Preneoplastic Alterations Define CLL DNA Methylome and Persist through Disease Progression and Therapy


Association for Cancer Research. 
Downloaded from https://bloodcancerdiscov.aacrjournals.org by guest on July 13, 2021. Copyright 2020 American Association for Cancer Research.
ABSTRACT

Most human cancers converge to a deregulated methylome with reduced global levels and elevated methylation at select CpG islands. To investigate the emergence and dynamics of the cancer methylome, we characterized genome-wide DNA methylation in preneoplastic monoclonal B-cell lymphocytosis (MBL) and chronic lymphocytic leukemia (CLL), including serial samples collected across disease course. We detected the aberrant tumor-associated methylation landscape at CLL diagnosis and found no significant differentially methylated regions in the high-count MBL-to-CLL transition. Patient methylomes showed remarkable stability with natural disease and posttherapy progression. Single CLL cells were consistently aberrantly methylated, indicating a homogeneous transition to the altered epigenetic state and a distinct expression profile together with MBL cells compared with normal B cells. Our longitudinal analysis reveals the cancer methylome to emerge early, which may provide a platform for subsequent genetically driven growth dynamics, and, together with its persistent presence, suggests a central role in disease onset.

SIGNIFICANCE: DNA methylation data from a large cohort of patients with MBL and CLL show that epigenetic transformation emerges early and persists throughout disease stages with limited subsequent changes. Our results indicate an early role for this aberrant landscape in the normal-to-preneoplastic transition that may reflect a pan-cancer mechanism.

See related commentary by Rossi, p. 6.

INTRODUCTION

In normal adult tissues, cell identity is associated with accurate maintenance of a distinct DNA methylation land-
1% to 2% of patients per year (15). A well-established prognostic factor in CLL is the mutational status of the IGHV region genes, with mutated IGHV showing a much better prognosis than CLL with unmutated IGHV (16, 17). The IGHV mutational status has been thought to reflect differences in the cell of origin, with a similarity in methylaton profiles of unmutated CLLs and pregerminal center B cells, and of mutated CLL with mature, postgerminal center memory B cells, suggesting that CLL emerges from a spectrum of B cells undergoing broad DNA methylation alterations (11, 16, 18, 19). In addition to these characteristic global changes, we previously identified a pervasive local disorder of methylation across genomic features in CLL, not present in normal tissues (20). Although general changes in methylation profiles during B-cell development and cancer have been described (6, 10, 11, 20–24), little is currently known about: (i) if and which additional methylation changes are necessary to transition from normal into a pooneoplastic state and further into cancer, (ii) how this altered cancer methylome is affected by therapy, and (iii) why it is found so ubiquitously across different types and stages of cancer. Furthermore, the chronologic origin of altered methylation with respect to cancer initiation and progression is not well understood but would be of relevance for early detection and could lead to novel therapeutic strategies.

To approach these questions, we used bulk and single-cell reduced representation bisulfite sequencing (RRBS; refs. 25–27) to profile normal mature B cells, as well as cells from patients in the pooneoplastic MBL phase and during CLL progression, including after treatment. We characterized the methylation status of samples collected from S3 patients supplemented with WBC counts as a measure of tumor burden, and hence the effect of treatment (average sampling period of 5.7 years). Further, we used single-cell transcriptomics to complement the DNA methylation results in the patients transitioning from MBL to CLL. Our analyses reveal that changes in methylene and transcriptome are established early on, already at the precursor stage, and remain remarkably stable throughout the disease and even after therapy.

Results

Unmutated and Mutated CLLs Converge to a Similar Methylome

To systematically study the DNA methylation dynamics across the disease course of CLL, we generated RRBS datasets from CD19+ CD5+ cells collected from 23 individuals with MBL, matched samples for 5 patients capturing both the MBL and their transition to CLL, and serial pre- and posttreatment samples from 25 patients collected following the diagnosis of CLL (28, 29) and compared these with published B-cell–lineage subpopulations (refs. 10, 30; Fig. 1A; Supplementary Table S1).

Genome-wide correlation of single CpG methylation showed a substantial similarity of unmutated and mutated methylation profiles ($r = 0.96$); however, compared with their putative cell of origin, the CLL IGHV subtypes showed different degrees of abnormality. Although the unmutated CLL showed more changes compared with naïve B cells, the mutated CLL exhibited a methylation landscape more similar to memory B cells than naïve to memory B cells (Fig. 1B). As noted above, CLls originate from a range of developmental stages with pregerminal center B cells thought to give rise to unmutated CLL and mature, postgerminal center memory B cells to mutated CLLs (10, 11, 16, 18, 19, 31). Evaluation of single samples in a phylogenetic tree analysis revealed that the unmutated and mutated CLL samples are characterized by a methylome that consistently differs from normal naïve and memory B cells, suggesting a convergent disease-associated methylome, irrespective of IGHV mutation status (Fig. 1C; Supplementary Fig. S1A). Together, these results suggest that both IGHV subtypes of CLL undergo methylation changes specific to CLL. However, some of these changes also appear to be normally acquired during B-cell maturation, as observed in the example of the EBFB locus (Fig. 1D).

