Cancer Cell
Article

Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia

Dan A. Landau,[1,2,3,13] Kendell Clement,[3,4,5,13] Michael J. Ziller,[3,4] Patrick Boyle,[3] Jean Fan,[6] Hongcang Gu,[3] Kristen Stevenson,[7] Carrie Sougnez,[3] Lili Wang,[1,2] Shuqiang Li,[8] Dylan Kotliar,[1] Wandi Zhang,[1] Mahmoud Ghandi,[3] Levi Garraway,[2,3] Stacey M. Fernandes,[2] Kenneth J. Livak,[8] Stacey Gabriel,[3] Andreas Gnirke,[3] Eric S. Lander,[3] Jennifer R. Brown,[2,9] Donna Neuberg,[7] Peter V. Kharchenko,[6,10] Nir Hacohen,[3,11] Gad Getz,[3,12,14] Alexander Meissner,[3,4,14,*] and Catherine J. Wu[1,2,10,14,*]
[1]Cancer Vaccine Center
[2]Department of Medical Oncology
Dana-Farber Cancer Institute, Boston, MA 02115, USA
[3]Broad Institute, Cambridge, MA 02139, USA
[4]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[5]Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139, USA
[6]Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA
[7]Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA 02115, USA
[8]Fluidigm, South San Francisco, CA 94080, USA
[9]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[10]Division of Hematology/Oncology, Children's Hospital, Boston, MA 02115, USA
[11]Center for Immunology and Inflammatory Diseases
[12]Cancer Center and Department of Pathology
Massachusetts General Hospital, Boston, MA 02114, USA
[13]Co-first author
[14]Co-senior author
*Correspondence: alexander_meissner@harvard.edu (A.M.), cwu@partners.org (C.J.W.)
http://dx.doi.org/10.1016/j.ccell.2014.10.012

SUMMARY

Intratumoral heterogeneity plays a critical role in tumor evolution. To define the contribution of DNA methylation to heterogeneity within tumors, we performed genome-scale bisulfite sequencing of 104 primary chronic lymphocytic leukemias (CLLs). Compared with 26 normal B cell samples, CLLs consistently displayed higher intrasample variability of DNA methylation patterns across the genome, which appears to arise from stochastically disordered methylation in malignant cells. Transcriptome analysis of bulk and single CLL cells revealed that methylation disorder was linked to low-level expression. Disordered methylation was further associated with adverse clinical outcome. We therefore propose that disordered methylation plays a similar role to that of genetic instability, enhancing the ability of cancer cells to search for superior evolutionary trajectories.

INTRODUCTION

Cancer evolution is a central obstacle to achieving cure, as treatment-resistant disease often emerges even in the context of highly effective therapies. Recent studies by us and others have demonstrated the contribution of genetic heterogeneity within each individual cancer to clonal evolution and its impact on clinical outcome (reviewed in Landau et al., 2014). In addition

Significance

Although it is well established that genetic intratumoral diversity fuels tumor evolution, relatively little is known about epigenetic diversity in primary cancer samples and its impact on evolution and outcomes. Using a variety of molecular platforms, we demonstrated a higher degree of intratumoral heterogeneity of DNA methylation in CLL. We have further shown that this heterogeneity stems from seemingly stochastic variation, reminiscent of the model of genetic heterogeneity in cancer, wherein stochastic variation is subjected to selection in tumor evolution. These data transform the way we view methylation differences between normal and cancer cells and will facilitate the crucial distinction between epigenetic alterations that result from background stochastic variation versus positive selection, in this leukemia and other cancers.

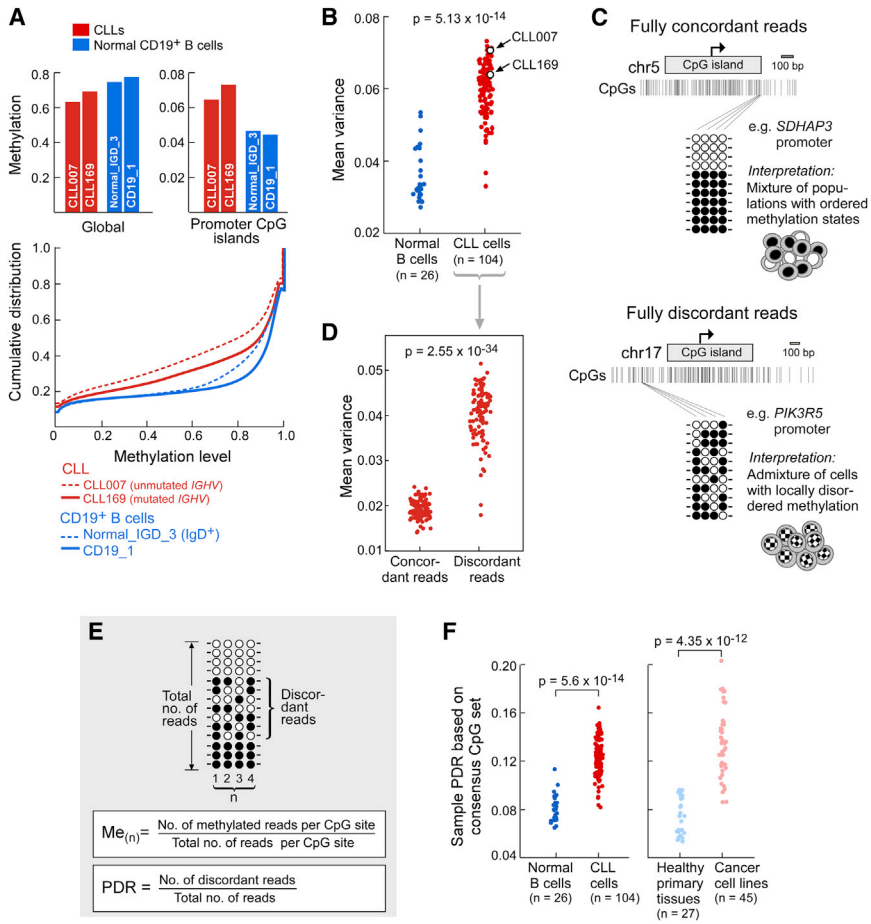


Figure 1. Higher DNA Methylation Intrasample Heterogeneity in CLL Arises from Locally Disordered Methylation

(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 (also see Figure S1).

(B) Mean intrasample 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 nonimprinted 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 intrasample CpG variance that arises from discordant compared with concordant reads across the 104 CLLs.

(E) CpG methylation and the 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 on the basis of a consensus set of 63,443 CpGs that are covered with greater than ten reads in >75% of all 202 RRBS samples.

See also Figure S1 and Tables S1 and S2.

to genetic mutations, somatic epigenetic alterations are also drivers of neoplastic transformation and fitness (Baylin, 2005; Baylin and Jones, 2011). Moreover, genetically uniform cells exhibit phenotypic variation in essential properties such as survival capacity and proliferative potential (Kreso et al., 2013; Spencer et al., 2009), likely reflecting epigenetic variation. Hence, a priority in cancer biology is to measure intratumoral heterogeneity at the epigenetic level and determine how somatic genetic and epigenetic heterogeneity together affect tumor evolution.

