \times 10^4$ aligned reads). Promoter PDR was associated with

significantly higher intra-tumoral expression information entropy in all 4 samples ($p < 1.4 \times 10^{-8}$, FIG. 14D-E, FIG. 15D), in a model that included transcript length as well as population average gene expression (see Methods), which is the variable associated most closely with technical noise in single cell transcriptome analyses (Shalek et al., 2014). These results remained significant even after the addition of promoter methylation to the model (FIG. 15E). As expression information entropy may be affected by variation in sampling of lowly expressed transcripts, Applicants compared the single cell expression patterns of genes with low or high promoter methylation disorder, but with similar population average expression levels (FIG. 14F). Applicants observed that high promoter PDR genes tend to be expressed in larger numbers of cells at lower expression magnitude, whereas low promoter PDR genes tend to be expressed in smaller numbers of cells at higher expression magnitude. Thus, promoter methylation disorder correlates with an intermediate transcriptional state that interferes with both complete silencing and high-level expression.

Example 11

Epigenetic Heterogeneity Contributes to Clonal Evolution

[0217] Increased epigenetic disorder is expected to result in a more plastic evolutionary landscape that facilitates the emergence of fitness-enhancing genetic and epigenetic alterations. The footprint of selection may be inferred across samples by assessing significantly differentially expressed genes, as genes that are recurrently differentially expressed across many samples are likely selected. Applicants identified 447 down-regulated genes (FDR, $Q < 0.01$) and found that their promoter PDR was higher compared with 1770 genes that were not significantly down-regulated ($Q > 0.2$, FIG. 16A). Further, Applicants defined a gene set ($n=1357$) with promoters exhibiting high PDR (mean promoter PDR > 0.1). This set was enriched for genes shown to be differentially methylated across various malignancies, for TP53 targets and for gene sets up-regulated in stem cells (FWER $Q < 0.1$, compared to genes with PDR < 0.1 , FIG. 16B). The 'increased PDR' gene set included genes important for stem-like properties (e.g., *lin28*, and *SOX2*) and genes that have well-defined roles in CLL biology (e.g., *DAPK1*, and *TERT*). Finally, CLLs with a higher number of subclonal mutations also had higher PDR, while no association was seen with the number of clonal mutations (FIG. 16C-D). Together, these findings suggest that locally disordered methylation creates a rich substrate for CLL evolution by stochastic variation amenable to positive selection.

[0218] To probe the relationship between genetic and epigenetic evolution, Applicants performed RRBS at two time points for 13 CLLs with characterized patterns of genetic evolution (median time between time points 3.4 yrs.; 4 unevolved, 9 evolved). The PDR increase between time points was higher in evolved vs. unevolved CLLs ($P=0.029$). In addition, Applicants identified 329 genes with promoters that were demethylated over time (greater than 10% decrease, $Q < 0.1$), and observed a significant enrichment for the same stem cell related gene-sets that were described herein ($Q < 1e-10$). Genes with promoters significantly hypermethylated over time ($n=159$) were enriched for genes methylated in lymphoma. In evolved CLLs, specific promoters revealed changes over time in methylation propor-

tions corresponding to increases in subclone size inferred from the genetic analysis. For example, an increase in size of a subclone harboring an *SF3B1* somatic mutation was observed in conjunction with progressive hypomethylation, in a similar proportion of cells, of the *TERT* promoter, a critical gene for CLL proliferation.

Example 12

Locally Disordered Methylation Impacts Stem Cell Genes and May Facilitate Leukemic Evolution

[0219] Increased epigenetic 'noise' would be expected to generate a more plastic evolutionary landscape that facilitates the emergence of fitness-enhancing genetic and epigenetic alterations. To explore the potential relationship between locally disordered methylation and selection, Applicants identified differentially methylated regions (DMRs) in promoters and CGIs, since the presence of recurrent epigenetic alterations might signal the presence of evolutionary convergence. In fact, these DMRs were associated with significantly higher PDR, suggestive of positive selection operating against a backdrop of stochastic epigenetic heterogeneity (FIG. 17A).

[0220] Furthermore, a gene-set enrichment analysis of genes with consistently high promoter PDR across CLL samples compared with genes with consistently low promoter PDR, revealed enrichment in TP53 targets (Perez et al., 2007), in genes differentially methylated across various malignancies (Acevedo et al., 2008; Sato et al., 2003) and in gene-sets associated with stem cell biology (Lim et al., 2010; Wong et al., 2008) (BH-FDR $Q < 0.1$; FIG. 18A, FIG. 17B, Table 4). Finally, regions that are specifically hypomethylated in human embryonic stem cells compared with a diverse collection of differentiated cells (Ziller et al., 2013), also showed decreased methylation and increased PDR in CLL compared to normal B cells, suggestive of a drift towards a more stem-like state (FIG. 18B). Collectively, these findings suggest that locally disordered methylation creates a rich substrate for CLL evolution by stochastic variation amenable to positive selection and by increasing the number of cells that carry the potential to propagate new genotypes to progeny populations. Indeed, CLLs with a higher number of subclonal mutations also exhibit higher PDR ($p=0.002$, FIG. 18C).

[0221] To directly observe the relationship between genetic and epigenetic evolution, Applicants studied RRBS data from 14 longitudinally sampled CLL patients with characterized patterns of genetic evolution (median time between samples 3.45 yrs; 9 CLLs with and 5 without evidence of genetic evolution, Table 5). CLLs that underwent genetic clonal evolution also had increased average promoter PDR over time (paired t-test, $p=0.037$, FIG. 18D), which may indicate a higher PDR in the subclone that expanded over time. In addition, genes with promoters that were demethylated over time, were significantly enriched for the same aforementioned stem cell-related gene-sets (Boquest et al., 2005; Jaatinen et al., 2006; Lim et al., 2010; Wong et al., 2008) (FIG. 18E, Table 6). Importantly, the correlation coefficient between Δ PDR and Δ Meth was markedly lower for gene promoters that were significantly demethylated or hypermethylated over time ($r=0.0937$ and $r=0.0987$, respectively), compared with the correlation coefficient for gene promoters without a significant change in methylation ($r=0.4163$; 144,161 promoters across 14 CLLs). These results suggest that gene promoters with significant changes in methylation over time were enriched with genes that underwent ordered methylation change, as expected from positive selection.

TABLE 4

Gene set enrichments of genes with promoters with consistently high PDR across 104 CLL samples (top 30 enrichments shown).		
Gene Set Name	Q_{high}	$Q_{highvslow}^I$
ZWANG_TRANSIENTLY_UP_BY_2ND_EGF_PULSE_ONLY	5.55E-52	2.89E-28
YOSHIMURA_MARK8_TARGETS_UP	1.73E-43	5.92E-11
ACEVEDO_METHYLATED_IN_LIVER_CANCER_DN	4.91E-22	3.32E-10
LIM_MAMMARY_STEM_CELL_UP	5.42E-38	3.32E-10
SATO_SILENCED_BY_METHYLATION_IN_PANCREATIC_CANCER_1	1.98E-28	2.00E-09
MCBRYAN_PUBERTAL_BREAST_4_5WK_UP	4.44E-25	2.23E-08
DURAND_STROMA_MAX_UP	1.11E-25	2.23E-08
LIU_PROSTATE_CANCER_DN	3.43E-27	2.25E-07
ONDER_CDH1_TARGETS_2_UP	8.01E-21	1.78E-06
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	5.67E-16	5.10E-06
WU_CELL_MIGRATION	1.44E-13	6.49E-05
SMID_BREAST_CANCER_RELAPSE_IN_BONE_DN	2.06E-14	9.04E-05
MIKKELSEN_ES_ICP_WITH_H3K4ME3	1.47E-06	9.04E-05
SERVITJA_ISLET_HNF1A_TARGETS_UP	2.34E-13	1.19E-04
WONG_ADULT_TISSUE_STEM_MODULE	1.52E-36	1.28E-04
ACEVEDO_LIVER_CANCER_WITH_H3K27ME3_DN	5.36E-12	1.65E-04
MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_DN	1.17E-16	1.67E-04
SMID_BREAST_CANCER_LUMINAL_B_DN	1.50E-12	1.71E-04
RIGGI_EWING_SARCOMA_PROGENITOR_UP	3.68E-23	2.14E-04
SCHAEFFER_PROSTATE_DEVELOPMENT_48HR_DN	3.02E-23	2.55E-04
MARTORIATI_MDM4_TARGETS_NEUROEPITHELIUM_DN	9.08E-15	2.73E-04
VART_KSHV_INFECTION_ANGIOGENIC_MARKERS_UP	4.12E-11	2.73E-04
SCHAEFFER_PROSTATE_DEVELOPMENT_48HR_UP	1.90E-22	2.93E-04
KATSANOU_ELAVL1_TARGETS_UP	1.37E-09	3.65E-04
MOHANKUMAR_TLX1_TARGETS_DN	4.84E-11	3.79E-04
PEREZ_TP63_TARGETS	4.83E-25	6.73E-04
BRUINS_UVC_RESPONSE_VIA_TP53_GROUP_A	6.89E-26	6.73E-04
GOZGIT_ESR1_TARGETS_DN	1.42E-25	7.30E-04
YAUCH_HEDGEHOG_SIGNALING_PARACRINE_UP	6.46E-10	1.16E-03
YAUCH_HEDGEHOG_SIGNALING_PARACRINE_DN	2.31E-09	1.33E-03

TABLE 5

Clinical characteristics of 14 patients for whom longitudinal samples were studied.								
CLL IDs	Age	Therapy		IGHV mutation status	Genetic evolution	ZAP70 status	FISH Cytogenetics	Years between samples
		Prior to timepoint 1	Between timepoints 1 & 2					
CLL018	71	None	None	Y	N	-	del(13q)	2.4
CLL020	54	None	None	Y	N	+	del(13q)	2.5
CLL019	52	None	None	Y	Y	-	del(13q)	3.2
CLL030	54	None	None	Y	N	+	del(13q)	3.5
CLL011	41	None	FCR	N	Y	+	del(13q)	5
CLL088	60	None	FCR, Alem + R	N	Y	-	tri12	4.5
CLL169	69	None	FR	Y	Y	+	del(13q)	4.7
CLL167	56	None	FR	Y	Y	-	del(13q), tri12	2.7
CLL016	59	None	FR	N	Y	+	del(13q)	3.4
CLL001	58	None	FR	N	Y	+	del(11q, 13q)	3.5
CLL006	67	FC, Chloram	Alem + R, FR, exp.	N	Y	-	del(13q), del(11q)	4.6
CLL014	65	R	FR	Y	N	-	del(13q)	2.9
CLL066	70	FR, Chloram	R-CVP	Y	N	-	del(13q)	3.5