To more systematically evaluate regions that are consistently altered in CLL, we identified differentially methylated regions (DMR) between (i) unmutated CLL versus naïve B cells ($n = 23,206$ DMRs) and (ii) mutated CLL versus memory B cells...
(n = 4,653 DMRs; Supplementary Table S2; ref. 32). To disentangle methylation changes associated with normal cell lineage–specific differentiation from potentially cancer-related changes, we classified the aggregate of these two sets of DMRs as B-cell related (n = 22,325) or CLL (n = 3,475), depending on whether they were classified as dynamically changing during normal B-cell development (Fig. 1E and F; Methods; ref. 30). The majority (85%) of the DMRs overlapped with developmental regions, whereas 15% were classified as CLL DMRs (Fig. 1E and F; Supplementary Fig. S1B). Based on the clustering, B-cell lineage–related DMRs showed a gradual shift, mostly toward hypomethylation, from naive to memory and both CLL subtypes, reflecting the normal B-cell developmental changes that are retained in CLL. In contrast, as expected, the set of CLL DMRs readily distinguished normal B cells from CLL (Fig. 1E). Moreover, genes that were associated with CLL DMRs were found to be overrepresented among pathways related to cell growth and survival, proliferation, and neoplastic transformation, suggesting possible regulatory relevance (Fig. 1G).

We additionally confirmed the DMRs to be a distinctive feature between normal B and CLL cells by analyzing replicates of CD5-positive and -negative naive and memory B cells from a set of three healthy donors. Genome-wide phylogenetic tree clustering and the correlation of methylation rates revealed two major clusters, separating the samples by naive and memory B cells but not by CD5 status (Supplementary Fig. S2A and S2B). Based on the presence of the CLL-specific DMRs, CD5-sorted healthy donor samples were found to cluster into the group of previously published reference B cells, hence demonstrating similar methylation in DMRs independent of CD5 status of naive or memory B-cell state (Supplementary Fig. S2C).

**CLL Methylome Remains Mostly Unchanged after Treatment**

To evaluate the stability of these DMRs and the dynamics of the CLL methylome over time, we analyzed longitudinal samples collected during natural CLL progression. CLL allows that leukemic burden can be approximately estimated by measuring the WBC count over time since it is, for many patients, primarily a circulating malignancy. To study the CLL methylome before and after the first treatment, we performed unsupervised phylogenetic clustering of the pre- and posttreatment (fludarabine, cyclophosphamide, and rituximab, FCR) samples of patients. Interestingly, we found no consistent methylation differences that separate pre- and posttreatment samples, and also no WBC-related effects could be seen in the clustering (Fig. 2A).

Next, we compared methylation levels across patient time points for selected chromatin states derived from published data of the lymphoblastoid cell line GM12878 (33). Despite vastly different growth patterns and subclonal dynamics (defined by prior genetic characterization; ref. 28), global methylation levels of various genomic features, such as heterochromatin, strong enhancers, and poised promoters, for serial samples from all 21 patients remained stable and were consequently independent of the dynamic changes in WBC counts (Fig. 2B, left; Supplementary Fig. S3A and S3B).

We observed substantial posttreatment reduction in WBCs creating population bottlenecks for the nine patients following treatment with FCR (Fig. 2B; Supplementary Fig. S3A). However, this was not associated with any notable DNA methylation changes over the three representative genomic features or the number of distinct epialleles present (Fig. 2B; Supplementary Fig. S3; ref. 34). Indeed, methylation levels were mostly independent of detected subclonal genetic evolution and from patterns of growth. Similar stability was also found at the previously identified DMRs, with once acquired changes appearing to persist during CLL progression (Fig. 2C). These deregulated regions, including the example of NFATC1 (Fig. 2D), further highlight the remarkable stability of methylation patterns.

We also compared the nine patients by focusing on the clinically most divergent time points, i.e., first and last...
The Altered DNA Methylation Landscape in CLL

A. Clonal expansion of B cells in CLL.

B. Correlation of CpG methylation changes with treatment response.

C. Heatmap of CLL DMRs and B-cell–related DMRs.

D. Chromosomal location of CpG islands.

E. Methylation changes at specific CpG sites in Patient 17.

F. Association of methylation changes with treatment response.

G. Methylation changes at V1 and V2 regions.

Downloaded from https://bloodcancerdiscovery.aacrjournals.org by guest on July 13, 2021. Copyright 2020 American Association for Cancer Research.
pretreatment and last pre- and first posttreatment. No joint DMRs between all first pretreatment versus last pretreatment time points could be detected. The variability between samples of the same patient and the lack of shared events appear to be more in line with patient-specific evolution than a common path across patients. Moreover, correlation analysis on the CpG level also confirmed a largely stable methylome across CLL evolution and even after treatment (Fig. 2E and F).

To explore if the observed methylation stability is therapy specific, we next analyzed four patients treated with the BCL2 inhibitor venetoclax (35). As with the FCR chemoimmunotherapy-treated patients, these CLL samples collected before and after venetoclax exposure clustered tightly together, within the group of chemoimmunotherapy-treated patients (Fig. 2A and G). We further confirmed the stability of the B-cell–related and CLL DMRs in this treatment cohort, in which we could only detect globally on average less than 6% of CpGs to vary between pre- and post-venetoclax treatment (Supplementary Fig. S4A–S4D).

Combined with the early emergence of the altered methylation landscape, our posttreatment results highlight the striking stability of the CLL methylome, with minimal changes over disease progression, including after treatment.