To examine these critical questions, we focused on chronic lymphocytic leukemia (CLL), a malignancy of mature B cells with well-documented epigenetic dysregulation of CLL-associated genes (Raval et al., 2007; Yuille et al., 2001). Stable differences have been observed in DNA methylation across CLL samples compared with normal B cells as well as between subtypes of CLL (e.g., with mutated versus unmutated *IGHV*) (Cahill et al., 2013; Harris et al., 2010; Kulis et al., 2012; Pei et al., 2012). We were motivated to perform an integrative study of intraleukemic genetic and DNA methylation heterogeneity in CLL because (1) recent studies have suggested that both epigenetic marks and genetic alterations can improve prognostic models of CLL (Kulis et al., 2012; Rossi et al., 2013); (2) higher methylation variability has been detected across cancer subtypes compared with healthy tissue-matched samples, including in other B cell malignancies (Berman et al., 2012; De et al., 2013; Hansen et al., 2011); and (3) the availability of whole-genome bisulfite sequencing (WGBS) and reduced-representation bisulfite sequencing (RRBS) now enables genome-wide investigation of DNA methylation at single base pair resolution and with local sequence context. In particular, RRBS constitutes a cost-effective approach that allows the study of large patient cohorts (Boyle et al., 2012).

We thus performed WGBS and RRBS on a large cohort of primary patient samples that were previously characterized by whole-exome sequencing (WES) (Landau et al., 2013), to assess intraleukemic DNA methylation heterogeneity in CLL.

RESULTS

Increased Intrasample DNA Methylation Heterogeneity in CLL Arises from Locally Disordered Methylation

To measure intrasample CLL DNA methylation heterogeneity, we compared WGBS data generated from two CLL cases and two healthy donor B cell samples (Figure 1A). We observed globally decreased methylation in CLL compared with normal B cells, with focally increased methylation of CpG islands (CGIs) (Figure 1A, top; Figures S1A–S1C available online), 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 (Figure 1A, bottom; Figures S1A–S1D), pointing to a large proportion of CpGs that are methylated in some cells in the sample and unmethylated in others. We reanalyzed published WGBS and Illumina 450 K methylation array data (Kulis et al., 2012) and confirmed the increased cell-to-cell variability in CpG methylation in CLL compared with normal B cells (Figures S1E–S1H).

We next applied RRBS to 104 primary CLL samples that had been previously characterized by WES (Landau et al., 2013) (Tables S1 and S2) and examined mean CpG variance. Consistent with the WGBS data, a greater than 50% increase in intrasample methylation heterogeneity was detected in CLL cells compared with 26 normal B cell samples (Figure 1B). We considered two possible sources for intrasample heterogeneity: variability between concordantly methylated fragments (i.e., whereby CpGs in an individual fragment are consistently methylated or unmethylated; Figure 1C, left) and variability within DNA fragments (i.e., discordant methylation by which CpGs in an individual fragment are variably methylated; Figure 1C, right).

On the basis of established observations that short-range methylation is highly correlated in normal physiological states (Eckhardt et al., 2006; Jones, 2012), we initially hypothesized that intrasample heterogeneity in CLL stems from variability between concordantly methylated fragments, reflecting a mixture of subpopulations with distinct but uniform methylation patterns. To test this, we focused on CpGs covered by reads containing four 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 100-mer WGBS reads, and an average of 307,041 [range 278,105–335,977] CpGs/sample covered by 29-mer RRBS reads). Contrary to the expected hypothesis, we found that 67.6 ± 3.2% (average ± SD) of the intratumoral methylation variance resulted from discordantly methylated reads across the 104 CLL samples ($p = 3.24 \times 10^{-35}$; Figure 1D). Similarly, the CLL WGBS confirmed a higher proportion of heterogeneously methylated CpGs in the discordant reads compared with the concordant reads (Figure S1E, right). These results demonstrate that methylation heterogeneity in CLL arises primarily from variability within DNA fragments, which we have therefore termed "locally disordered methylation."

We performed several analyses to exclude potential alternative explanations to these findings, including the impact of contaminating nonmalignant cells (Figure S1I), allele-specific methylation (Figures S1J–S1L), the contribution of reads that cover an ordered transition point from one methylation state to another (Figure S3L), and technical biases (see Supplemental Experimental Procedures). 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 affected by somatic copy number variations (see Supplemental Experimental Procedures and Figure S1O).

To quantify the magnitude of this phenomenon across large collections of normal and malignant human tissues, we analyzed RRBS data not only from the 104 CLL and 26 B cell samples but also from 45 solid and blood cancer cell lines and from 27 primary human tissue samples (Table S2). We 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 (Figure 1E). As expected, we found that the average PDR was higher in CLL compared with normal B cells ($p = 5.60 \times 10^{-14}$). Similarly, we found higher PDR in cancer cell lines compared with a diverse collection of healthy human tissue samples ($p = 4.35 \times 10^{-12}$; Figure 1F). These results support the idea that locally disordered methylation is a general property of the malignant process.

Locally Disordered Methylation Broadly Affects the CLL Genome

To determine whether specific elements in the genome harbor higher levels of locally disordered methylation in CLL compared with normal B cells, we calculated the average PDR across the 104 CLL samples and 26 healthy donor B cell samples (Table S3).

In normal B cells, PDR levels were lowest in regions with major roles in gene regulation (promoters, CGIs, 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 (Figure 2A), regardless of whether they were relatively hypermethylated (e.g., CGIs) or hypomethylated (e.g., intergenic regions) compared with normal B cells (Figure 2B). This phenomenon appeared to be neither specific to a subregion of CGIs or promoters (e.g., CGI borders; Figure 2C) nor restricted to a subtype of CGI (Figure S2A). Increased PDR in CLL was also observed in highly repetitive DNA sequences (e.g., long interspersed elements [LINE] and long terminal repeat retrotransposons; Figure 2A, RRBS data, and Figure S2B, WGBS data), which largely account for the global DNA hypomethylation observed in cancer (Ehrlich, 2009).

Alterations in the DNA methylation regulatory machinery could affect 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, and IDH1-S210N) showed increased PDR compared with the 101 CLL samples wild-type for these genes (Figure S2C).

Locally Disordered Methylation Appears to Be a Largely Stochastic Process

Two observations in the data suggest that PDR measures a process that stochastically increases variation in methylation, a notion that 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 with B cells supports 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, we observed a larger relative PDR increase in CLL in regions with lower PDR in normal B cells. To formally measure the level of disorder, we undertook a parallel analysis to calculate Shannon's information entropy of intrasample methylation variation (Figure S3A). We determined this entropy to be higher in CLL than in normal B cells (as well as higher in cancer cell lines compared with normal tissues), consistent with an increase in stochastic "noise" (Figures S3B and S3C).