TABLE 5-continued

Clinical characteristics of 14 patients for whom longitudinal samples were studied.								
CLL IDs	Age	Therapy		IGHV mutation status	Genetic evolution	ZAP70 status	FISH Cytogenetics	Years between samples
		Prior to timepoint 1	Between timepoints 1 & 2					
CLL040	60	FCR	FCR, Alem + R	N	Y	+	del(13q), del (11q)	3

Abbreviations:

Y—Yes,

N—No,

Mut—Mutated,

FISH—Fluorescence In Situ Hybridization,

F—Fludarabine,

C—Cyclophosphamide,

R—Rituximab,

V—Vincristine,

Chloram—Chlorambucil,

Alem—Alemtuzumab;

Rev—Revlimid;

exp—experimental

TABLE 6

Gene set enrichments of genes with significant promoter methylation changes over time (top 30 enrichments for demethylation and methylation are shown).			
Gene Set Name	Q_{high}	Q_{high} vs. genes with Methylation no change change	
MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	1.82E-17	3.42E-22	decrease
BENPORATH_SUZ12_TARGETS	3.76E-15	9.10E-21	decrease
BENPORATH_ES_WITH_H3K27ME3	4.89E-14	4.70E-19	decrease
PEREZ_TP53_TARGETS	2.45E-11	3.96E-17	decrease
ACEVEDO_METHYLATED_IN_LIVER_CANCER_DN	3.16E-08	4.60E-17	decrease
BENPORATH_EED_TARGETS	4.47E-12	7.05E-17	decrease
DODD_NASOPHARYNGEAL_CARCINOMA_UP	5.63E-08	2.73E-16	decrease
MIKKELSEN_MCV6_HCP_WITH_H3K27ME3	6.78E-12	1.06E-14	decrease
SMID_BREAST_CANCER_BASAL_DN	3.88E-09	3.35E-14	decrease
MEISSNER_NPC_HCP_WITH_H3K4ME2_AND_H3K27ME3	6.78E-12	4.87E-14	decrease
MIKKELSEN_MEF_HCP_WITH_H3K27ME3	7.35E-11	5.53E-14	decrease
ZWANG_TRANSIENTLY_UP_BY_2ND_EGF_PULSE_ONLY	9.93E-05	2.06E-13	decrease
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP	6.96E-10	2.83E-13	decrease
WONG_ADULT_TISSUE_STEM_MODULE	6.82E-09	2.89E-13	decrease
BENPORATH_PRC2_TARGETS	6.96E-10	6.13E-13	decrease
MEISSNER_NPC_HCP_WITH_H3K4ME2	3.91E-10	1.28E-12	decrease
LIM_MAMMARY_STEM_CELL_UP	5.58E-09	2.10E-12	decrease
MIKKELSEN_NPC_HCP_WITH_H3K27ME3	6.18E-10	4.62E-12	decrease
GOZGIT_ESR1_TARGETS_DN	1.56E-06	1.56E-11	decrease
ONDER_CDH1_TARGETS_2_UP	1.99E-08	2.55E-11	decrease
CUI_TCF21_TARGETS_2_DN	1.62E-07	6.88E-11	decrease
MEISSNER_BRAIN_HCP_WITH_H3K27ME3	3.98E-08	1.52E-09	decrease
ZWANG_TRANSIENTLY_UP_BY_1ST_EGF_PULSE_ONLY	1.70E-03	1.91E-09	decrease
DACOSTA_UV_RESPONSE_VIA_ERCC3_DN	3.23E-06	8.12E-09	decrease
CHYLA_CBFA2T3_TARGETS_UP	3.83E-05	2.48E-08	decrease
CHEN_METABOLIC_SYNDROM_NETWORK	5.55E-04	3.21E-08	decrease
LEE_BMP2_TARGETS_UP	5.39E-05	5.16E-08	decrease
MEISSNER_NPC_HCP_WITH_H3_UNMETHYLATED	7.13E-06	6.43E-08	decrease
LIU_PROSTATE_CANCER_DN	5.55E-05	7.34E-08	decrease
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_DN	9.93E-05	1.08E-07	decrease
MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	1.82E-17	3.42E-22	decrease
BENPORATH_SUZ12_TARGETS	1.49E-19	2.99E-25	increase
BENPORATH_ES_WITH_H3K27ME3	2.27E-18	2.05E-23	increase
MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	4.03E-18	6.63E-22	increase
BENPORATH_EED_TARGETS	3.55E-15	1.21E-19	increase
MIKKELSEN_MCV6_HCP_WITH_H3K27ME3	5.84E-14	2.80E-16	increase
DODD_NASOPHARYNGEAL_CARCINOMA_UP	6.53E-08	1.40E-15	increase
BENPORATH_PRC2_TARGETS	1.72E-11	3.77E-14	increase
MEISSNER_NPC_HCP_WITH_H3K4ME2	7.06E-12	5.84E-14	increase
MIKKELSEN_MEF_HCP_WITH_H3K27ME3	2.32E-11	6.76E-14	increase
MEISSNER_NPC_HCP_WITH_H3K4ME2_AND_H3K27ME3	4.49E-10	2.21E-11	increase

TABLE 6-continued

Gene set enrichments of genes with significant promoter methylation changes over time (top 30 enrichments for demethylation and methylation are shown).			
Gene Set Name	Q_{high}	Q_{high} vs. genes with Methylation no change	change
ZWANG_TRANSIENTLY_UP_BY_2ND_EGF_PULSE_ONLY	7.57E-04	7.91E-11	increase
MARTENS_TRETINOIN_RESPONSE_UP	1.48E-05	1.04E-10	increase
GOZGIT_ESR1_TARGETS_DN	8.29E-06	6.73E-10	increase
SCHAEFFER_PROSTATE_DEVELOPMENT_48HR_UP	9.59E-07	7.99E-10	increase
MEISSNER_NPC_HCP_WITH_H3_UNMETHYLATED	9.09E-08	8.06E-10	increase
LEE_BMP2_TARGETS_UP	1.01E-06	8.06E-10	increase
CHEMNITZ_RESPONSE_TO_PROSTAGLANDIN_E2_DN	1.71E-06	6.84E-09	increase
ZWANG_TRANSIENTLY_UP_BY_1ST_EGF_PULSE_ONLY	2.06E-03	8.66E-09	increase
MEISSNER_BRAIN_HCP_WITH_H3K27ME3	1.84E-07	1.31E-08	increase
BLALOCK_ALZHEIMERS_DISEASE_UP	1.40E-04	2.97E-08	increase
MIKKELSEN_NPC_HCP_WITH_H3K27ME3	1.01E-06	6.16E-08	increase
SMID_BREAST_CANCER_BASAL_UP	8.09E-05	7.55E-08	increase
GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_UP	7.07E-05	1.05E-07	increase
DAWSON_METHYLATED_IN_LYMPHOMA_TCL1	9.59E-07	1.68E-07	increase
WONG_ENDOMETRIUM_CANCER_DN	1.16E-05	2.55E-07	increase
BRUINS_UVC_RESPONSE_VIA_TP53_GROUP_A	5.02E-04	2.68E-07	increase
GINESTIER_BREAST_CANCER_ZNF217_AMPLIFIED_DN	1.48E-05	3.12E-07	increase
CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_5	3.55E-05	3.62E-07	increase
MARTINEZ_TP53_TARGETS_DN	1.91E-04	9.25E-07	increase
BENPORATH_SUZ12_TARGETS	1.49E-19	2.99E-25	increase
BENPORATH_ES_WITH_H3K27ME3	2.27E-18	2.05E-23	increase

Example 13

Locally Disordered Methylation Impacts Clinical Outcome

[0222] The presented data support a model in which locally disordered DNA methylation facilitates tumor evolution through increased genetic and epigenetic plasticity. Thus, Applicants hypothesized that increased PDR would be associated with a shorter remission time after treatment, which was previously linked with clonal evolution (Landau et al., 2013).

[0223] Applicants therefore examined failure-free survival after treatment (FFS, failure defined as retreatment or death) in 49 patients included in the cohort that were treated after tumor sampling for RRBS. A higher mean sample promoter PDR (>mean for cohort) was significantly associated with shorter FFS (median FFS of 16.5 vs. 44 months, hazard ratio=2.5 [95% CI: 1.1-5.7], $p=0.028$, FIG. 19A; 52% and 65% of patients, respectively, were treated with fludarabine based immunochemotherapy, $p=0.39$). A regression model including established CLL risk indicators (IGHV unmutated status, del(17p) and del(11q)) showed an adjusted hazard ratio of 2.81 (95% CI 1.05-7.53, $p=0.039$, FIG. 19B) for high promoter PDR. Similar results were obtained after inclusion of additional variables in the model including mutation burden and average promoter methylation (Table

7). Samples with higher promoter PDR were also more likely to have a subclonal driver mutation as previously defined (Landau et al., 2013) ($p=0.01$). When the presence of a subclonal driver was added to the regression model, the increased risk associated with the elevated PDR was no longer preserved (FIG. 19B). These results support the notion that epigenetic ‘noise’ may function primarily as a facilitating feature, allowing the emergence of subclonal drivers, which then contribute to the adverse clinical outcome.

[0224] A further extension of this model proposes that locally disorder methylation enhances the evolutionary capacity of CLL by optimizing the process of genetic diversification. This framework would necessitate coincidence of a novel somatic mutation with an epigenetic state permissive to the propagation of the new genotype to a progeny population. In cellular populations with a preserved epigenetic landscape (FIG. 20—top), the proportion of cells capable of actively participating in the evolutionary process is predicted to be small. On the other hand, in a more malleable epigenetic landscape (FIG. 20—bottom) as is expected with a high level of locally disordered methylation, a greater proportion of cells can give birth to new subclones. This process would accelerate genetic evolution, provide a greater adaptive capacity for the cancer population and result in adverse clinical outcome with therapy, as was saw in the CLL cohort (FIG. 20).