Variations in the CLL Methylome Appear Stochastic among Patients

Because only a few patient-specific methylation dynamics were observed, we assessed if their occurrence exceeded random dynamics present among normal B-cell subtypes. Focusing on the chemoimmunotherapy-treated patient samples, we first compared the number of dynamic CpGs between first and last pretreatment, and last pre- and first posttreatment CLL samples with differences between biological replicates of naïve and memory B cells (Supplementary Fig. S5A; Supplementary Table S3). Although we did not detect any correlation with the time to treatment, we observed the fraction of dynamic CpGs to be slightly higher in posttreatment samples. Overall, only 1 of 21 patients stood out with higher variability (\( r = 0.94 \) pretreatment and \( r = 0.93 \) for pre- to posttreatment comparison); however, this is very similar to the variation observed at the transition of naïve to memory B cells (\( r = 0.95 \)). Moreover, the relative number of CpGs that exhibited substantial differences was less than 5% of all considered CpGs for most CLL cases, and those CpGs were frequently located in heterochromatin regions and outside of CpG islands (Supplementary Fig. S5B). Most of the dynamic CpGs were restricted to individual patients, and 99.9% of CpGs were shared within a maximum of six and four patients for pre- and posttreatment comparisons, respectively (Supplementary Fig. S5C), thus again confirming limited variation and a high degree of stability of CLL methylome over time.

To further separate random single dynamic position events from consistently altered regions during disease progression and following treatment, we focused on patient-specific DMRs identified either between first and last pretreatment for all 21 patients (Supplementary Fig. S5D, left), or between the last pre- and first posttreatment for the nine patients (Supplementary Fig. S5D, right). We found a median of 106 pre- and 143 posttreatment DMRs per sample. Of these, the vast majority (89% and 86%) appeared in regions shared with normal B-cell development. The remaining DMRs comprised only 8% of the aforementioned dynamic CpGs. However, with 50% of the dynamic CpGs still localized to CLL-specific regions, our data suggest the presence of randomly modulated individual CpGs rather than stretches of adjacent CpGs. Furthermore, about half of the CLL DMRs were located in heterochromatin, supporting the assumption that these may represent a secondary effect (Supplementary Fig. S5E).

Finally, gene set enrichment analysis revealed pathways supported by only a few genes (i.e., B-cell and T-cell pathways supported by three genes and p53 by two genes) or pathways with no apparent link to CLL (Supplementary Fig. S5E).

In sum, although a low number of patient-specific methylation changes accompany the individual tumor evolution, we observed remarkable stability and similarity of the acquired CLL methylome across patients.

The Altered Methylome Is Already Detectable at the MBL Stage

Because we observed an altered methylome already present within the first time points of the characterized CLL specimens, we next turned to specimens collected from our patients with MBL to evaluate the cancer precursor methylome. We again performed unsupervised phylogenetic clustering but now including the high-count MBL samples. Despite clinical classification as a precursor state, all of the MBL cases were found to cluster directly among the group of CLL samples and not to branch earlier from the trunk of this tree (Fig. 3A). Most strikingly, all five matched MBL–CLL cases appeared to be as similar to each other as the biological replicates of different healthy B-cell types and much more similar to each other than to the other CLL cases. Thus, patient-specific methylation signatures appeared to be stronger than any preleukemic versus leukemic methylation signature, which would have otherwise resulted in the separate clustering of the MBL samples from the CLL samples.

Strikingly, the identified CLL-associated DMRs were also present in our patients with MBL (Supplementary Fig. S6A and S6B). We extended the analysis to search for additional consistently occurring methylation changes that could potentially drive the MBL-to-CLL transition. However, no statistically significant DMRs could be detected between the methylomes of the individuals that transitioned from MBL to CLL. As a representative example of this shared landscape between MBL and CLL, we show the methylation patterns for the gene NFATC1, which has been reported as overexpressed in CLL due to loss of epigenetic repression (36) and is an upstream effector of BCL2, which itself is frequently deregulated due to chromosomal translocations in B-cell malignancies (Fig. 3B). Through a correlation analysis at single CpG resolution of the methylomes of the matched MBL and CLL pairs, we further observed the striking similarity between MBL and CLL methylomes. These results revealed only minor, if any, targeted remodeling of the methylome between the precursor and CLL stages within a given patient (Fig. 3C; Supplementary Fig. S6C). To appreciate this high similarity, we note that comparably high correlations are otherwise found between biological replicates of flow cytometrically isolated normal B-cell subpopulations.
Figure 3. CLL methylation signatures are already present in MBL. A, Phylogenetic tree of normal B cells, MBL, CLL (first pretreatment sample per case), and matched MBL–CLL pairs using all CpGs (n = 3.5 million). Left, all MBL cases clustered together with CLL cases. Each line represents a sample. Right, unsupervised hierarchical clustering of normal B cells and matched MBL–CLL pairs. Matched MBL–CLL cases clustered as closely as biological replicates of normal B cells; mean joint fork distance (log2 FC) of matched versus mismatched MBL–CLL versus normal B cells: −0.03 versus −1.06. B, RRBS-based genome browser tracks. Average methylation is shown as a smoothed line across the NFATC1 locus, grouped by naïve B cells (light gray, n = 3 samples), memory B cells (dark gray, unmutated (red) and mutated CLL (orange), and MBL (brown). Naïve and memory B cells are hypermethylated in the gene body, with only a small drop at the transcription start site. Unmutated and mutated CLL are hypomethylated across the promoter and gene body. C, Correlation of CpG methylation levels in matched MBL–CLL pairs (n = 5), in biological replicates of normal B cells (naïve B cells and memory B cells), at CpG-level resolution (n = 1,801,907 CpGs). Plots show a high correlation between biological replicates of normal B cells and a similarly high correlation between MBL–CLL pairs. Numbers give quantification of hypermethylated (>0.25) and hypomethylated (<−0.25) CpGs. For comparison of CpG differences with mismatched samples, the quantification of all CLL samples versus one MBL is shown. r = 0.97−0.98 for matched pairs, r = 0.85−0.9 for mismatched pairs, r = 0.99 for naive B cells, and r = 0.97 for memory B cells. D, CpG-level resolution differences between matched MBL–CLL pairs and among biological replicates of normal B cells. For all MBL–CLL comparisons, a maximum of 4.2% of positions show a difference of >0.25 (n = 1,801,907 CpGs). Methylation differences among biological replicates were less than 2% (naïve B cells) and 3.6% to 9.5% (memory B cells) of CpGs with a difference of >0.25. Per box-plot, the median value is indicated by the centerline, with first and third quartiles as outlines of boxes, and 1.5× interquartile range as whiskers; data beyond the end of the whiskers are omitted. E, Number of CpGs with a difference of >0.25 in MBL–CLL comparisons and frequency of recurrent observations across the five pairs (light gray, unique for one pair, to dark grey, observed in four of five pairs, no five of five detected). The majority of CpGs are recurrently differentially methylated. F, Comparison of average chromatin state methylation among DNA from normal B cells, MBL, and CLLs. Black horizontal lines, matched MBL–CLL pairs (n = 1,801,907 CpGs). G, Comparison of chromatin state proportions of discordant methylation rates (proportion of discordant reads, PDR) between MBL and CLLs. Black horizontal lines, matched MBL–CLL pairs.
Also, at single CpG resolution, we found that at most 4% of all covered CpGs showed a difference of 0.25 or greater between MBL and CLL samples of the same patient (Fig. 3D). An expanded analysis to examine whether individual CpGs were conserved targets across patients revealed this is not the case (Fig. 3E). Finally, we compared methylation and read-discordance levels for different chromatin states. This showed that differences and variability compared with normal B cells affect MBL and CLL cases to the same degree, again highlighting that the similarity of MBL and CLL is not merely based on patient identity (Fig. 3F and G).