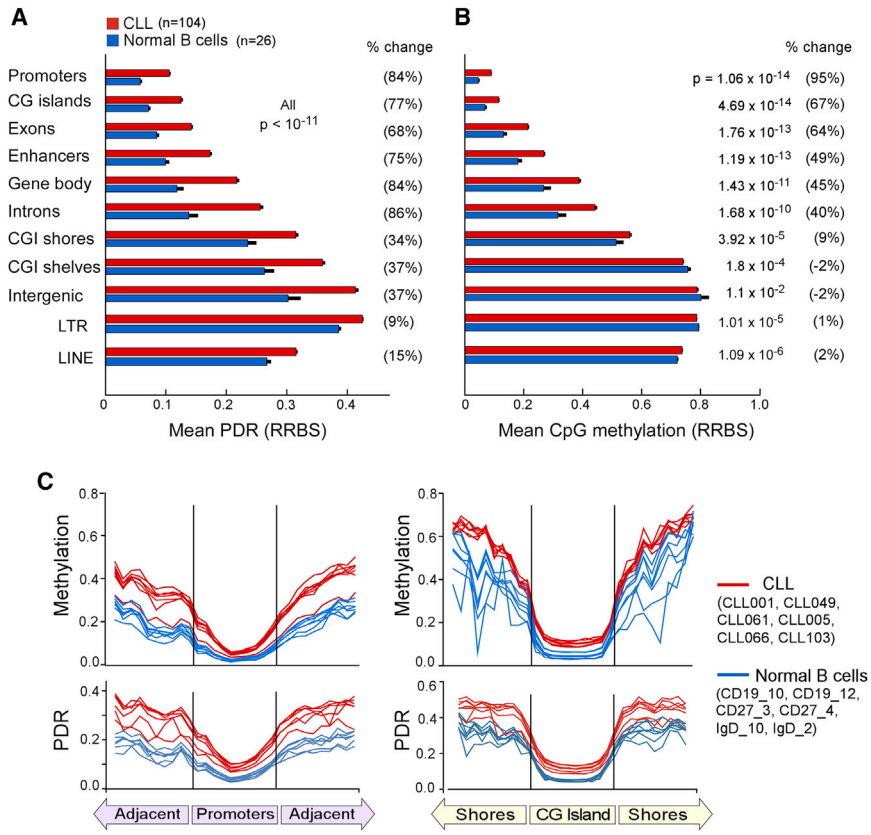


Figure 2. Locally Disordered Methylation Affects All Genomic Regions in CLL, Including CGIs and Repeat Regions

(A and B) Comparison of mean PDR (A) and mean CpG methylation (B) per genomic region between CLLs and normal B cells using RRBS data (Table S3 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 six CLL and six 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 2KB upstream and downstream are also shown. Bottom: an analogous analysis of CGIs and adjacent shore regions.

See also Figure S2 and Table S3.

To model the relationship between methylation and PDR under completely stochastic conditions, we plotted the expected distribution of PDR for any level of methylation assuming a purely random assignment of methylation states at each individual CpG (Figure 3A; see Supplemental Experimental Procedures). 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 (Figure 3B). In outlier genes (i.e., those with less promoter PDR than expected on the basis of the promoter methylation level; n = 195 [1.4%]; Table S4 and Figure S3D), imprinted genes were enriched (Morison et al., 2005) as expected, because these are hemimethylated under normal physiological conditions (n = 10, Fisher's 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 established roles 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.

Similar to promoters, methylation of ~1,900 LINE repeat elements also displayed a similar relationship between methylation and PDR (Figure 3C). A comparable distribution was observed for other genomic features (Figure S3E) and with RRBS data (Figure S3F). 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) (Figure S3G).

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 generated largely 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's correlation coefficient r = 0.90, $p = 1.01 \times 10^{-38}$; Figure 3D). When this analysis was repeated with genes grouped on the basis of 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 previously described distribution in Figure 3B (Figure S3H). 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. We likewise observed sample average LINE repeat elements PDR to be correlated with a decrease in methylation ($r = -0.32$, $p = 6.99 \times 10^{-4}$; Figure 3E).

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 (ΔPDR) and difference in methylation (ΔMeth). Because previous reports have indicated that a large degree of methylation disorder occurs during normal differentiation (Landan et al., 2012), we sought to compare the correlation between ΔPDR and ΔMeth among pairs of cancer and normal samples with the correlation between pairs of healthy human tissues. Indeed, the correlation coefficient between ΔPDR and ΔMeth was significantly higher when CLL samples were paired to either normal B cells or to other healthy primary tissue samples, compared with the pairing of healthy primary tissues against either normal B cells or other

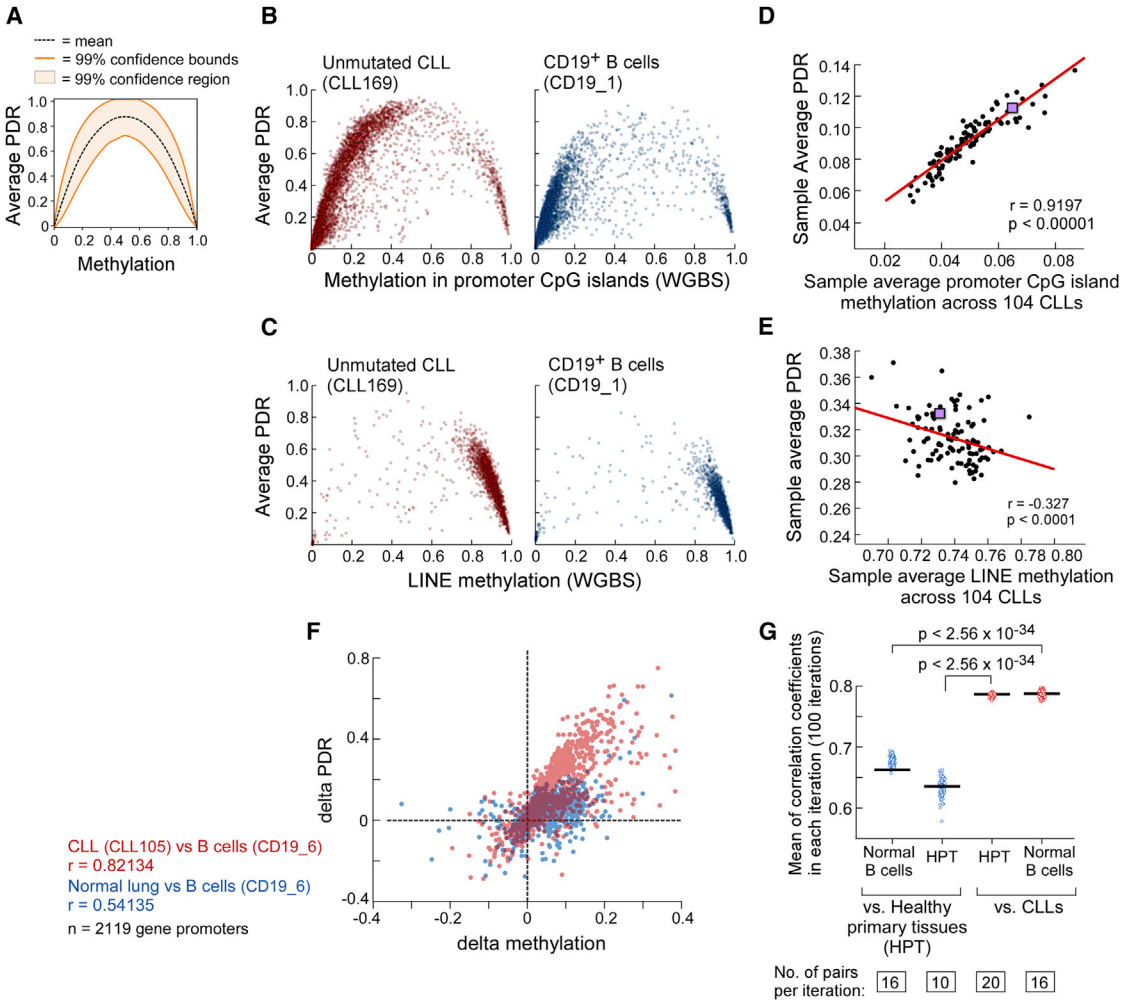


Figure 3. Locally Disordered Methylation in CLL Is Consistent with a Stochastic Process

(A) We 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 four 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 (more than ten CpGs per island) in both the CLL and the normal B cell samples. Outliers represent 1.4% of events (see Figure S3D and Table S4).