TABLE 7

Stepwise regression model for prediction of clinical outcome.			
	Unadjusted HR [95% CI]	Stepwise selection Final model (without subclonal driver) HR [95% CI]	Stepwise selection Final model (including subclonal driver as candidate) HR [95% CI]
Promoter PDR: cutpoint at the Mean >0.1033 vs. ≤ 0.1033	2.51 [1.10-5.17] $p = 0.029$	3.48 [1.37-8.86] $p = 0.009$	

TABLE 7-continued

Stepwise regression model for prediction of clinical outcome.			
	Unadjusted HR [95% CI]	Stepwise selection Final model (without subclonal driver) HR [95% CI]	Stepwise selection Final model (including subclonal driver as candidate) HR [95% CI]
IGHV Mutated vs. Unmutated	0.29 [0.11-0.77] p = 0.013	0.16 [0.05-0.47] p = 0.0009	0.20 [0.07-0.58] p = 0.003
Presence of del11q	1.26 [0.55-2.86] p = 0.58		
Presence of del17p	3.46 [1.39-8.62] p = 0.008	2.51 [0.84-7.51] p = 0.10	3.24 [0.99-10.54] p = 0.051
Presence of a subclonal driver	4.80 [1.79-12.92] p = 0.002	NA	6.54 [2.16-19.86] p = 0.0009
Promoter methylation: cutpoint at mean >0.0735 vs. ≤0.0735	1.81 [0.83-3.99] p = 0.14		
Mutation number: cutpoint at mean >18.8 vs. ≤18.8	1.89 [0.85-4.23] p = 0.012	2.57 [1.04-6.35] p = 0.040	3.42 [1.39-8.39] p = 0.007

Example 14

Excluding Alternative Explanations for High PDR
Other than Locally Disordered Methylation

[0225] Applicants considered several possible alternative explanations to these findings. First, the contaminating non-malignant cell fraction of samples may contribute to the PDR, even though the overall purity of the CLL samples was consistently high (90.2% median purity). However, when Applicants compared samples with purity above and below the overall average (86.6%), PDR was higher in the former (mean±SEM, 0.2259±0.0047 vs. 0.2062±0.0066, t-test p=0.009), indicating that indeed the malignant cells in the samples contribute to the high PDR (FIG. 4I). Second, Applicants considered the possibility that elevated PDR may affect only one allele in the sample as part of allele-specific methylation (ASM). To test this, Applicants identified germline SNPs that did not involve CpGs across 53 randomly selected CLL samples in the cohort. Of these germline SNPs, 4,486 had equivalent coverage of both genotypes in the RRBS reads (ratio of 0.4-0.6 in variant reads/total reads). At these sites, discordant reads were found to contain both alternative genotypes in an increasing proportion of SNPs in association with an increased total number of discordant reads per locus (FIG. 4J), converging towards 1. This result demonstrates that locally disordered methylation likely affects both parental alleles. Furthermore, even within a given genotype different discordancy patterns were seen (FIG. 4K), revealing that high PDR results indeed from locally disordered methylation and not simply from allele-specific methylation patterns. In this context, it is important to note that X/Y chromosomes were excluded from the entire analysis.

[0226] In addition to the germline variants, Applicants carried out a similar analysis with regards to somatic single nucleotide mutations, by integrating WGS and WGBS data for CLL007 and CLL169. After excluding C>T mutations, and limiting the analysis to regions with >4 CpGs per read on average (to ensure accurate estimation of PDR) and to mutations with >20× coverage in the WGS (to ensure

accurately distinguishing clonal vs. subclonal events), Applicants identified 52 and 66 high confidence mutations for analysis, respectively (91% and 79% of these mutations were either intronic or intergenic mutations in CLL007 and CLL169, respectively). The correlation between the average methylation values of the clonally mutated alleles and the matching germline alleles was high (CLL169—number of clonal mutations evaluated=30, r=0.96, p=1.9×10⁻¹⁷, CLL007—number of clonal mutations evaluated=10, r=0.94, p=3.6×10⁻⁵). Similarly, the correlation between the PDR of the clonally mutated alleles and the matched germline alleles was also high (CLL169: r=0.72, p=5.6×10⁻⁷; CLL007: r=0.65, p=0.04). While the correlation of average methylation values remained high between the mutated alleles and the matched germline alleles for subclonal mutations (CLL169—number of subclonal mutations evaluated=36, r=0.47, p=0.008, CLL007—number of subclonal mutations evaluated=42, r=0.81, p=5.3×10⁻¹¹), the correlation between the PDR values of the two alleles was lower (r=0.09 and 0.45, p=0.5 and p=0.002, respectively), with a trend towards higher PDR in the mutated subclonal allele (20.5% and 34.6% increase in PDR in mutated alleles, for CLL169 and CLL007, respectively, with p=0.2 and 0.048). Collectively, these data show that disordered methylation involved both the mutant and germline alleles, with a trend towards higher PDR in subclonally mutated alleles.

[0227] Moreover, if high PDR results from ASM, then it would be expected to find predominately 1 or 2 consistent patterns of discordancy, across all reads covered for a particular locus. However, a histogram of the number of distinct discordancy pattern in loci that have a significant number of discordant reads (10-20) across ten randomly selected CLL samples, shows a normal distribution centered at 5 discordant patterns, consistent with a model of stochastic disorder rather than ASM (FIG. 4L). This latter finding also confirms that most of the PDR does not result from reads that cover an ordered transition point from one methylation state to another, which is also expected to yield 1 recurrent discordancy pattern.

[0228] Another potential explanation for increased PDR could be related to methQTL (Gibbs et al., 2010). This is

unlikely to account for the genome-wide pervasive process described for the following reasons: i) this effect is expected to be of importance in a tumor with a high mutation load. However, CLL is a malignancy with one of the lowest mutational loads, 1000-2000 mutations per genome (Wang et al., 2011). Extrapolating from the study by Gibbs et al., which evaluated ~1.5M germline SNPs and only found association with 4-5% of CpGs, the mutational load in CLL at best will only affect 0.005% of CpGs. This is expected to have a small effect in comparison to the pervasive disorder in methylation patterns (e.g., in CLL169 WGBS, 73.39% of CpGs have PDR >0.1). ii) Cancer cell lines, which harbor 1-3 orders of magnitude more somatic mutations than primary CLLs, harbor marginally higher rates of PDR. iii) Finally, the PDR pattern would more likely result from methQTLs of subclonal mutations, as clonal mutations would behave largely like germline SNPs and therefore are unlikely to result in increase in PDR in cancer vs. normal tissue, given their number in the CLL genome. To assess for the confounding effect of methQTL on PDR, which may be related to subclonal mutations, Applicants compared the correlation to PDR between clonal mutations and subclonal mutations and found that the distance from clonal mutations shows a stronger negative correlation to PDR, compared to the distance from subclonal mutations (FIG. 12B-C). Although methQTL may have long-range effects, at least a third supposedly act in cis (defined in Gibbs et al., as <1 MB). These results, therefore, are not consistent with a significant impact of methQTL.

[0229] Finally, technical artifacts were also considered as a potential cause of locally disordered methylation. Incomplete bisulfite conversion is an unlikely explanation for these findings as bisulfite conversion rates were high in both CLL and normal B cell samples (average of 99.66% and 99.72%, respectively) as measured by the rate of unmethylated cytosines in a non-CpG context (Bock et al., 2005). Furthermore, incomplete conversion is expected to decrease PDR preferentially in highly methylated region, however, Applicants observed an increase in PDR in CLLs in regions with both low and high methylation.

[0230] PCR amplification biases in the RRBS procedure are not likely to contribute significantly to this result. First, Applicants have no reason to expect differential impact on CLL samples and normal B cells. Second, the consistency of the finding in WGBS where duplicate reads were discarded makes this technical bias an unlikely source for locally disordered methylation. Indeed the Pearson's correlation of PDR in promoter CpGs covered by both RRBS and WGBS at >30x was high (CLL169; $r=0.856$, CLL007; $r=0.855$, and Normal_IGD_3; $r=0.737$). Finally, given that there is no reason to expect duplicate reads to affect concordant reads less than discordant reads, duplicate reads are expected to decrease PDR, as the overall number of concordant reads is higher than discordant reads (87.1±2% of RRBS reads evaluated are concordant, evaluated in randomly selected 5 samples (CLL003, CLL005, CLL006_TP1, CLL001_TP1 and CLL001_TP2). To quantify PCR amplification biases, Applicants measured the ratio of reads for each of the heterozygous SNP and found a similar representation of both parental alleles (FIG. 4M). In addition, measured methylation values for germline Imprinted Control Regions (ICRs) (Woodfine et al., 2011) and found that these loci approximated 50% methylation, as expected (FIG. 4N).

[0231] Finally, although CLL genomes are mostly diploid (Brown et al., 2012), and therefore the analysis is not expected to be significantly impacted by somatic copy number variations (sCNV), Applicants examined the PDR in regions of sCNV in WGBS of CLL007 and CLL169. Altogether in these tumors, 4 sCNVs were detected (using SNP array analysis as described previously (Landau et al., 2013)). As shown in FIG. 4O, both the overall PDR and the promoter PDR do not differ substantially in the sCNVs compared to the remainder of the genome.

Methods

[0232] Sample Acquisition.

[0233] Heparinized blood samples were obtained from patients and healthy adult volunteers enrolled on clinical research protocols at the Dana-Farber/Harvard Cancer Center (DF/HCC), approved by the DF/HCC Human Subjects Protection Committee. The diagnosis of CLL according to WHO criteria was confirmed in all cases by flow cytometry, or by lymph node or bone marrow biopsy. Peripheral blood mononuclear cells (PBMC) from normal donors and patients were isolated by Ficoll/Hypaque density gradient centrifugation. Mononuclear cells were cryopreserved with FBS/10% DMSO and stored in vapor-phase liquid nitrogen until the time of analysis. The patients included in the cohort represent the broad clinical spectrum of CLL (Table 1). Informed consent on DFCI IRB-approved protocols for genomic sequencing of patients' samples was obtained prior to the initiation of sequencing studies.

[0234] Reduced Representation Bisulfite Sequencing (RRBS).