Lastly, as most of our patients with MBL had already relatively elevated WBC counts, we further extended our investigation to include CD5+ sorted cells from three patients with low-count MBL (Supplementary Fig. S7A). Although the signal is expectedly not as strong due to the rare proportion of cells sequenced and the potential contamination with CD5+ normal B cells, we do detect evidence of the same characteristic epigenetic alterations as observed in high-count MBL and CLL (Supplementary Fig. S7B–S7D).

Taken together, our comprehensive analysis of 53 patients and 101 pretreatment RRBS datasets suggests that the transition to the cancer methylome occurs early in the disease. Analysis of patient-matched MBL and CLL shows that no consistent additional DNA methylation changes are seemingly associated with disease progression.

**Heterogeneous Expression Patterns Are Present Per Patient but Are Stable across Natural Disease Progression**

Complementing our methylation analysis, we profiled the transcriptomes of approximately 60k single cells isolated from healthy donors and the five matched MBL–CLL specimens (Fig. 4A). Unsupervised clustering revealed nine distinct clusters: four clusters representing peripheral blood mononuclear cells of the two healthy donors and the remaining five, with each representing one of five patient B cells (Fig. 4B). From the healthy donors, cell clusters of myeloid and lymphoid origin were readily identifiable based on their marker gene expression. In contrast, the five clusters from patients were identified as CLL/MBL-mixed clusters that showed expression of some B-cell marker genes, although less pronounced (Fig. 4C). When looking more specifically at differentially expressed genes, we found lower and heterogeneous expression for some characteristic B-cell markers and similarly heterogeneous upregulation of genes such as KLF2 and CD27 in the patient samples (Fig. 4D). Of note, the MBL and CLL cells per patient were transcriptionally indistinguishable.

Because the transcription-based clustering could not distinguish the MBL and CLL cells, we instead used barcode information per cell for annotation (Fig. 4E). We observed a remarkable overlap for most clusters, supporting the striking similarity of the MBL and CLL transcriptomes. Only the MBL and CLL cells from patient A were slightly separated in the UMAP visualization. However, upon evaluation of the highest-ranked marker genes for each MBL–CLL set, we found a surprisingly high concordance of expression of even the most differential expressed genes between MBL and CLL cells (Fig. 4F).

Based on the single-cell expression profiles, the differences between MBL and CLL appear to be marginal, which agrees with the lack of separation by clustering on single-cell transcriptomes. Combined with the lack of DNA methylation changes in the MBL-to-CLL transition, it points to an earlier molecular event that sets the normal cells already on the path to tumorigenesis.

**Individual MBL and CLL Cells Show Little DNA Methylation Heterogeneity**

Our bulk data indicated an early conserved switch in the DNA methylation landscape across patients with CLL, and our single-cell transcriptome data demonstrate the transcriptional similarity between matched MBL and CLL. However, bulk measurements cannot completely distinguish the contributions of diverse cellular subpopulations to the overall picture. Subclonal evolution and genetic heterogeneity are common in CLL (37–39). This understanding motivated us to investigate single-cell methylation maps from two patients with CLL, two patients with MBL, and age-matched B cells collected from two healthy adult volunteers, uncovering a stable level of mean methylation per cell on a global scale (Fig. 5A; Supplementary Table S1; ref. 40).