(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 ± 3,102.8 promoter CGIs per sample were evaluated; see also Figure S3E).

(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 ΔPDR and ΔMeth, we paired representative CLL and normal B cell samples. For each promoter (>20 CpGs per promoter, n = 2,119), ΔMeth and ΔPDR 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 ΔPDR and ΔMeth in cancer-related changes versus normal physiological state changes.

(G) To confirm this finding across the entire data set, 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. See also Figure S3 and Table S4.

healthy tissue samples (Figures 3F and 3G). 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.

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

Some regions of the genome may be more prone to stochastic variation in methylation (Pujadas and Feinberg, 2012). We found 3-fold higher promoter PDR in regions with the lowest gene

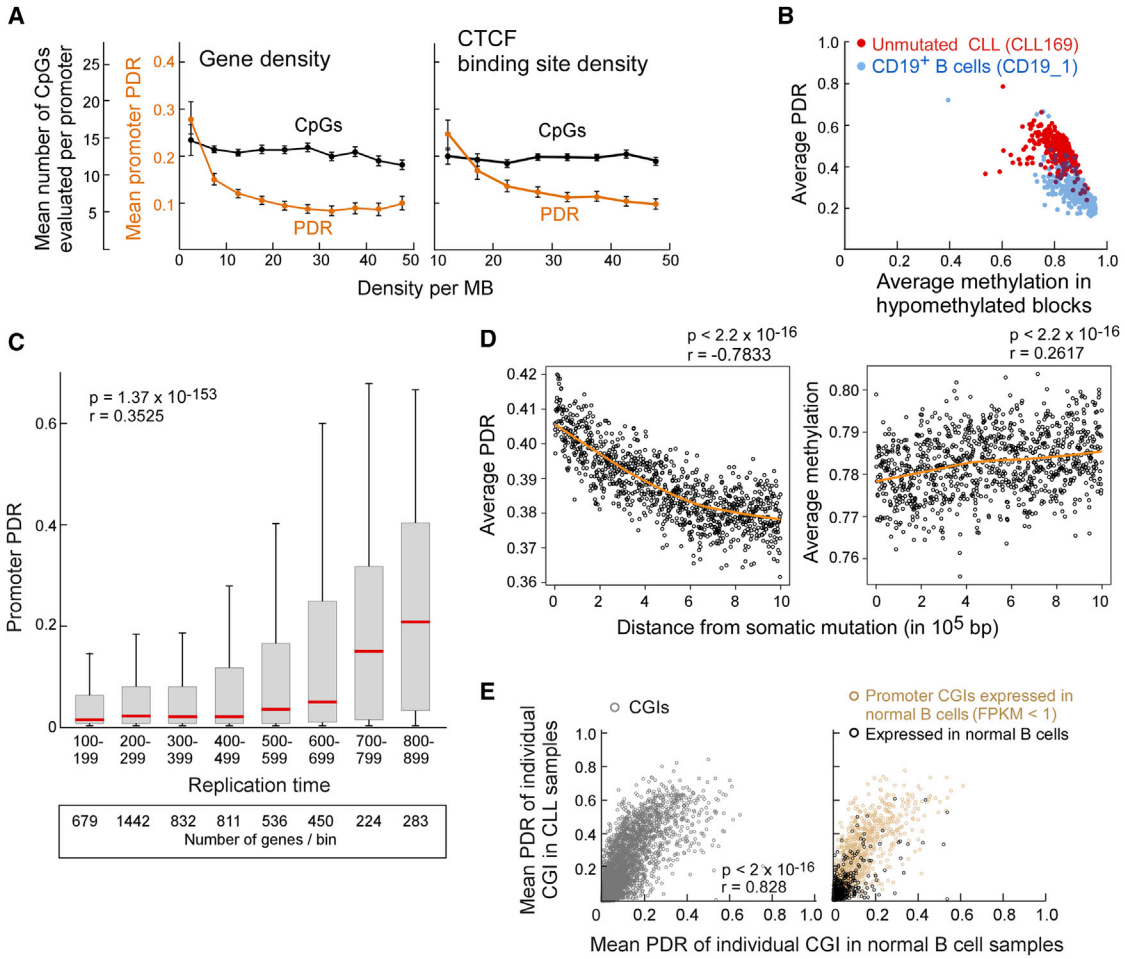


Figure 4. Locally Disordered Methylation Affects Preferentially Gene-Poor Regions and Can Be Traced Back to Nonexpressed Genes in Normal B Cells

(A) Promoter PDR (orange, error bars represent 95% CIs 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; see also Figure S4A for comparison with a matched set of control genomic blocks).

(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 denote the locally weighted scatterplot smoothing. See Figures S4B and S4C for an analysis performed separately for clonal and subclonal mutations.

(E) Left: promoter CGI PDR is correlated between CLL and normal B cell samples (Pearson, evaluated with 5,811 consistently covered CGIs). Right: promoter CGI PDR in B cells and CLLs is shown for genes expressed and not expressed in normal B cells (FPKM < 1, n = 1,002 from RNA-seq data of seven healthy donor B cell samples).

See also Figure S4.

density compared with those with highest gene density (with similar correlations to CTCF density; Figure 4A). 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, we observed a significant PDR increase in CLL (Figures 4B and S4A). Finally, in concert with these findings, we observed higher promoter PDR in genes with later replication time across the 104 CLL samples ($r = 0.35$, $p = 1.3 \times 10^{-153}$; Figure 4C), 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 (Figures 4D, S4B, and S4C).