[0235] Genomic DNA from CLL samples, normal B cell samples and cancer cell line samples were used to produce RRBS libraries. These were generated by digesting genomic DNA with MspI to enrich for CpG-rich fragments, and then were ligated to barcoded TruSeq adapters (Illumina) to allow immediate subsequent pooling. This was followed by bisulfite conversion and PCR, as previously described (Boyle et al., 2012). Libraries were sequenced and 29mers were aligned to the hg19 genome using MAQ version 0.6.6 (Li et al., 2008). Reads were further filtered if: i) The read did not align to an autosome, ii) The read failed platform/vendor quality checks (samtools flag 0x200), and/or iii) the read did not align to an MspI cut site. The methylation state of each CpG was determined by comparing bisulfite-treated reads aligning to that CpG with the genomic reference sequence. The methylation level was computed by dividing the number of observed methylated cytosines (which did not undergo bisulfite conversion) by the total number of reads aligned to that CpG (FIG. 3E). In addition, the number of CpG measurements on each read was noted. In order to identify locations in the genome where concordant methylation (in either methylated or unmethylated states) occurs, Applicants devised a measure called the Proportion of Discordant Reads (PDR). This measure can be computed for a specific genomic location or for the entire genome. After reads are aligned to the reference genome, the methylation state of each CpG on a read is determined. If all the CpGs on a specific read are methylated, or all of the CpGs on a read are unmethylated, the read is classified as concordant; otherwise it is classified as discordant. At each CpG, the PDR is equal to the number of discordant reads that cover that location divided by the total number of reads that cover that location (FIG. 3E). The PDR across the entire genome

or for a specified genomic region is given by averaging the values of individual CpGs, as calculated for all CpGs within the region of interest with read depth greater than 10 reads and that are covered by reads that contain at least 4 CpGs. It is important to note that PDR and variances were also calculated with means weighted by depth of coverage of a particular CpG with consistently similar results. For example, overall variance weighted by the number of read depth per CpG shows similar difference in variance of 0.0696 [0.0679-0.0714] for CLL samples, vs. 0.0437 [0.0399-0.0475] for normal B cell samples ($p=2.61 \times 10^{-13}$). Weighted average of PDR for CLL samples was 0.2476 [0.2431-0.2520] vs. 0.1402 [0.1275-0.1528] for normal B cell samples ($p=1.06 \times 10^{-14}$). The CLL and normal B cell RRBS raw data are deposited in dbGaP (phs000435.v2.p1), and processed data format files containing PDR and methylation values for each CpG evaluated in the CLL and normal B cell samples are deposited in GEO (GSE58889). RRBS of primary diverse human tissue samples were previously reported (<http://www.roadmapepigenomics.org>). Reads were realigned and methylation was determined using identical protocols to the rest of the samples.

[0236] Whole Genome Bisulfite Sequencing (WGBS).

[0237] Genomic DNA was fragmented to 100-500 bp fragments using a Covaris S2 sonicator (Woburn, Mass.). DNA fragments were cleaned-up, end-repaired, A-tailed and ligated with methylated paired-end adapters (from ATDBio, Southampton, UK). Libraries were sequenced and WGBS reads were aligned using BSMAP version 2.7 (Xi and Li, 2009) to the hg19/GRCh37 reference assembly. Subsequently, CpG methylation calls were made using custom software, excluding duplicate, low-quality reads, as well as reads with more than 10% mismatches. Applicants note that as previously reported (Kulis et al., 2012), non-CpG methylation levels were minimal (0.08% in both CLL samples). Only CpGs covered by >10 reads were considered for further analysis. A methylation-calling pipeline was implemented in Perl and determines CpG methylation state by observing bisulfite conversion at read locations aligned to a CpG in the reference genome. Previously published WGBS data for 2 CLL samples and 3 normal B cell samples (Kulis et al., 2012) were downloaded with permission from the European Genome-Phenome Archive. The raw sequencing reads were processed in identical fashion to the in-house produced WGBS libraries. Additional processing steps for WGBS reads included trimming by 4 bp to ensure high data quality, and filtering out reads that: i) did not align to an autosome, ii) failed platform/vendor quality checks (samtools flag 0x200), iii) had poor alignment score (samtools flag 0x2), iv) had poor alignment of the read mate (samtools flag 0x8), v) aligned to the same location as another read (read duplicate), or vi) contained nucleotides at a CpG location that could not have been produced by bisulfite conversion. The determination of the concordant vs. discordant classification was performed in identical fashion as with RRBS reads. The CLL and normal B cell WGBS data are deposited in dbGaP (phs000435.v2.p1), and processed data format files containing PDR and methylation values for each CpG evaluated in the sample are deposited in GEO (GSE58889).

[0238] RNA-Sequencing.

[0239] RNA-sequencing of CLL and normal B cell samples was performed as previously described (Landau et al., 2013). For single cell RNAseq, the C1 Single-Cell Auto

Prep System (Fluidigm, San Francisco, Calif.) was used to perform SMARTer (Clontech, Mountain View, Calif.) whole transcriptome amplification (WTA), on up to 96 individual cells per sample from 4 primary CLL patient samples. WTA products were then converted to Illumina sequencing libraries using Nextera XT (Illumina, San Diego, Calif.) (Ramskold et al., 2012).

[0240] Statistical Analysis.

[0241] Statistical analysis was performed with MATLAB (MathWorks, Natick, Mass.), R version 2.15.2 and SAS version 9.2 (SAS Institute, Cary, N.C.). Categorical variables were compared using the Fisher Exact test, and continuous variables were compared using the Student's t-test, Wilcoxon rank sum test, or Kruskal-Wallis test as appropriate. Linear modeling for expression as a predicted variable, based on methylation and PDR was performed using built in R linear model function. FFS (failure-free survival from first treatment after sampling) was defined as the time to the 2nd treatment or death from the 1st treatment following sampling, was calculated only for those patients who had a 1st treatment after the sample and was censored at the date of last contact for those who had only one treatment after the sample, and estimated using the method of Kaplan and Meier. The difference between groups was assessed using the log-rank test. Unadjusted and adjusted Cox modeling was performed to assess the impact of established CLL high-risk predictors and the presence of a subclonal driver. Models were adjusted for known prognostic factors including the presence of a 17p deletion, the presence of a 11q deletion and IGHV mutational status. Cytogenetic abnormalities were primarily assessed by FISH; if FISH was unavailable, genomic data were used. For unknown IGHV mutational status an indicator was included in adjusted modeling and was not found to be significant. Similarly, unadjusted and adjusted Cox modeling was performed to assess the impact of mutational burden and average promoter methylation in addition to established CLL prognostic factors. Given the large number of potential variables, a stepwise selection procedure was used to determine a final multivariable model considering all factors listed above. All p-values are two-sided and considered significant at the 0.05 level unless otherwise noted. The CLL and normal B cell sequencing data were deposited in dbGaP (phs000435.v2.p1), and the processed data deposited in GEO (GSE58889).

[0242] Established CLL Prognostic Factor Analysis.

[0243] Immunoglobulin heavy-chain variable (IGHV) homology (unmutated was defined as greater than or equal to 98% homology to the closest germline match) and ZAP-70 expression (high risk defined as >20% positive) were determined (Rassenti et al., 2008). Cytogenetics were evaluated by FISH for the most common CLL abnormalities (del(13q), trisomy 12, del(11q), del(17p), all probes from Vysis, Des Plaines, Ill., performed at the Brigham and Women's Hospital Cytogenetics Laboratory, Boston Mass.). Samples were scored positive for a chromosomal aberration based on consensus cytogenetic scoring (Smoley et al., 2010).

[0244] DNA Isolation from CLL and Normal B-Cell Subpopulations.

[0245] Genomic DNA was extracted from CLL cells or normal B cell populations utilizing the ROCHE DNA Isolation Kit (Roche Applied Science, Indianapolis, Ind.). Control CD19⁺ B cell samples were isolated from buffy coats of healthy adult volunteers using a two-step enrichment pro-

cedure. B cells were first enriched using the RosetteSep Human B cell Enrichment System (StemCell Technologies Inc., Vancouver, British Columbia, Canada) and then further purified by immunomagnetic bead selection (CD19⁺ beads, Miltenyi Biotec, Cambridge, Mass.). From these purified CD19⁺ cells, naive B cells (CD19⁺CD27-IgD⁺) and memory B cells (CD19⁺CD27-IgD⁻) were isolated by flow cytometric sorting (FACSaria II, BD Biosciences) using CD27-PC5 (Beckman Coulter, Brea, Calif.) and IgD-CY7 (Biolegend, San Diego, Calif.) antibodies. Standard protocols for DNA quality control for genomic studies were applied, as recently described (Berger et al., 2011; Chapman et al., 2011; Landau et al., 2013).

[0246] Reanalysis of Whole-Exome DNA Sequencing (WES) Data from CLL Samples.

[0247] Applicants re-analyzed WES from 104 of 160 previously reported CLLs and their matched germline samples (Landau et al., 2013), deposited in dbGaP (phs000435.v2.p1). Details of whole-exome library construction and analysis have been detailed elsewhere (Fisher et al., 2011; Landau et al., 2013). Briefly, output from Illumina software (Illumina, San Diego, Calif.) was processed by the “Picard” data processing pipeline to yield BAM files containing aligned reads with well-calibrated quality scores (Chapman et al., 2011; DePristo et al., 2011). Somatic alterations were identified using a set of tools within the “Firehose” pipeline, developed at the Broad Institute (www.broadinstitute.org/cancer/cga) (Berger et al., 2011; Chapman et al., 2011). Somatic single nucleotide variations (sSNVs) were detected using MuTect (Cibulskis et al., 2013). Applicants used the ABSOLUTE algorithm to calculate the purity, ploidy, and absolute DNA copy-numbers of each sample (Carter et al., 2012) and clonal/sub-clonal status of each alteration inferred using a probabilistic approach (Escobar and West, 1995; Landau et al., 2013). Applicants note that the spectrum of mutations in these samples was consistent with prior publications (Quesada et al., 2012), with C>T transitions constituting the most frequent sSNVs (average of 41.8±15% of all sSNV across all 104 CLL WES analyzed in this study). There was no significant correlation between the proportion per sample of any specific subtype of sSNV and PDR ($-0.1 < r < 0.1$, $p > 0.3$).

[0248] Whole Genome Sequencing of CLL Sample CLL169 and CLL007.

[0249] Library construction was performed using 1-3 micrograms of native DNA from primary tumor (peripheral blood) and germline (saliva) samples. The DNA was sheared to a range of 101-700 bp using the Covaris E210 Instrument and was then phosphorylated and adenylated according to the Illumina protocol. Adaptor ligated purification was done by preparatory gel electrophoresis, and size was selected by excision of two bands (500-520 bp and 520-540 bp, respectively), yielding two libraries per sample with average of 380 bp and 400 bp, respectively. The libraries were then sequenced with the Illumina GA-II or Illumina HiSeq sequencer with 76 or 101 bp reads, achieving an average of ~30× coverage depth. The resulting data were analyzed with the current Illumina pipeline, which generates data files (BAM files) that contain the reads and quality parameters. Sequencing data are available in the dbGaP database (<http://www.ncbi.nlm.nih.gov/gap>) under accession number phs000435.v2.p1. Somatic single nucleotide variations (sSNVs) were detected using MuTect (Cibulskis et al., 2013). Replication times were adopted from Chen et al. (Chen et

al., 2010). S50 values (for a defined genome region, S50 corresponds to the fraction of the S phase at which 50% of the sequence reads that map in this region were obtained) were rescaled to vary from 100 (early) to 1000 (late) as previously described (Lawrence et al., 2013). Although replication times reported by Chen et al., were not measured directly in CLL cells or B cells, previous studies have shown that replication time is fairly consistent across different cell types (Karnani et al., 2007). Furthermore, Chen and colleagues confirmed a high correlation with previously measured replication time in other cell types including human lymphocytes.