Analysis of our previously defined DMRs showed the presence of aberrant methylation levels in all MBL/CLL cells with sufficient coverage (Fig. 5B). When comparing naive to memory with MBL and CLL cells, a gradual gain of methylation in B-cell–related and CLL hyper-DMRs was observed. Conversely, hypomethylated B-cell–related and CLL DMRs appeared slightly stronger in separating normal from diseased (MBL and CLL) cells (Fig. 5B). Phylogenetic clustering separated CLL, memory B, and naïve B cells, with no differences between the B-cell subpopulations with or without the presence of CD5 (Fig. 5C; ref. 12). Moreover, we observed a clear separation between MBL and CLL versus normal, with each forming a tight cluster in line within the observed stability of the methylome per patient. Of note, memory B cells, despite many shared features with the CLL methylome, cluster distinctly next to the naïve B cells and apart from the MBL and CLL cells.

Our genome-wide single-cell methylation analysis thus complements our bulk data by further showing the clear methylation difference between MBL and CLL compared with sorted B-cell subtypes.

**DISCUSSION**

We show that the aberrant cancer methylome in CLL is already established at the preneoplastic MBL stage and is consistently present at the time of diagnosis across samples collected from 3 low-count and 20 high-count MBL, 5 matched MBL–CLL pairs, and 25 patients with CLL. Although normal B-cell maturation shows some similarities with the CLL methylome, these normal developmental changes are likely insufficient to transform cells into proliferative MBL and CLL. Nonetheless, the shared targets make a better understanding of the underlying mechanism and biological reason highly relevant. We also find a limited set of cancer-specific targets that can be readily applied to distinguish all normal B-cell subtypes from MBL and CLL. These CLL DMRs are
Figure 4. Single-cell transcriptome analysis. A, Summary table with details of donors, tissue source, number of cells per sample, and IGHV status. Patients were profiled in MBL as well as the CLL state using the 10x Genomics Chromium droplet single-cell RNA sequencing. HD, healthy donor. B, UMAP displaying the groups found using the Louvain algorithm. The healthy donor cells split into B cells, T cells, myeloid cells, and natural killer (NK) cells. The MBL and CLL cells of patients build distinct groups but are within a patient not distinguishable. C, Normalized gene expression level and the number of positive cells of marker genes used to identify normal cell types in B. B-cell–specific genes show an aberrant expression profile in clusters derived from patient cells. D, Normalized gene expression level and the number of positive cells of genes identified as marker genes between B cells and all patient cells (Wilcoxon rank-sum test). E, UMAP of MBL, CLL, and B cells with artificially introduced identification of MBL and CLL cells. MBL and CLL cells cluster farther apart from the B cells than from other cells of the same patient and are highly overlapping for almost all patients. F, Heatmap displaying single-cell gene expression of highest-ranking marker genes between MBL and CLL cells of the same patient (Wilcoxon rank-sum test). Expression levels are very similar among cells of the same patient as compared with other patients or B cells, a parameter that is also supported by hierarchical clustering.

overrepresented among pathways involved in proliferation, cell survival, and growth. Although it remains technically challenging to experimentally explore if, for instance, the addition of just these CLL DMRs alone is sufficient to drive the tumorigenic transition or facilitate extended and rapid proliferation, we anticipate that these targets are certainly worthy of future exploration, including with emerging epigenome editing tools.

Our results provide a comprehensive picture of the DNA methylation alterations in MBL and CLL and demonstrate that the switch to an abnormal landscape has consistently occurred before any of our measured time points. This notably expands findings from prior array-based studies (41, 42) and complements recent work on the genetic evolution across the 21 CLL samples (28). Similar early alterations of the methylome have also been noted in colorectal
cancer where the aberrant methylation landscape is already detectable in premalignant colorectal adenomas and amplifies upon the colorectal cancer state (43). Taken together with the near universality of the altered cancer methylome (5), the possibly conserved early emergence points to an important role for the epigenetic change in a tumorigenic transition. Although it is difficult to establish causality, we speculate that the altered landscape may provide a receptive platform for the disease progression. Alternatively, the cancer methylome may simply be a consequence of a developmental program that may regulate numerous cellular attributes, including methylation (5). In the latter case, it may well be other features driving the tumorigenic transition, and the methylome is only one biological consequence of the entire program. Although this seems possible, it should be noted that this altered DNA methylation landscape is maintained across patients sometimes for decades and found in nearly all cancer types, raising the question of why it is not diverging if it has no functional role. Another possibility that we can consider here is that the altered methylome presents an optimized epigenetic state to maintain viability with maximum proliferation and the minimal energy requirement for DNA methylation maintenance.

Aside from these considerations, we note that the MBL and CLL methylome and transcriptome are extremely stable once acquired. In contrast to the dramatic fluctuations in tumor burden (estimated by the changes in the level of WBC counts) across disease course, methylation levels are not consistently affected by clonal expansion or treatment-induced bottlenecks. The latter may reflect that cells surviving treatment represent either the average of all subclones or that limited methylation heterogeneity is present across all subclones. From a practical standpoint, the stability of the methylome in patients with CLL limits its utility to track disease progression. Still, it may be valuable for early detection and helpful to assess the efficiency of treatments. We observed neither any notable consistency of dynamic CpGs along with the MBL to CLL nor CLL progression and treatment, indicating that the few observed dynamics over the disease progression are possibly an accumulative secondary effect. During disease progression, considerable increases in WBC counts are only juxtaposed with subtle methylome changes. These largely constant methylation levels within each patient indicate that increased clonal expansion occurred without substantial additional departure from the preexisting, already aberrant landscape. The stability of the altered state is further supported by our single-cell
transcriptomes from the five patients that transition from MBL to CLL without any major expression dynamics. Finally, our finding that the cancer methylome also remains mostly unaffected by conventional chemoinmunotherapy or the BCL2 inhibitor venetoclax may keep patients at an elevated risk for relapse, in line with the fact that CLL is rarely cured, although this treatment landscape is continuously evolving (17).