As many features of chromatin and spatial organization may be shared between the CLL and normal B cell genomes, we

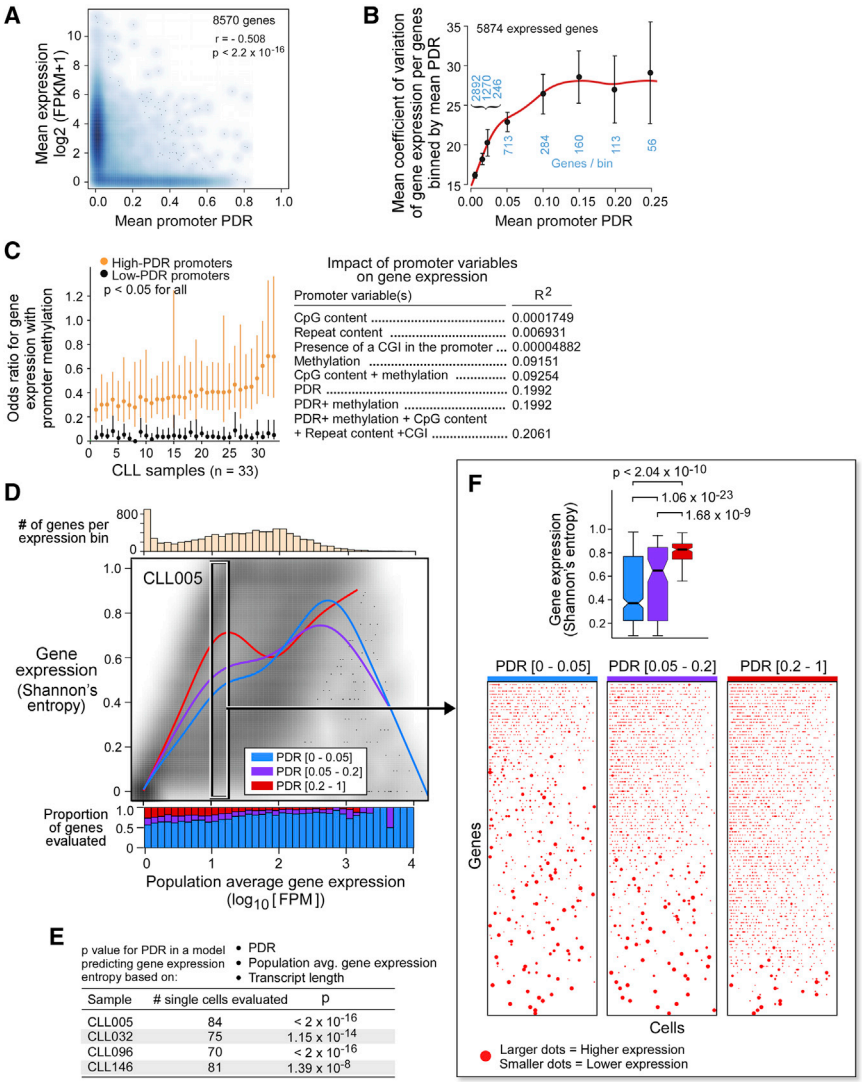


Figure 5. Locally Disordered Methylation Is Associated with Transcriptional Variation

(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 RNA-seq, the number of genes evaluated within each expression range provided in Figure S5A, and mean expression and methylation correlation is provided in Figure S5B.

(B) PDR and expression variability as measured with CV of 5,874 transcribed genes (FPKM > 1). Black circles (brackets) denote mean CV (95% CI) for genes within PDR bins (number of genes per bin in blue). Red line is the cubic smoothing spline of CV and PDR values (unbinned). Note that the analysis was limited to transcribed genes to avoid an artificial enhancement of the CV that occurs with very low mean expression values. Because >97.5% of transcribed genes had PDR < 0.3, we limited the x axis to PDR < 0.3.

(C) Left: OR (bars denote 95% CI) for gene expression (FPKM > 1) with a methylated promoter (average methylation > 0.8) versus an unmethylated promoter (average methylation < 0.2) is calculated for genes with high (orange, 27.5 ± 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 intrasample gene expression heterogeneity (assessed by Shannon's information entropy) across the range of population average expression (fragments per million [FPM]), by single-cell RNA-seq of 84 cells from CLL005 (see Figure S5D for analysis of three additional CLL samples). 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 on the basis of PDR, population average expression, and transcript length across the four CLL samples.

(F) Single-cell gene expression patterns for genes within a narrow population average expression range of 1.0 to 1.2 (black rectangle in 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, whereas genes with high PDR (bottom right) are frequently expressed at low expression magnitudes across many cells.

See also Figure S5 and Tables S5 and S6.

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 ($r = 0.83$, $p < 2 \times 10^{-16}$; Figure 4E, left). Thus, the promoters with highest PDR in CLL already have increased PDR in normal B cells. Consistent with the notion that nonexpressed 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 (Figure 4E, right).

Locally Disordered Methylation and Gene Expression

To examine the relationship between locally disordered DNA methylation and gene expression in more detail, we analyzed matched RRBS and RNA sequencing (RNA-seq) profiles of 33 CLL samples (Table S5; PDR and methylation calculated on the basis of an average ± SD of 12.1 ± 4.8 CpGs per promoter). As in normal B cells, in the 33 CLL samples, PDR was inversely correlated with gene expression ($r = -0.51$, $p < 2 \times 10^{-16}$; Figures 5A, S5A, and S5B). Notably, whereas promoter PDR was negatively correlated with mean transcript levels, it was positively correlated with intersample variation in transcript levels (Figure 5B). Although 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 (CV) of gene expression in a linear model ($p < 2 \times 10^{-16}$ for both).

To further examine the impact of locally disordered methylation in CLL on expression levels, we calculated the odds ratio

(OR) 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 ± SD promoter PDR was 0.10 ± 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 OR was 0.043 (range 0.036–0.050). In contrast, genes with high PDR promoters (greater than the mean PDR) had a greater likelihood of undergoing transcription (OR 0.396, range 0.259–0.698, Wilcoxon $p = 6.5 \times 10^{-11}$; Figure 5C), despite comparable promoter methylation levels. As a representative example, we 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 OR analysis above, we observed undetectable expression in CLL062 (FPKM of 0.03) and measurable RNA expression in CLL074 (FPKM of 5.6) (Figure S5C).

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, we used linear models to predict expression on the basis of methylation information. Across the 33 samples, a univariate model that predicts expression on the basis of average promoter methylation yielded an adjusted R^2 value of 0.092, whereas one using promoter PDR yielded an average adjusted R^2 value of 0.202. Inclusion of additional features such as CpG and repeat content only modestly improved the predictive power of the model (average adjusted $R^2 = 0.214$; Table S6). Indeed, the addition of PDR information to a model that uses promoter methylation to predict gene expression as a continuous variable (evaluated for 320,574 matched values of expression and methylation from 33 CLLs) resulted in a significant improvement, with more than doubling of the model's explanatory power (increase in adjusted R^2 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 (Figure 5C, right).

Single-Cell Gene Expression Patterns of Genes with Disordered Promoter Methylation

We next isolated 96 individual cells from four $CD19^+CD5^+$ purified CLL samples and generated single-cell full-length transcriptomes using SMART-seq (Clontech; 75–84 cells analyzed per sample after excluding cells with $< 1 \times 10^4$ aligned reads; Table S2). Promoter PDR was associated with significantly higher intratumoral expression information entropy in all four samples ($p < 1.4 \times 10^{-8}$; Figures 5D, 5E, and S5D), in a model that included transcript length as well as population average gene expression (see Supplemental Experimental Procedures), 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 (Figure S5E). Because expression information entropy may be affected by variation in sampling of lowly expressed transcripts, we compared the single-cell expression patterns of genes with low or high promoter methylation disorder but with similar population average expression levels (Figure 5F). We 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.

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

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, we identified differentially methylated regions (DMRs) in promoters and CGIs, because 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 (Figure S6A).

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; Figures 6A and S6B; Table S7). 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 with normal B cells, suggestive of a drift toward a more stem-cell-like state (Figure 6B). 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 ($p = 0.002$; Figure 6C).