[0250] RNA-Sequencing of CLL Samples and Analysis.

[0251] 5 µg of total RNA was poly-A selected using oligo-dT beads to extract the desired mRNA, and used to construct dUTP libraries as previously described (Landau et al., 2013). Samples were pooled and sequenced using either 76 or 101 bp paired end reads. RNAseq BAMs were aligned to the hg19 genome using the TopHat suite. FPKM values were generated with the Cufflinks suite (<http://cufflinks.cbc.bcm.edu/>). These data are deposited in dbGaP (phs000435.v2.p1).

[0252] Methylation Array Analysis.

[0253] Data for previously published 450K methylation arrays (Kulis et al., 2012) were downloaded with permission from the European Genome-Phenome Archive. Data from the 450 k Human Methylation Array were analyzed by GenomeStudio (Illumina) and R using the lumi package available through Bioconductor.

[0254] Single Cell RNA-Sequencing of CLL Samples:

[0255] Four primary cryopreserved peripheral blood CLL samples were thawed and stained with anti-CD19 FITC and anti-CD5 PE antibodies (Beckman Coulter, Indianapolis, Ind.). 7-AAD (Invitrogen, Grand Island, N.Y.) was added before FACS sorting as a viability control. Live CD19⁺ CD5⁺ tumor cells were preliminarily sorted into a collection tube. Subsequently, the bulk cell concentration was adjusted to 250 cell/µl and applied to the C1 Single-Cell Auto Prep System for single cell capture with a 5-10 micron chip (Fluidigm, San Francisco, Calif.). The capture rate was measured at >80%. Following capture, whole transcriptome amplification (WTA) was immediately performed using the C1 Single-Cell Auto Prep System with the SMARTer Kit (Clontech, Mountain View, Calif.) on up to 96 individual cells. The C1 WTA products were then converted to Illumina sequencing libraries using Nextera XT (Illumina). RNA-Seq was performed on a MiSeq instrument (Illumina).

[0256] Analysis of Single-Cell RNA-Seq Data.

[0257] Paired-ended reads were aligned against UCSC hg19 human annotation (Mar. 6, 2013 version) using Tophat 2.0.10 (Kim et al., 2013), and read counts for each gene were determined using HTSeq 0.5.4 (Anders et al., 2014). A subset of cells with more than 10,000 total reads across all genes was selected for further analysis (73-87% of cells). To determine population average gene expression (performed separately for each of the 4 primary CLL samples), the read counts observed in each cell were normalized by the effective library size, determined by edgeR (Robinson et al., 2010) ‘calcNormFactors’ method.

[0258] To test for significance of association of PDR with expression heterogeneity, first the fraction of positive cells (fpc) was calculated per gene (a cell is defined as positive if >0 reads aligned to the gene). Subsequently, Shannon’s information entropy (ent) was calculated $ent = -1 \times (fpc \times \log$

$2(\text{fpc}) + (1 - \text{fpc}) \times \log_2(1 - \text{fpc})]$. The association with PDR was tested using generalized additive models (implemented by gam R package). The following types of models were tested:

$\text{ent} \sim s(\text{population average expression}) + \text{PDR} + \text{transcript length}$

$\text{ent} \sim s(\text{population average expression}) + \text{PDR} + \text{transcript length} + \text{methylation}$

where $s()$ indicates local regression. The population average expression values were entered into the models on $\log_{10}$ scale (adding 1).

[0259] Genome Annotations Definitions.

[0260] Promoters were defined as 1 Kb upstream and 1 Kb downstream of hg19 Refgene gene transcription start sites (TSSs). The set of CpG Islands (CGIs) were defined using biologically-verified CGIs (Illingworth et al., 2010). Enhancer regions were defined as the union of the 'Distal Regulatory Modules' class from all cell types as previously identified (Ramskold et al., 2012). CTCF binding sites were annotated based on published CTCF binding ChIP-seq experiments using 27 healthy donor transformed B cells ChIP-seq experiments (Wang et al., 2012). Applicants curated a list of CTCF binding sites based on sites that were detected in at least 75% of these B cell samples, and then calculated the CTCF binding site per megabase across the human genome. The location of repeat elements was identified based on the RepBase database version 18.09 for hg19 (<http://www.girinst.org/server/archive/RepBase18.09/>).

Hypomethylated regions in embryonic stem cells were defined as previously described (Ziller et al., 2013), and the analysis was limited to regions with at least 20 CpGs. Differentially-methylated regions (DMRs) were called using a two-sample t-test with significance of $p < 0.01$ and in which the difference between the weighted average region methylation levels was greater than 10%. Well-covered regions with at least 5 CpGs in at least 80% of the samples were used for the analysis, as previously described (Bock et al., 2011).

Summary Statistics of Methylation Across DMRs Between CLL and Normal B Cells.

[0261]

DMR category	Mean methylation		Standard deviation		Number of elements
	CLL	Normal B cells	CLL	Normal B cells	
Promoters	3.06E-01	1.71E-01	1.51E-01	1.40E-01	213
hypermethylated in IGHV mutated CLL					
Promoters hypomethylated in IGHV mutated CLL	4.95E-01	6.40E-01	2.11E-01	2.02E-01	28
CGIs hypomethylated in IGHV mutated CLL	4.01E-01	5.54E-01	3.08E-01	2.97E-01	41
CGIs hypomethylated in IGHV mutated CLL	3.27E-01	1.87E-01	1.61E-01	1.51E-01	558
CGIs hypomethylated in IGHV unmutated CLL	4.70E-01	6.69E-01	3.15E-01	2.81E-01	115
CGIs hypermethylated in IGHV unmutated CLL	2.84E-01	1.24E-01	1.63E-01	1.55E-01	817
Promoters hypomethylated in IGHV unmutated CLL	5.57E-01	7.11E-01	2.55E-01	2.56E-01	145
Promoters hypermethylated in IGHV unmutated CLL	2.59E-01	1.02E-01	1.53E-01	1.42E-01	332

[0262] Modeling Locally Disordered Methylation.

[0263] In order to describe the expected PDR for a given set of reads covering the same set of CpGs, Applicants developed a model to describe the likelihood of finding a certain number of discordant reads, given a methylation value for the set of reads. The input parameters for the model were the number of CpGs covered by the reads, the average methylation value of the covered CpGs, and the number of reads covering the CpGs. Applicants modeled the methylation state of each CpG on each read as an independent Bernoulli trial, with the probability of getting a methylated CpG being set to the overall empirical methylation average. The probability of seeing a specified number of discordant reads was then unity minus the probability of observing a specified number of concordant reads (a probability derived directly from the independent Bernoulli trials for each CpG).

[0264] Using this model, Applicants were able to predict the maximum likelihood for PDR for a set of reads covering a certain number of CpGs, with a certain methylation value. In addition to finding the maximum likelihood PDR, Applicants were able to assign a P-value for the probability of finding a specified number of discordant reads, given the number of CpGs covered by the reads, the average methylation value, and the total number of reads. Applicants plotted the 99% confidence interval using this model in FIG. 10A.

[0265] Germline Variants Detection for Allele-Specific Analyses.

[0266] Germline variants were detected using the UnifiedGenotyper in the Genome Analysis Toolkit (<http://www.broadinstitute.org/gatk/>), using default options, followed by the filtering of SNPs using Variant Quality Score Recalibration, and hard-filtering of indels (DePristo et al., 2011; McKenna et al., 2010). Germline variants were annotated using SeattleSeq137 (<http://snp.gs.washington.edu/SeattleSeqAnnotation137/>).

[0267] Gene Set Enrichment Analysis.

[0268] Gene set enrichment analysis was limited to the C2 gene set collection (Subramanian et al., 2005). To assess gene set enrichments in genes that exhibit consistently elevated PDR (greater than mean promoter PDR of 0.1 in >75% of 104 CLL samples) a Fisher's exact test was used to measure the enrichment of these genes in each gene-set,

followed by a Benjamini-Hochberg FDR procedure. Similarly, to compare enrichments between the set of genes with high promoter PDR and low promoter PDR (less than mean promoter PDR of 0.1 in >75% of 104 CLL samples), a Fisher's exact test was used, followed by a Benjamini-Hochberg FDR procedure. This latter procedure was done to avoid potential biases related to the CpG content of different promoters as previously described. By comparing enrichments of two gene sets both covered by RRBS, these biases are likely to have minimal impact. A similar procedure was undertaken for gene set enrichment analysis of genes with significant change in methylation in the longitudinal samples ($Q < 0.1$). By comparing these gene-sets with genes that did not have a significant change in methylation ($Q > 0.2$), Applicants were able to assess the gene set enrichment while limiting the impact of biases related to CpG content of different gene promoters.