Genetic and epigenetic diversity of normal tissues, tumors, and even clonally amplified cell populations have been most broadly assessed in-depth so far using bulk sequencing. To date, the degree of heterogeneity in methylation levels among distinct genomic regions within a single cell has not been investigated to our knowledge. Here, we applied single-cell methylome analysis and could show that aberrant methylation affects single cells to a surprisingly similar extent. Despite these convincing findings, it needs to be stated that due to technical limitations, such as stochastically missing values caused by unequal coverage, parts of the genome have not been investigated. Nevertheless, we could confirm that hypomethylation is more pronounced across individual MBL and CLL cells, suggesting that the methylation machinery targets a consistent and specific set of regions, though single CpGs are generally affected discordantly (20). As a result, single-cell analysis can help classify tumors and their microenvironment, which provides a more holistic disease picture and may guide more precise treatments in the future.

In sum, our comprehensive exploration over the disease course of CLL, including its precursor stage MBL, highlights several important lessons toward a better mechanistic understanding of the cancer methylome. First, the transition to the altered methylome occurs very early, possibly as a nongenetic precursor lesion that is not yet tumorigenic. Second, it should play at least a role in the current chemoimmunotherapy and BCL2 inhibition approach as it is actually effective in the majority of patients evaluated and at all stages with remarkable stability. And finally, its persistence after treatment, though currently not investigated to our knowledge, provides a more holistic disease picture and may guide more precise treatments in the future.

Finally, we converted the resulting CpG level files to bigWig files, filtering out all CpGs covered with less than five reads. An average of 34M CpGs was covered per sample at an average depth of 14×.

**Methods**

**Human Samples**

Heparinized blood samples were obtained from normal donors and patients enrolled in clinical research protocols approved by the Human Subjects Protection Committee of the Dana-Farber Cancer Institute (DFCI), at University of California, San Diego and the Mayo Clinic (CLL Research Consortium), and through the International Cancer Genome Consortium (42) after obtaining written-informed consent. Treatment indication for all 21 patients in the discovery cohort was determined based on International Workshop on Chronic Lymphocytic Leukemia criteria (12, 44). Peripheral blood mononuclear cells (PBMC) from normal donors and patients were isolated by Ficoll/Hypaque density gradient centrifugation. Mononuclear cells were used fresh or cryopreserved with 10% DMSO FBS and stored in vapor-phase liquid nitrogen until the time of analysis. CD19+ B cells from normal volunteers and CLL samples with WBC ≤50 × 10⁹/L were isolated by immunomagnetic selection (Miltenyi Biotec) and stained with anti–CD19-phycocerythrin (PE; BioLegend) prior to FACS sorting for live single cells in the presence of DAPI. MBL cells and naive and memory B cells from age-matched healthy donors were isolated as follows: Cryopreserved PBMCs were thawed and stained with anti–CD19-PE, CD5-FITC, and CD27-Allophycocyanin (APC; BioLegend). Cells were gated for naive B cells (CD19+, CD27-, and CD5), memory B cells (CD19+, CD27+, and CD5+), or MBL (CD19+ and CD5+; Supplementary Fig. S7A).

**Bulk RRBS Library Generation and Data Processing**

RRBS libraries were generated from 25 to 100 ng of input DNA using the Ovation Methyl-Seq System (NuGen) following the manufacturer’s recommendation. We used NuGen unique molecular identifier (UMI) technology to measure the rate of PCR duplicates on one patient (four samples) and found the duplicates rate to be below 2%, even at an input of only 25 ng of DNA. On average, 15.7M fragments, resulting in 31.4M paired-end 101-base pair (bp) reads, were sequenced per sample on an Illumina HiSeq2500. These reads were aligned to the human hg19 genome using BSMap (45) with flags -v 0.05 -s 16 -w 100 -S 1 -p 8 -u. An average of 21.1M reads per sample was aligned correctly. Custom scripts written in Perl were used to count the number of times a CpG was observed to be methylated. The methylation percentage for each CpG was calculated as the number of times the CpG appeared methylated divided by the total times the CpG was covered in sequencing reads. Finally, we converted the resulting CpG level files to bigWig files, filtering out all CpGs covered with less than five reads. An average of 34M CpGs was covered per sample at an average depth of 14×.