To directly observe the relationship between genetic and epigenetic evolution, we studied RRBS data from 14 longitudinally sampled CLL patients with characterized patterns of genetic evolution (median time between samples 3.45 years; 9 CLLs with and 5 without evidence of genetic evolution; Table S8). CLLs that underwent genetic clonal evolution also had increased average promoter PDR over time (paired t test, $p = 0.037$; Figure 6D), 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) (Figure 6E; Table S9). Importantly, the correlation

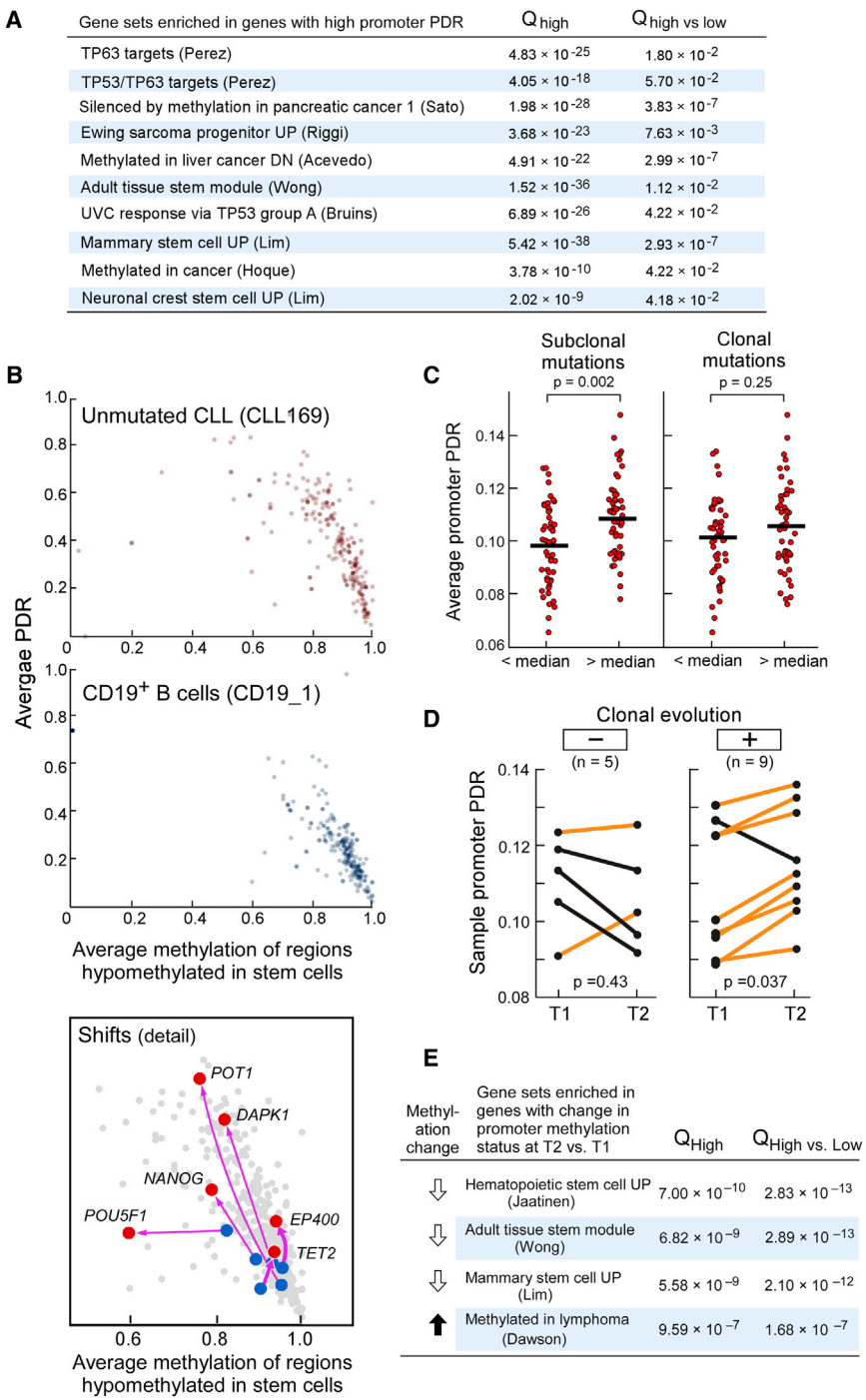


Figure 6. Locally Disordered Methylation May Interact with Evolution through Drift toward a Stem-like State

(A) Gene set enrichment analysis comparing 1,668 genes with consistently high promoter PDR (>0.1 in >75% of samples) with 5,392 genes with consistently low promoter PDR (<0.1 in >75% of samples, selected ten gene sets displayed; see Table S7 for the top 30 enrichments). 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 versus low-PDR genes was calculated using Fisher's exact test followed by BH-FDR ("Q(high versus 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*), six CGIs (e.g., *DAPK1*), six 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 versus low numbers of subclonal (median 7.5 sSNVs) and clonal mutations (median 10 sSNVs).

(D) Fourteen CLLs were sampled longitudinally at two time points (T1 and 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).

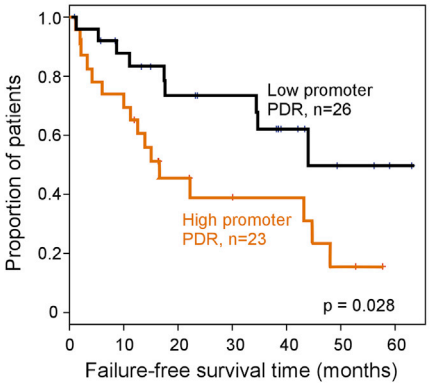
(E) Gene set enrichment of the 899 genes from the 14 cases with significant promoter methylation change between time points T1 and T2 (absolute change > 10%, FDR BH Q < 0.1) in genes with promoter demethylation over time (456 genes), and in genes with promoter methylation over time (443 genes; see Table S9 for top 30 enrichments). See also Figure S6 and Tables S7–S9.

coefficient between ΔPDR and ΔMeth 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 significant changes 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 for genes that underwent ordered methylation change, as expected from positive selection.

Locally Disordered Methylation Affects Clinical Outcome

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

We therefore examined failure-free survival after treatment (FFS; failure defined as retreatment or death) in 49 patients included in the cohort who were treated after tumor sampling for RRBS. A higher mean sample promoter PDR (greater than the mean for the cohort) was significantly associated with shorter FFS (median FFS of 16.5 versus 44 months, hazard ratio 2.5, 95% confidence interval [CI] 1.1 to 5.7, p = 0.028, Figure 7A; 52% and 65% of patients, respectively, were treated with fludarabine-based immunochemotherapy, p = 0.39). A regression



B

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

Figure 7. Locally Disordered Methylation Is Associated with Adverse Clinical Outcome

(A) Kaplan-Meier plot showing FFS 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 be performed only for the 49 patients who received therapy after RRBS sampling.

(B) Multivariate 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.

See also Table S10.

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, Figure 7B) for high promoter PDR. Similar results were obtained after the inclusion of additional variables in the model, including mutation burden and average promoter methylation (Table S10). 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 (Figure 7B). 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.