REFERENCES

- [0269] Acevedo, L. G., Bieda, M., Green, R., and Farnham, P. J. (2008). Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors. *Cancer Res.* 68, 2641-2651.
- [0270] Akiyama, Y., Watkins, N., Suzuki, H., Jair, K. W., van Engeland, M., Esteller, M., Sakai, H., Ren, C. Y., Yuasa, Y., Herman, J. G., et al. (2003). GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Mol Cell Biol.* 23, 8429-8439.
- [0271] Anders, S., Pyl, P. T., and Huber, W. (2014). HTSeq—A Python framework to work with high-throughput sequencing data. *bioRxiv*.
- [0272] Balazsi, G., van Oudenaarden, A., and Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell.* 144, 910-925.
- [0273] Baylin, S. B. (2005). DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol.* 2 Suppl 1, S4-11.
- [0274] Baylin, S. B., and Jones, P. A. (2011). A decade of exploring the cancer epigenome—biological and translational implications. *Nat Rev Cancer.* 11, 726-734.
- [0275] Berger, M. F., Lawrence, M. S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A. Y., Sboner, A., Esqueva, R., Pflueger, D., Sougnez, C., et al. (2011). The genomic complexity of primary human prostate cancer. *Nature.* 470, 214-220.
- [0276] Berman, B. P., Weisenberger, D. J., Aman, J. F., Hinoue, T., Ramjan, Z., Liu, Y., Nourshahr, H., Lange, C. P., van Dijk, C. M., Tollenaar, R. A., et al. (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet.* 44, 40-46.
- [0277] Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6-21.
- [0278] Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z. D., Ziller, M., Croft, G. F., Amoroso, M. W., Oakley, D. H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell.* 144, 439-452.
- [0279] Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics.* 21, 4067-4068.
- [0280] Boquest, A. C., Shahdadfar, A., Fronsdaal, K., Sigurjonsson, O., Tunheim, S. H., Collas, P., and Brinckmann, J. E. (2005). Isolation and transcription profiling of purified uncultured human stromal stem cells: alteration of gene expression after in vitro cell culture. *Mol Biol Cell.* 16, 1131-1141.
- [0281] Boyle, P., Clement, K., Gu, H., Smith, Z. D., Ziller, M., Fostel, J. L., Holmes, L., Meldrum, J., Kelley, F., Gnirke, A., et al. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol.* 13, R92.
- [0282] Brown, J. R., Hanna, M., Tesar, B., Werner, L., Pochet, N., Asara, J. M., Wang, Y. E., Dal Cin, P., Fernandes, S. M., Thompson, C., et al. (2012). Integrative genomic analysis implicates gain of PIK3CA at 3q26 and MYC at 8q24 in chronic lymphocytic leukemia. *Clin Cancer Res.* 18, 3791-3802.
- [0283] Cahill, N., Bergh, A. C., Kanduri, M., Goransson-Kultima, H., Mansouri, L., Isaksson, A., Ryan, F., Smedby, K. E., Juliusson, G., Sundstrom, C., et al. (2013). 450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. *Leukemia.* 27, 150-158.
- [0284] Carter, S. L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P. W., Onofrio, R. C., Winckler, W., Weir, B. A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 30, 413-421.
- [0285] Chapman, M. A., Lawrence, M. S., Keats, J. J., Cibulskis, K., Sougnez, C., Schinzel, A. C., Harview, C. L., Brunet, J. P., Ahmann, G. J., Adli, M., et al. (2011). Initial genome sequencing and analysis of multiple myeloma. *Nature.* 471, 467-472.
- [0286] Chen, C. L., Rappailles, A., Duquenne, L., Huvet, M., Guilbaud, G., Farinelli, L., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., Hyrien, O., et al. (2010). Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.* 20, 447-457.
- [0287] Chim, C., Pang, R., and Liang, R. (2008). Epigenetic dysregulation of the Wnt signalling pathway in chronic lymphocytic leukaemia. *J Clin Pathol.* 61, 1214-1219.
- [0288] Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyer, M., Lander, E. S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 31, 213-219.
- [0289] De, S., Shaknovich, R., Riester, M., Elemento, O., Geng, H., Kormaksson, M., Jiang, Y., Woolcock, B., Johnson, N., Polo, J. M., et al. (2013). Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet.* 9, e1003137.
- [0290] DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43, 491-498.
- [0291] Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R.,

- Down, T. A., et al. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet.* 38, 1378-1385.
- [0292] Ehrlich, M. (2009). DNA hypomethylation in cancer cells. *Epigenomics.* 1, 239-259.
- [0293] Escobar, M., and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association.* 90, 577-588.
- [0294] Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T. M., Young, G., Fennell, T. J., Allen, A., Ambrogio, L., et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 12, R1.
- [0295] Gibbs, J. R., van der Brug, M. P., Hernandez, D. G., Traynor, B. J., Nalls, M. A., Lai, S. L., Arepalli, S., Dillman, A., Rafferty, I. P., Troncoso, J., et al. (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS genetics.* 6, e1000952.
- [0296] Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell.* 144, 646-674.
- [0297] Hansen, K. D., Timp, W., Bravo, H. C., Sabuncuyan, S., Langmead, B., McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat Genet.* 43, 768-775.
- [0298] Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y., et al. (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol.* 28, 1097-1105.
- [0299] Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., and Bird, A. P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 6, e1001134.
- [0300] Inokuchi, K., Miyake, K., Takahashi, H., Dan, K., and Nomura, T. (1996). DCC protein expression in hematopoietic cell populations and its relation to leukemogenesis. *J Clin Invest.* 97, 852-857.
- [0301] Jaatinen, T., Hemmoraanta, H., Hautaniemi, S., Niemi, J., Nicorici, D., Laine, J., Yli-Harja, O., and Partanen, J. (2006). Global gene expression profile of human cord blood-derived CD133+ cells. *Stem Cells.* 24, 631-641.
- [0302] Jantus Lewintre, E., Reinoso Martin, C., Montaner, D., Marin, M., Jose Terol, M., Farras, R., Benet, I., Calvete, J., Dopazo, J., and Garcia-Conde, J. (2009). Analysis of chronic lymphocytic leukemia transcriptomic profile: differences between molecular subgroups. *Leuk Lymphoma.* 50, 68-79.
- [0303] Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 13, 484-492.
- [0304] Jones, P. A., and Baylin, S. B. (2007). The epigenomics of cancer. *Cell.* 128, 683-692.
- [0305] Kamani, N., Taylor, C., Malhotra, A., and Dutta, A. (2007). Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. *Genome research.* 17, 865-876.
- [0306] Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology.* 14, R36.
- [0307] Kim, J., Woo, A. J., Chu, J., Snow, J. W., Fujiwara, Y., Kim, C. G., Cantor, A. B., and Orkin, S. H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell.* 143, 313-324.
- [0308] Kreso, A., O'Brien, C. A., van Galen, P., Gan, O. I., Notta, F., Brown, A. M., Ng, K., Ma, J., Wienholds, E., Dunant, C., et al. (2013). Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science.* 339, 543-548.
- [0309] Kulis, M., Heath, S., Bibikova, M., Queiros, A. C., Navarro, A., Clot, G., Martinez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet.* 44, 1236-1242.
- [0310] Landan, G., Cohen, N. M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S., Zalcenstein, D. A., Goldfinger, N., Zundelovich, A., et al. (2012). Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet.* 44, 1207-1214.
- [0311] Landau, D. A., Carter, S. L., Getz, G., and Wu, C. J. (2014). Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia.* 28, 34-43.
- [0312] Landau, D. A., Carter, S. L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M. S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell.* 152, 714-726.
- [0313] Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 499, 214-218.
- [0314] Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. *N Engl J Med.* 363, 2424-2433.
- [0315] Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851-1858.
- [0316] Lim, E., Wu, D., Pal, B., Bouras, T., Asselin-Labat, M. L., Vaillant, F., Yagita, H., Lindeman, G. J., Smyth, G. K., and Visvader, J. E. (2010). Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* 12, R21.
- [0317] Maegawa, S., Gough, S. M., Watanabe-Okochi, N., Lu, Y., Zhang, N., Castoro, R. J., Estecio, M. R., Jelinek, J., Liang, S., Kitamura, T., et al. (2014). Age-related epigenetic drift in the pathogenesis of MDS and AML. *Genome Res.* 24, 580-591.
- [0318] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altschuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297-1303.
- [0319] Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E.,

- Nusbaum, C., Jaffe, D. B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 454, 766-770.
- [0320] Menke, A. L., Clarke, A. R., Leitch, A., Ijpenberg, A., Williamson, K. A., Spraggon, L., Harrison, D. J., and Hastie, N. D. (2002). Genetic interactions between the Wilms' tumor 1 gene and the p53 gene. *Cancer Res*. 62, 6615-6620.
- [0321] Morison, I. M., Ramsay, J. P., and Spencer, H. G. (2005). A census of mammalian imprinting. *Trends Genet*. 21, 457-465.
- [0322] Ohnishi, K., Semi, K., Yamamoto, T., Shimizu, M., Tanaka, A., Mitsunaga, K., Okita, K., Osafune, K., Arioka, Y., Maeda, T., et al. (2014). Premature Termination of Reprogramming In Vivo Leads to Cancer Development through Altered Epigenetic Regulation. *Cell*. 156, 663-677.
- [0323] Pei, L., Choi, J. H., Liu, J., Lee, E. J., McCarthy, B., Wilson, J. M., Speir, E., Awan, F., Tae, H., Arthur, G., et al. (2012). Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia. *Epigenetics*. 7, 567-578.
- [0324] Perez, C. A., Ott, J., Mays, D. J., and Pietenpol, J. A. (2007). p⁶³ consensus DNA-binding site: identification, analysis and application into a p63MH algorithm. *Oncogene*. 26, 7363-7370.
- [0325] Pujadas, E., and Feinberg, A. P. (2012). Regulated noise in the epigenetic landscape of development and disease. *Cell*. 148, 1123-1131.
- [0326] Quesada, V., Conde, L., Villamor, N., Ordonez, G. R., Jares, P., Bassaganyas, L., Ramsay, A. J., Bea, S., Pinyol, M., Martinez-Trillos, A., et al. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 44, 47-52.
- [0327] Ramskold, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., Daniels, G. A., Khrebtkova, I., Loring, J. F., Laurent, L. C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*. 30, 777-782.
- [0328] Rassenti, L., Jain, S., Keating, M., Wierda, W., Grever, M., Byrd, J., Kay, N., Brown, J., Gribben, J., Neuberg, D., et al. (2008). Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood*. 112, 1923-1930.
- [0329] Raval, A., Tanner, S., Byrd, J., Angerman, E., Perko, J., Chen, S., Hackanson, B., Grever, M., Lucas, D., Matkovic, J., et al. (2007). Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. *Cell*. 129, 879-890.
- [0330] Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26, 139-140.
- [0331] Rossi, D., Rasi, S., Spina, V., Bruscaggini, A., Monti, S., Ciardullo, C., Deambrogi, C., Khiabanian, H., Serra, R., Bertoni, F., et al. (2013). Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood*. 121, 1403-1412.
- [0332] Sato, N., Fukushima, N., Maitra, A., Matsubayashi, H., Yeo, C. J., Cameron, J. L., Hruban, R. H., and Goggins, M. (2003). Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. *Cancer Res*. 63, 3735-3742.
- [0333] Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., Chen, P., Gertner, R. S., Gaublomme, J. T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*. 510, 363-369.
- [0334] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*. 27, 379-423.
- [0335] Shipony, Z., Mukamel, Z., Cohen, N. M., Landan, G., Chomsky, E., Zelig, S. R., Fried, Y. C., Ainsbinder, E., Friedman, N., and Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature*. 513, 115-119.
- [0336] Siegmund, K. D., Marjoram, P., Woo, Y. J., Tavaré, S., and Shibata, D. (2009). Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc Natl Acad Sci USA*. 106, 4828-4833.
- [0337] Smoley, S. A., Van Dyke, D. L., Kay, N. E., Heerema, N. A., Dell'Aquila, M. L., Dal Cin, P., Koduru, P., Aviram, A., Rassenti, L., Byrd, J. C., et al. (2010). Standardization of fluorescence in situ hybridization studies on chronic lymphocytic leukemia (CLL) blood and marrow cells by the CLL Research Consortium. *Cancer Genet Cytogenet*. 203, 141-148.
- [0338] Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M., and Sorger, P. K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*. 459, 428-432.
- [0339] Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., Paulovich, A., Pomeroy, S., Golub, T., Lander, E., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 102, 15545-15550.
- [0340] Timp, W., and Feinberg, A. P. (2013). Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer*. 13, 497-510.
- [0341] Ushijima, T., Watanabe, N., Okochi, E., Kaneda, A., Sugimura, T., and Miyamoto, K. (2003). Fidelity of the methylation pattern and its variation in the genome. *Genome research*. 13, 868-874.
- [0342] Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 22, 1680-1688.
- [0343] Wang, L., Lawrence, M. S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D. S., Zhang, L., et al. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 365, 2497-2506.
- [0344] Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D. J., Campan, M., Young, J., Jacobs, I., et al. (2007). Epigenetic stem cell signature in cancer. *Nat Genet*. 39, 157-158.
- [0345] Woodfine, K., Huddleston, J. E., and Murrell, A. (2011). Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. *Epigenetics & chromatin*. 4, 1.