**Multiplexed Single-Cell RRBS Library Generation and Data Processing**

Single-cell RRBS libraries were prepared by combining the first steps (cell lysis and physical separation of DNA and mRNA) of the single-cell methylome and transcriptome sequencing protocol (46) with multiplexed single-cell RRBS (26) using double MspI-HaeIII digestion. Single cells were sorted into 5 μL 5′-RLT plus buffer (QIAGEN) containing 1 U SUPERase in RNase inhibitor (Invitrogen) in 96-well PCR plates, flash-frozen on dry ice, and stored at −80°C. Upon thawing, 5 μL of QIAGEN RLT plus buffer and 10 μL M-280 streptavidin beads conjugated to a biotinylated oligo-dT primer were added to each well. After 30 minutes at 25°C, the plates were transferred to a magnet to capture bead-bound mRNA, and the DNA-containing supernatant was transferred to a new 96-well plate. Beads in the original wells were washed twice with 15 μL of washing buffer (50 mmol/L Tris-HCl, pH 8, 75 mmol/L KCl, 3 mmol/L MgCl₂, 10 mmol/L EDTA, and 0.5% Tween-20), and each wash was added to the DNA plate. To clean up the DNA, 1 volume of a 1:5 dilution of AMPure XT SPRI beads (Beckman Coulter) in 20% PEG/2.5 mol/L NaCl and 20 μL Proteinase K (0.8 μg/μL, NEB) were added. After 30 minutes at 25°C with mixing, the beads were washed with 80% ethanol and genomic DNA eluted with 8 μL H₂O, with the beads remaining in the well during library prep. After addition of 2 μL 1x CutSmart buffer (NEB) containing 10 μL of MspI (NEB), or 5 U of MspI plus 5 U of HaeIII (NEB), DNA was digested for 2 hours at 37°C, followed by heat inactivation for 15 minutes at 65°C. MspI sites were filled in and fragment ends adenylated by adding 2 μL 1x CutSmart containing 2.5 U Klenow fragment (3′-5′exo-, NEB), 0.4 μL of dNTP mixture (10 mmol/L dATP, 1 mmol/L dCTP, and 1 mmol/L dGTP) followed by a two-step incubation for 25 minutes at 30°C and 30 minutes at 37°C and heat inactivation at 70°C for 10 minutes. After addition of 3 μL 1x CutSmart containing 800 U T4 DNA ligase (NEB), 0.1 μL of 100 mmol/L ATP (Roche), 1.5 μL of 0.1 μL of 1 mmol/L custom 5′-substituted and indexed (online barcode) adapter, overnight ligation at 16°C, and heat inactivation (20 minutes at 65°C), 24 separately indexed ligation reactions were pooled. After addition of 3 μL sheared and dephosphorylated Escherichia coli...
carrier DNA (27), DNA was cleaned up with 1.8 volumes of AMPure XP beads (Beckman Coulter), eluted off the beads, and bisulfite converted (EpiTect Fast Bisulfite kit, QIAGEN) following the manufacturer’s recommendations with extended conversion time (20 minutes each cycle). Each pool of RRBS libraries from 24 single cells was PCR amplified using KAPA HiFi Uracl+ DNA Polymerase, a universal P5, and a pool-specific indexed P7 primer for a total of 17 cycles. The thermoprofile was 98°C denaturation for 45 seconds, 6 cycles of 98°C for 20 seconds, 58°C annealing for 30 seconds, and 72°C extension for 1 minute, followed by 11 cycles of 98°C for 20 seconds, 65°C annealing for 30 seconds, and 72°C extension for 1 minute, and a final extension at 72°C for 5 minutes. To minimize size bias during sequencing, multiple PCR products, each representing 24 single cells, were pooled together and size selected on a 2% NuSieve agarose gel into two fractions (150–400 bp and 400–800 bp) that were sequenced in separate lanes with a 10% spike-in of a library with a balanced base composition, which is typically 2 lanes (1.5 plus 0.5 lanes for the low and high size cut, respectively) for 96 cells. On average, 4.5M fragments, resulting in 9M paired-end 75-bp reads, were generated per sample on an Illumina HiSeq4000.

Sequencing reads were demultiplexed using the inline barcode, adapters were trimmed, and reads were trimmed for quality. These reads were aligned to the human hg19 genome using BMap with flags -v 0.1 -s 12 -w 100 -S 1 −q 20 −u −R. An average of 8.4M reads (4.2M pairs) per sample was aligned. To determine the methylation state of all CpGs captured and assess the bisulfite conversion rate, we used the mcall module in the MOABS software suite with standard parameter settings (47). Finally, we converted the resulting CpG level files to bigwig files, filtering out all CpGs covered with more than 250 reads resulting in an average of 1.1M CpGs covered per sample.

10x Single-Cell RNA Library Generation and Data Processing

PBMCs were thawed in Roswell Park Memorial Institute 1640 medium supplemented with 10% FBS and centrifuged at 1,500 rpm for 5 minutes. Each sample was filtered through a 70-μm filter. Cells were resuspended in PBS-0.04% BSA and stained with anti-human CD5 (FITC), CD19 (PE), CD27 (APC), and 7-aminoacridine (Becton Dickinson). Cells were sorted through a 70-μm filter and sorted in parallel on the same day using two FacsAria II cytometers into 1.5-mL Eppendorf tubes with 10-μL PBS–0.04% BSA at a concentration of 1 × 10⁶ cells/mL. Samples from the same patient were processed and sorted in parallel on the same day using two FacsAria II cytometers (Becton Dickinson). Cells were sorted through a 70-μm nozzle into 1.5-mL Eppendorf tubes with 10-μL PBS–0.04% BSA and immediately stored on ice. Cellular suspensions were loaded on a 10x Genomics Chromium Controller platform (10x Genomics, Inc.) to generate a single-cell Gel bead in Emulsion (GEM). Single-cell RNA sequencing (scRNA-seq) libraries were prepared as previously described (51). In brief, the phylogenetic trees were inferred using the fastme.bal function in the R package ape, which is based on the minimal evolution method. Trees were computed by applying the fastme.bal function on the Euclidean distance matrix of the methylation rates of all samples in the tree. To always capture the maximal information, the subset of CpGs considered was adapted to the sample shows, resulting in n = (i) 28,343,743; (ii) 5,227,401; and (iii) 3,490,971 CpGs for (i) normal B cells, (ii) normal B cells + first time point CLL, and (iii) normal B cells + first time point CLL + MBL, normal B cells + all CLL time points, respectively.