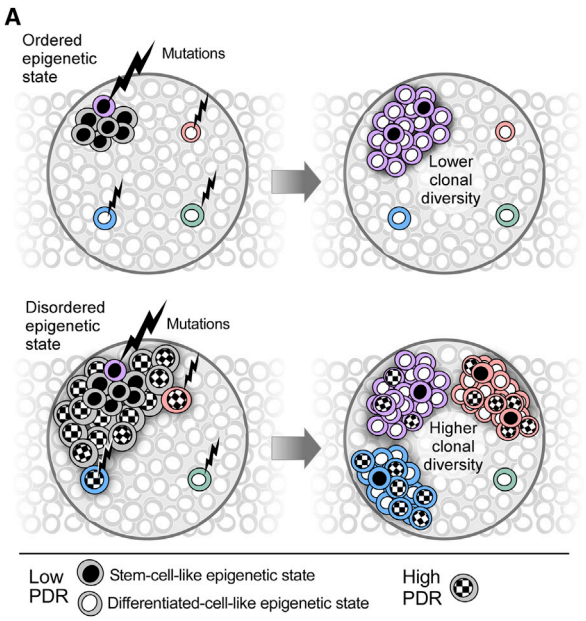
DISCUSSION

Cancer epigenomes have been long appreciated to differ from their normal tissue counterparts (Baylin and Jones, 2011). Global hypomethylation of cancer DNA was described as early as the 1980s, with frequent focal hypermethylation of key regulatory regions (Jones and Baylin, 2007). Recent genome-wide mapping have further highlighted alterations likely to contribute to the malignant process such as the epigenetic silencing of tumor suppressor genes and the activation of genes in stem-like cellular programs (Akiyama et al., 2003; Jones and Baylin, 2007; Widschwendter et al., 2007).

We now report the analysis of DNA methylation in primary leukemia cells that reveals another fundamental difference between cancer and normal methylomes: locally disordered methylation arising from a stochastic process, which leads to a high degree of intrasample methylation heterogeneity. These findings further advance key concepts described in several prior reports (Berman et al., 2012; Hansen et al., 2011; Landan et al., 2012; Maegawa et al., 2014; Pujadas and Feinberg, 2012; Siegmund et al., 2009). Thus, as previously suggested (Timp and Feinberg, 2013), cancer epigenomes may accommodate a higher amplitude of epigenetic "noise" and thereby allow cancer cells a greater degree of population diversity. Analogous to the role of genetic instability, which fuels cancer plasticity by facilitating the acquisition of somatic alterations at random locations across the genome (Hanahan and Weinberg, 2011), we propose that stochastic methylation changes enhance epigenetic plasticity and likewise enable tumor cells to better explore the evolutionary space in search of superior fitness trajectories.

These data alter the way we understand differential methylation in cancer. First, the insight that stochastic variation underlies the bulk of CLL methylome heterogeneity signifies that changes in methylation measured between cancer and normal cells do not likely reflect a uniform change in methylation state of a given region but rather a disordered methylation change involving differing, isolated CpGs, affecting many cells in the cancer population. Second, these data suggest improved methods from which we can identify fitness-enhancing DMRs. We can draw from the lessons of the computational analyses of large cancer genome sequencing data sets, in which a better understanding of the variation in the distribution of gene mutations has led to an improved ability to distinguish "driver" from "passenger" mutations (Lawrence et al., 2013). In an analogous fashion, we anticipate that appreciation of the extent of locally disordered methylation provides an appropriate background model against which a departure from the stochastic regime would indicate positively selected DMRs. We note that only a small proportion of methylation events fall outside the predictions of the stochastic model, suggesting very few of the changes in methylation undergo positive selection.

These data moreover demonstrate that locally disordered methylation is associated with a more "noisy" transcriptional landscape, with a decoupling of the relationship between promoter methylation and gene expression. Our analysis suggests that some of the epigenetic variability is likely associated with stemlike cell programs, which have been implicated in cancer (Kim et al., 2010; Ohnishi et al., 2014). Indeed, we detected a concurrent decrease in methylation and an increase in PDR, affecting regions that were identified to be hypomethylated in human embryonic stem cells, consistent with the notion that stochastic noise may lead to a drift toward a hybrid stem-somatic cell state (Timp and Feinberg, 2013). Furthermore, in CLLs that were directly observed to undergo genetic diversification and evolution over time, stem cell-related genes with higher promoter PDR also underwent demethylation over time. Thus, increased stochastic variation may blur the lines between populations with different proliferative potentials and thus increase



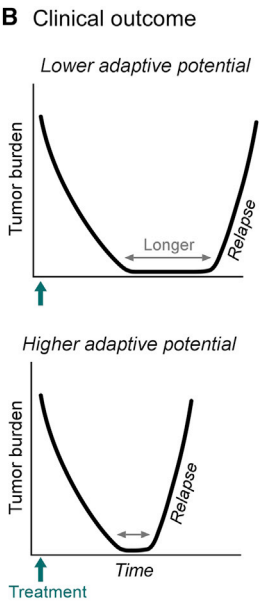


Figure 8. Proposed Interaction between Methylation Disorder and Clonal Evolution

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), the proportion of cells that are therefore able to actively participate in the evolutionary process is small. However, in a more malleable epigenetic landscape, such as expected to result from a 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.

the diversity of adaptive mechanisms available to the cell, a hedging strategy for enhanced survival (Balázsi et al., 2011).

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 (Figure 8, 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 (Figure 8, 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 we saw in our CLL cohort.

What is the basis of increased locally disordered methylation in CLL? Although the exact mechanism remains to be fully elucidated, we speculate that the considerably higher replication rate in CLL compared with their normal differentiated counterparts could contribute to accumulation of stochastic lapses in methylation inheritance in cancer cells, given the estimated error rate of 0.08% to 4% for a given CpG per cell division (Bird, 2002; Ushijima et al., 2003). This maybe further compounded by the occurrence of genetic lesions in essential components of the methylation machinery. In addition, the finding that locally disorder methylation in CLL tended to be highest in gene-poor and late-replicating regions suggests that some genomic regions exhibit even higher error rates, consistent with the previously observed high cancer intersample methylation variability in these regions (Hansen et al., 2011).

Our data suggest that evolution and diversity of DNA methylation in CLL result from stochastic events. This insight should improve our model for background methylation changes in cancer and allow more rigorous identification of positively selected methylated regions. Locally disordered DNA methylation is likely to have a similar role to genetic instability, providing a mechanism for cancer cells to find superior evolutionary trajectories during tumorigenesis and in response to therapy.

EXPERIMENTAL PROCEDURES

Sample Acquisition

Peripheral blood samples were obtained from patients with CLL and healthy adult volunteers. Informed consent on Dana-Farber Cancer Institute institutional review board-approved protocols for genomic sequencing of patients' samples was obtained prior to the initiation of sequencing studies. Genomic DNA was extracted from CLL cells or normal B cell populations.

WGBS

WGBS was performed as described in Supplemental Experimental Procedures. Subsequently, CpG methylation calls were made using custom software, excluding duplicate and low-quality reads. Previously published WGBS data for two CLL samples and three normal B cell samples (Kulis et al., 2012) were downloaded with permission and processed in identical fashion to the in-house-produced WGBS libraries.

RRBS

RRBS was performed as described in Supplemental Experimental Procedures. RRBS of primary diverse human tissue samples were previously reported (http://www.roadmapepigenomics.org). Reads were aligned, and methylation was determined using identical protocols to the rest of the samples.