[0346] Wong, D. J., Liu, H., Ridky, T. W., Cassarino, D., Segal, E., and Chang, H. Y. (2008). Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell*. 2, 333-344.

[0347] Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics*. 10, 232.

[0348] Yuille, M., Condie, A., Stone, E., Wilsher, J., Bradshaw, P., Brooks, L., and Catovsky, D. (2001). TCL1 is activated by chromosomal rearrangement or by hypomethylation. *Genes Chromosomes Cancer*. 30, 336-341.

[0349] Ziller, M. J., Gu, H., Muller, F., Donaghey, J., Tsai, L. T., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 500, 477-481.

[0350] Having thus described in detail preferred embodiments of the present invention, it is to be understood that the invention defined by the above paragraphs is not to be limited to particular details set forth in the above description as many apparent variations thereof are possible without departing from the spirit or scope of the present invention.

1. A method of assessing a subject in need thereof's cancer treatment prognosis comprising:

detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells in a tumor sample from the subject;

comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells; and

assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells,

wherein the presence or prevalence of inconsistent methylation status along the sequences or across the multiple gene copies indicates that the subject is more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination thereof.

2. The method according to claim 1,

wherein the DNA methylation is detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing, or

wherein the DNA methylation is detected by microarray, such as a genome-wide microarray, or

wherein comparing comprises calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism, or calculating information entropy, or

wherein the one or more regions of neighboring CpG sites having locally disordered methylation comprise CpG sites within a CpG island, or

wherein the one or more regions of neighboring CpG sites having locally disordered methylation are within a promoter.

3. (canceled)

4. (canceled)

5. (canceled)

6. (canceled)

7. The method according to claim 1, wherein the method further comprises detecting a subclonal genetic mutation.

8. A method of treating cancer in a subject in need thereof, comprising performing the method of claim 7, and; if a subclonal genetic mutation is detected in or near the one or more regions of neighboring CpG sites having the locally disordered methylation, ceasing or altering chemotherapeutic treatment, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation), or if no subclonal genetic mutation is detected in or near the one or more regions of neighboring CpG sites having the locally disordered methylation, initiating or continuing treatment with a chemotherapeutic treatment.

9. (canceled)

10. A method of treating cancer in a subject in need thereof, comprising performing the method of claim 1, and a) administering an antitumor agent to the subject if no or few inconsistencies in methylation status are identified or b) administering fewer antitumor agents to a subject having a low level of inconsistencies in methylation status and more antitumor agents to a subject having a high level of inconsistencies in methylation status.

11. The method according to claim 10, subpart a), wherein the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if a high level of inconsistencies in methylation status is identified.

12. The method according to claim 1, wherein the detecting step is performed before treatment with the antitumor agent, or

wherein the detecting step is performed after treatment with the antitumor agent.

13. (canceled)

14. A method for treating cancer in a subject in need thereof, comprising performing a prognostic method according to claim 1, administering an antitumor agent to the subject, and repeating the prognostic method, wherein the antitumor agent is administered after the initial prognostic method and before the subsequent prognostic method, and continuing to treat the subject with the antitumor agent if the level of inconsistent methylation status is substantially the same in the initial and subsequent prognostic methods or lower in the subsequent prognostic method than in the initial prognostic method.

15. The method according to claim 14, wherein the method comprises ceasing or altering treatment with an antitumor agent, or initiating a non-chemotherapeutic treatment (e.g., surgery or radiation) if the level of inconsistent methylation status is different in the subsequent prognostic method compared to the initial prognostic method, or

wherein the antitumor agent is administered to the subject for at least 3 months, at least 6 months, at least 9 months, at least 12 months, at least 24 months or at least 36 months before performing the subsequent diagnostic method, preferably at least 12 months.

16. (canceled)

17. A method of calculating a proportion of discordant reads (PDR) in a first tumor sample from a subject in need thereof comprising:

detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells in the tumor sample;

comparing the DNA methylation status of sequencing reads in multiple regions of neighboring CpG sites along a sequence of CpG sites in DNA of the plurality of cells; and

determining a relative number of cells in the tumor sample having inconsistent methylation status across the sequence of CpG sites as compared to the total number of cells in the tumor sample or a number of cells in the tumor sample having consistent methylation status across the sequence of CpG sites, or determining a level of inconsistent methylation status across the sequence of CpG sites in cells in the tumor sample.

18. The method according to claim 17, wherein the DNA methylation is detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing, or

wherein the DNA methylation is detected by microarray, such as a genome-wide microarray, or

wherein a PDR greater than 0.15 indicates that the subject is more likely to (1) develop resistance to an antitumor agent; (2) relapse after treatment with an antitumor agent; (3) develop a metastatic tumor; or (4) any combination of the above, or

wherein method further comprises treating the subject with an antitumor agent if the PDR is less than 0.15.

19. (canceled)

20. (canceled)

21. (canceled)

22. The method according to claim 17, wherein the method further comprises performing a second PDR calculation on a second tumor sample from the subject.

23. The method according to claim 22, wherein the method further comprises the step of comparing the PDR from the first tumor sample to the PDR from the second tumor sample, or

wherein the first and second tumor samples are from different tumors.

24. (canceled)

25. The method according to claim 22, wherein the first and second tumor samples are obtained from the same tumor at different times, such as at least 3 months apart, at least 6 months apart, at least 9 months apart, or at least 12 months apart.

26. The method according to claim 25, wherein the subject is treated with an antitumor agent between obtaining the first tumor sample and obtaining the second tumor sample.

27. A method of treating cancer in a subject in need thereof, comprising conducting the method of claim 26, and stopping or changing the treatment with the antitumor agent if the PDR from the second tumor sample is greater than the PDR from the first tumor sample, or continuing the treatment with the antitumor agent if the PDR from the second tumor sample is the same as or less than the PDR from the first tumor sample.

28. (canceled)

29. A method of treating a subject suffering from a tumor comprising:

calculating a first PDR, variance, epipolymorphism or information entropy from a tumor sample obtained from the subject according to the method of claim 1;

treating the subject with an antitumor agent;

calculating a second PDR, variance, epipolymorphism or information entropy from a tumor sample obtained from the same tumor in the subject; and

comparing the second PDR, variance, epipolymorphism or information entropy to the first PDR, variance, epipolymorphism or information entropy;

wherein the antitumor agent is administered after obtaining the first tumor sample and before obtaining the second tumor sample, and

wherein treatment with the antitumor agent is stopped or changed if the second PDR, variance, epipolymorphism or information entropy is different than the first PDR, variance, epipolymorphism or information entropy.

30. The method according to claim 29, wherein the subject is treated with the antitumor agent for 3 months, 6 months, 9 months or 12 months before the second PDR, variance, epipolymorphism or information entropy is calculated.

31. The method according to claim 1, wherein the tumor sample is from a solid tumor, such as adrenocortical carcinoma, bone tumors, brain cancer, breast cancer, cervical cancer, colorectal carcinoma, desmoid tumors, desmoplastic small round cell tumors, endocrine tumors, esophageal cancer, Ewing sarcoma family tumors, gastric cancer, germ cell tumors, head or neck cancer, hepatoblastoma, hepatocellular carcinoma, lung cancer, melanoma, mesothelioma, nasopharyngeal carcinoma, neuroblastoma, non-rhabdomyosarcoma soft tissue sarcoma, osteosarcoma, ovarian cancer, pancreatic cancer, prostate cancer, retinoblastoma, rhabdomyosarcoma, skin carcinoma, testicular cancer, thyroid carcinoma, uterine cancer and Wilms tumors, or

wherein the tumor sample is from a hematological cancer, such as leukemia, preferably chronic lymphocytic leukemia, or

wherein the antitumor agent is selected from an angiogenesis inhibitor, such as angiostatin K1-3, DL- α -difluoromethyl-ornithine, endostatin, fumagillin, genistein, minocycline, staurosporine, and ($\pm$)-thalidomide; a DNA intercalator/cross-linker, such as Bleomycin, Carboplatin, Carmustine, Chlorambucil, Cyclophosphamide, cis-Diammineplatinum(II) dichloride (Cisplatin), Melphalan, Mitoxantrone, and Oxaliplatin; a DNA synthesis inhibitor, such as ($\pm$)-Amethopterin (Methotrexate), 3-Amino-1,2,4-benzotriazine 1,4-dioxide, Aminopterin, Cytosine 3-D-arabino-furanoside, 5-Fluoro-5'-deoxyuridine, 5-Fluorouracil, Ganciclovir, Hydroxyurea, and Mitomycin C; a DNA-RNA transcription regulator, such as Actinomycin D, Daunorubicin, Doxorubicin, Homoharringtonine, and Idarubicin; an enzyme inhibitor, such as S(+)-Camptothecin, Curcumin, (-)-Deguelin, 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside, Etoposide, Formestane, Fostriecin, Hispidin, 2-Imino-1-imidazoli-dineacetic acid (Cyclocreatine), Mevinolin, Trichostatin A, Tyrphostin AG 34, and Tyrphostin AG 879; a gene regulator, such as 5-Aza-2'-deoxycytidine, 5-Azacytidine, Cholecalciferol (Vitamin D3), 4-Hydroxytamoxifen, Melatonin, Mifepristone, Raloxifene, all trans-Retinal (Vitamin A aldehyde), Retinoic acid, all trans (Vitamin A acid), 9-cis-Retinoic Acid, 13-cis-Retinoic acid, Retinol (Vitamin A), Tamoxifen, and Troglitazone; a microtubule inhibitor, such as Colchicine, Dolastatin 15, Nocodazole, Paclitaxel, Podophyllotoxin, Rhizoxin, Vinblas-