RNA-Seq

RNA-seq of CLL and normal B cell samples was performed as previously described (Landau et al., 2013). For single-cell RNA-seq, the C1 Single-Cell Auto Prep System (Fluidigm) was used to perform SMARTer (Clontech) whole-transcriptome amplification (WTA), on up to 96 individual cells per sample from four primary CLL patient samples. WTA products were then converted to Illumina sequencing libraries using Nextera XT (Illumina) (Ramsköld et al., 2012).

Statistical Analysis

Statistical analysis was performed with MATLAB (The MathWorks), R version 2.15.2 (R Foundation for Statistical Computing), and SAS version 9.2 (SAS Institute). A complete description of the materials and methods is provided in Supplemental Experimental Procedures. The CLL and normal B cell

sequencing data were deposited in the database of Genotypes and Phenotypes (dbGaP) (phs000435.v2.p1) and the processed data deposited in Gene Expression Omnibus (GEO) (GSE58889).

ACCESSION NUMBERS

The GEO accession number for the data reported in this paper is GSE58889. The dbGaP accession number for the sequencing data reported in this paper is phs000435.v2.p1.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and ten tables and can be found with this article online at http://dx.doi.org/10.1016/j.ccell.2014.10.012.

AUTHOR CONTRIBUTIONS

D.A.L, K.C., N.H., G.G., A.M., and C.J.W. conceived and designed the experiments. D.A.L, P.B., H.G., L.W., D.K., W.Z., K.J.L., S.L., and A.G. performed the experiments. D.A.L, K.C., M.Z., J.F., K.S., M.G., D.N., and P.V.K. analyzed the data. Additional contribution to funding and sample collection and processing was provided by C.S., L.G., E.S.L., S.M.F., and J.R.B. The paper was written by D.A.L, K.C., A.M., and C.J.W.

ACKNOWLEDGMENTS

We thank all members of the Broad Institute's Biological Samples, Genetic Analysis, and Genome Sequencing Platforms, who made this work possible (NHGRI-U54HG003067). We thank Rahul Satija, Angela Brooks, Scott Carter and Brad Bernstein for their valuable input and insights. We thank John Daley, Suzan Lazo-Kallanian, Jonna Grimsby, and Niall Lennon for their assistance in the single-cell RNA-seq. We thank Adam Kiezun for his guidance regarding germline SNP detection and Michael Lawrence for the replication time data. D.A.L. is supported by a Postdoctoral Fellowship from the American Cancer Society (ACS) and by the NIH Big Data to Knowledge initiative (BD2K, 1K01ES025431-01). K.C. is supported by the National Science Foundation Graduate Research Fellowship under Grant No. 112237. L.W. is supported by a Lymphoma Research Foundation Postdoctoral Fellowship. J.F. is supported by the National Science Foundation Graduate Research Fellowship (DGE1144152). J.R.B. is a Scholar in Clinical Research of the Leukemia & Lymphoma Society (LLS) and is supported by a LLS Translational Research Program award as well as an ACS Research Scholar Grant and the Melton and Rosenbach Funds. A.M. is a New York Stem Cell Foundation Robertson Investigator and received support from NIH grants U01ES017155 and 1R01DA036898. C.J.W. is a Scholar of LLS and the recipient of a Quest for Cures Award from LLS. She acknowledges support from the Blavatnik Family Foundation, the American Association for Cancer Research (Stand Up to Cancer Innovative Research Grant), the National Heart, Lung, and Blood Institute (1RO1HL103532-01), and the National Cancer Institute (1R01CA155010-01A1). K.J.L. and S.L. are employees of Fluidigm Corporation.

Received: March 4, 2014
Revised: September 16, 2014
Accepted: October 24, 2014
Published: December 8, 2014

REFERENCES

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.

Akiyama, Y., Watkins, N., Suzuki, H., Jair, K.W., van Engeland, M., Esteller, M., Sakai, H., Ren, C.Y., Yuasa, Y., Herman, J.G., and Baylin, S.B. (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.

Balázsi, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. Cell *144*, 910–925.

Baylin, S.B. (2005). DNA methylation and gene silencing in cancer. Nat. Clin. Pract. Oncol. *2* (*Suppl 1*), S4–S11.

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.

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, 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.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Boquest, A.C., Shahdadfar, A., Frønsdal, K., Sigurjonsson, O., Tunheim, S.H., Collas, P., and Brinchmann, 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.

Boyle, P., Clement, K., Gu, H., Smith, Z.D., Ziller, M., Fostel, J.L., Holmes, L., Meldrim, J., Kelley, F., Gnirke, A., and Meissner, A. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol. *13*, R92.

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.

Cahill, N., Bergh, A.C., Kanduri, M., Göransson-Kultima, H., Mansouri, L., Isaksson, A., Ryan, F., Smedby, K.E., Juliusson, G., Sundström, 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.

Chim, C.S., Pang, R., and Liang, R. (2008). Epigenetic dysregulation of the Wnt signalling pathway in chronic lymphocytic leukaemia. J. Clin. Pathol. *61*, 1214–1219.

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.

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.

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. Epigenomics *1*, 239–259.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, 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.

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.

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.

Jaatinen, T., Hemmoranta, 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.

Jantus Lewintre, E., Reinoso Martín, C., Montaner, D., Marín, M., José Terol, M., Farrás, R., Benet, I., Calvete, J.J., Dopazo, J., and García-Conde, J. (2009). Analysis of chronic lymphotic leukemia transcriptomic profile: differences between molecular subgroups. Leuk. Lymphoma *50*, 68–79.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683–692.

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.

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.

Kulis, M., Heath, S., Bibikova, M., Queirós, A.C., Navarro, A., Clot, G., Martínez-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.

Landan, G., Cohen, N.M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S., Zalcenstein, D.A., Goldfinger, N., Zundelevich, A., et al. (2012). Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat. Genet. *44*, 1207–1214.

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.

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.

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.

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.

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.

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.

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.

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.

Morison, I.M., Ramsay, J.P., and Spencer, H.G. (2005). A census of mammalian imprinting. Trends Genet. *21*, 457–465.

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.

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.

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.

Pujadas, E., and Feinberg, A.P. (2012). Regulated noise in the epigenetic landscape of development and disease. Cell *148*, 1123–1131.

Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, 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.

Raval, A., Tanner, S.M., Byrd, J.C., Angerman, E.B., Perko, J.D., Chen, S.S., Hackanson, B., Grever, M.R., Lucas, D.M., Matkovic, J.J., et al. (2007). Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. Cell *129*, 879–890.

Rossi, D., Rasi, S., Spina, V., Bruscaggin, 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.

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.

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.

Shipony, Z., Mukamel, Z., Cohen, N.M., Landan, G., Chomsky, E., Zeliger, S.R., Fried, Y.C., Ainbinder, E., Friedman, N., and Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature *513*, 115–119.

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. U S A *106*, 4828–4833.

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.

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.

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 Res. *13*, 868–874.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., and Laird, P.W. (2007). Epigenetic stem cell signature in cancer. Nat. Genet. *39*, 157–158.

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.

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

Ziller, M.J., Gu, H., Müller, 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.