tine, Vincristine, Vindesine, and Vinorelbine (Navelbine); a neoantigen; and an unclassified antitumor agent, such as 17-(Allylamino)-17-demethoxygeldanamycin, 4-Amino-1,8-naphthalimide, Apigenin, Brefeldin A, Cimetidine, Dichloromethylene-diphosphonic acid, Leuprolide (Leuporelin), Luteinizing Hormone-Releasing Hormone, Pifithrin- α , Rapamycin, Sex hormone-binding globulin, Thapsigargin, and Urinary trypsin inhibitor fragment (Bikunin) rituximab (Rituxan®), alemtuzumab (Campath®), Ipilimumab (Yervoy®), Bevacizumab (Avastin®), Cetuximab (Erbix®), panitumumab (Vectibix®), and trastuzumab (Herceptin®), Vemurafenib (Zelboraf®) imatinib mesylate (Gleevec®), erlotinib (Tarceva®), gefitinib (Iressa®), Vismodegib (Erivedge™), 90Y-ibritumomab tiuxetan, 131I-tositumomab, ado-trastuzumab emtansine, lapatinib (Tykerb®), pertuzumab (Perjeta™), ado-trastuzumab emtansine (Kadcyla™), regorafenib (Stivarga®), sunitinib (Sutent®), Denosumab (Xgeva®), sorafenib (Nexavar®), pazopanib (Votrient®), axitinib (Inlyta®), dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (Bosulif®), ofatumumab (Arzerra®), obinutuzumab (Gazyva™), ibrutinib (Imbruvica™), idelalisib (Zydelig®), crizotinib (Xalkori®), erlotinib (Tarceva®), afatinib dimaleate (Gilotrif®), ceritinib (LDK378/Zykadia), TOSITUMOMAB and 131I-tositumomab (Bexxar®), ibritumomab tiuxetan (Zevalin®), brentuximab vedotin (Adcetris®), bortezomib (Velcade®), siltuximab (Sylvant™), trametinib (Mekinist®), dabrafenib (Tafinlar®), pembrolizumab (Keytruda®), carfilzomib (Kymprolis®), Ramucirumab (Cyramza™), Cabozantinib (Cometriq™), vandetanib (Caprelsa®), a cytokine such as interferons (INFs), interleukins (ILs), or hematopoietic growth factors, INF- α , IL-2, Aldesleukin, IL-2, Erythropoietin, Granulocyte-macrophage colony-stimulating factor (GM-CSF), granulocyte colony-stimulating factor, a targeted therapy such as toremifene (Fareston®), fulvestrant (Faslodex®), anastrozole (Arimidex®), exemestane (Aromasin®), letrozole (Femara®), ziv-aflibercept (Zaltrap®), Alitretinoin (Panretin®), temsirolimus (Torisel®), Tretinoin (Vesanoid®), denileukin difitox (Ontak®), vorinostat (Zolinza®), romidepsin (Istodax®), bexarotene (Targretin®), pralatrexate (Folotylin®), lenalidomide (Revlimid®), belinostat (Beleodaq™), lenalidomide (Revlimid®), pomalidomide (Pomalyst®), Cabazitaxel (Jevtana®), enzalutamide (Xtandi®), abiraterone acetate (Zytiga®), radium 223 chloride (Xofigo®), or everolimus (Afinitor®), a checkpoint inhibitor such as an inhibitor of the programmed death-1 (PD-1) pathway, an anti-PD1 antibody (Nivolumab), an anti-cytotoxic T-lymphocyte-associated antigen (CTLA-4) antibody, an antibody targeting the CD28 CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR, an inhibitor targeting a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3, an epigenetic targeted drug such as HDAC inhibitors, kinase inhibitors, DNA methyltransferase inhibitors, histone demethylase inhibitors, or histone methylation inhibitors, epigenetic drugs such as Azacitidine (Vidaza), Decitabine (Dacogen), Vorinostat (Zolinza), Romidepsin (Istodax), or Ruxolitinib (Jakafi).

32. (canceled)

33. (canceled)

34. A method for identifying an agent that decreases the potential evolutionary capacity of a tumor comprising growing a first culture of hyperproliferative cells and a second culture of hyperproliferative cells, wherein the first culture is grown in the presence of an agent and the second culture is grown in the absence of the agent; detecting DNA methylation status at one or more regions of neighboring CpG sites in a plurality of cells from the first culture and a plurality of cells from the second culture; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the first culture; comparing the DNA methylation status of neighboring CpG sites along one or more sequences of neighboring CpG sites in DNA of the plurality of cells and/or comparing the DNA methylation status of corresponding CpG sites across multiple gene copies in the plurality of cells of the second culture; and assessing the consistency of methylation status along the sequences of neighboring CpG sites and/or across multiple gene copies in the plurality of cells, wherein the agent decreases the potential evolutionary capacity of cancer if the level of inconsistent methylation status is less in the first culture than in the second culture.

35. The method according to claim 34, wherein the DNA methylation is detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing, or

wherein the DNA methylation is detected by microarray, such as a genome-wide microarray.

36. (canceled)

37. The method according to claim 34, wherein the comparing step comprises calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism or calculating methylation entropy.

38. The method according to claim 37, wherein the comparing step comprises calculating a proportion of discordant reads and wherein the agent decreases the potential evolutionary capacity of cancer if the PDR of the first culture is less than the PDR of the second culture, or

wherein comparing comprises calculating variance and wherein the agent decreases the potential evolutionary capacity of cancer if the variance of the second tumor sample is less than the variance of the first tumor sample, or

wherein comparing comprises calculating epipolymorphism and wherein the agent decreases the potential evolutionary capacity of cancer if the epipolymorphism of the second tumor sample is less than the epipolymorphism of the first tumor sample, or

wherein comparing comprises calculating information entropy and wherein the agent decreases the potential evolutionary capacity of cancer if the information entropy of the second tumor sample is less than the information entropy of the first tumor sample.

39. (canceled)

40. (canceled)

41. (canceled)
42. The method according to claim 34, wherein the hyperproliferative cells in the first and second cultures comprise cells from a cell line, e.g., a tumor cell line, or wherein the hyperproliferative cells in the first and second cultures are cells from a tumor sample obtained from a subject, preferably a human, or cells cultured from such a sample, or wherein the first and second cultures are the same culture, wherein the second culture is a sample of the hyperproliferative cells before addition of the antitumor agent and the first culture is a sample of the hyperproliferative cells after addition of the antitumor agent, or wherein the first culture is cultured in the presence of the antitumor agent for at least 6 hours, at least 12 hours, at least 18 hours, at least one day, at least two days, at least three days, at least four days, at least five days, at least six days or at least one week, preferably at least one day, prior to detecting methylation status.
43. (canceled)
44. (canceled)
45. (canceled)
46. A method for identifying an agent that decreases the potential evolutionary capacity of a tumor, comprising performing a first prognostic method according to claim 1 on a first tumor sample from a subject, such as a laboratory animal; administering the agent to the subject; and performing a second prognostic method according to claim 1 on a second tumor sample from the subject; wherein the agent is administered between the first and second prognostic methods; wherein the agent decreases the potential evolutionary capacity of cancer if level of inconsistent methylation status is less in the second tumor sample than in the first tumor sample.
47. The method according to claim 46, wherein the DNA methylation is detected by massive parallel sequencing with bisulfite conversion, e.g., whole-genome bisulfite sequencing or reduced representation bisulfite sequencing, or wherein the DNA methylation is detected by microarray, such as a genome-wide microarray.
48. (canceled)
49. The method according to claim 46, wherein the comparing step comprises calculating the proportion of discordant reads, calculating variance, calculating epipolymorphism or calculating methylation entropy.
50. The method according to claim 49, wherein comparing comprises calculating a proportion of discordant reads and wherein the agent decreases the potential evolutionary

capacity of cancer if the PDR of the second tumor sample is less than the PDR of the first tumor sample, or

wherein comparing comprises calculating variance and wherein the agent decreases the potential evolutionary capacity of cancer if the variance of the second tumor sample is less than the variance of the first tumor sample, or

wherein comparing comprises calculating epipolymorphism and wherein the agent decreases the potential evolutionary capacity of cancer if the epipolymorphism of the second tumor sample is less than the epipolymorphism of the first tumor sample, or

wherein comparing comprises calculating information entropy and wherein the agent decreases the potential evolutionary capacity of cancer if the information entropy of the second tumor sample is less than the information entropy of the first tumor sample.

51. (canceled)

52. (canceled)

53. (canceled)

54. The method according to claim 46, wherein the agent is administered to the subject for at least one day, at least two days, at least three days, at least 4 days, at least five days, at least six days, at least one week, at least two weeks, at least three weeks, at least one month, preferably at least one week, or

wherein the laboratory animal is a mouse, a rabbit, a rat, a guinea pig, or a hamster, or

wherein the first and second tumor samples are from a solid tumor, such as adrenocortical carcinoma, bone tumors, brain cancer, breast cancer, cervical cancer, colorectal carcinoma, desmoid tumors, desmoplastic small round cell tumors, endocrine tumors, esophageal cancer, Ewing sarcoma family tumors, gastric cancer, germ cell tumors, head or neck cancer, hepatoblastoma, hepatocellular carcinoma, lung cancer, melanoma, mesothelioma, nasopharyngeal carcinoma, neuroblastoma, non-rhabdomyosarcoma soft tissue sarcoma, osteosarcoma, ovarian cancer, pancreatic cancer, prostate cancer, retinoblastoma, rhabdomyosarcoma, skin carcinoma, testicular cancer, thyroid carcinoma, uterine cancer and Wilms tumors, or

wherein the first and second tumor samples are from a hematological cancer, such as leukemia, preferably CLL.

55. (canceled)

56. (canceled)

57. (canceled)

* * * * *