Data Analysis

If not stated otherwise, all statistics and plots are generated using R version 3.5.1. “Feather Spray.” In all boxplots, the centerline is median; boxes, first and third quartiles; whiskers, 1.5x interquartile range; and data beyond the whiskers' end are omitted.

Bed files were processed using UCSCTools and bedtools (v2.25.0).

Additional Data

Whole genome bisulfite sequencing (WGBS) data of normal B cells were obtained from the European Genome-phenome Archive (EGA) under accession EGAS00001001196 for comparison in the phylogenetic and average methylation analysis. Methylation data were filtered for minimum coverage of 10× read coverage, and coordinates were converted to hg19 using bedtools liftOver (49). Additional WGBS data of normal B cells were obtained from Beekman and colleagues (30) and downloaded from http://resources.idibaps.org/paper/the-reference-epigenome-and-regulatory-chromatin-landscape-of-chronic-lymphocytic-leukemia. Methylation data were filtered for a minimum of 10× read coverage, and coordinates were converted to hg19 using bedtools liftOver. Data were used in phylogenetic comparisons, average methylation analysis, and all comparative analysis, e.g., detecting differential methylated regions and genomic region visualizations. To ensure accurate comparison among samples, all published WGBS data were reduced to positions covered in any of our patient RRBS data, resulting in a set of approximately 5 million comparable CpGs. Single-cell RRBS and RNA-seq data of the two CLL samples were obtained from Gaiti and colleagues (40).

Feature Annotations

Chromatin states were defined by the standard 15-state model using the ChromHMM algorithm (33) and were downloaded from the UCSC Genome Browser (50). The average methylation rate for each sample and per chromatin state was calculated as the mean of all methylation rates overlapping with a particular chromatin state. CpG islands were downloaded from the UCSC Genome Browser; shores were defined as adjacent 2 kb regions and shelves as the next adjacent 2 kb regions. Gene annotations were downloaded from the UCSC Genome Browser (gencode v19), and promoter regions were defined as 5,000 nt upstream to 2,500 nt downstream of annotated transcription start site. DMRs were assigned to genes if overlapping with the promoter or gene body with at least one shared base. For unique annotation of DMRs and if a DMR overlapped more than one feature, the ranking was: promoter, gene body, last intergenic; or CpG island, shelf, and last shore. For chromatin state annotation of DMRs, the 15 chromatin states were collapsed into the 6 main categories (Active Promoter, Posed Promoter, Enhancer, Transcription, Insulator, and Heterochromatin), and each DMR was assigned to the region with its maximal overlap.

Mutagen and subclone information was taken from ref. 28.

Phylogenetic Tree

The phylogenetic analysis of DNA methylation was performed as previously described (51). In brief, the phylogenetic trees were inferred using the fastme.bal function in the R package ape, which is based on the minimal evolution method. Trees were computed by applying the fastme.bal function on the Euclidean distance matrices of the methylation rates of all samples in the tree. To always capture the maximal information, the subset of CpGs considered was adapted to the sample shows, resulting in n = (i) 28,343,743; (ii) 5,227,401; and (iii) 3,490,971 CpGs for (i) normal B cells, (ii) normal B cells + first time point CLL, and (iii) normal B cells + first time point CLL + MBL, normal B cells + all CLL time points, respectively.

Scatter Plots and Correlation

Scatter plots were created using the smoothScatter function of R, and correlations were calculated using the cor function of R. For the first figure, the average methylation per group was used (n = 3 for naïve and memory B cells, n = 20 and 21 for unmutated and mutated CLL, respectively). Missing values were removed from the mean calculation. For the matched MBL–CLL correlation, missing values were
not omitted from the naive, memory, and the five matched MBL and CLL samples, resulting in 1,801,907 CpGs that were used for correlation analysis and scatter plots. Statistics for the full CLL cohort are given in Supplementary Fig. S3A.

**Genome Region Visualization**

Visualization of methylation levels per CpG at genomic regions was done using the plotTracks function of the R package Gviz (52). Track data were grouped by cell type, and for the curve, representation was plotted as a smoothed line.

**Differential Methylation Analysis**

DMRs were called using metilene version 0.2–7 (32). DMRs were defined to have an absolute minimum difference in methylation of 0.25 with a maximum distance of 100 nt between CpGs within a DMR and a minimum of 3 CpGs per DMR (parameter $-M100 -m 3 -d 0.25$). DMRs were calculated between (i) the normal B cells from Beekmann and colleagues (30), (ii) CLL samples as well as between the first and last pretreatment and the last pre- and first posttreatment time points, and (iii) for each sample individually between the first and last pretreatment and the last pre- and first posttreatment time points. More specifically, DMRs were calculated for normal B cells to CLL between the naive B cells ($n = 3$) and the first time point of the CLL samples with unmutated IGHV as well as between the memory B cells ($n = 3$) and the first time point of the CLL samples with mutated IGHV. Only positions on the autosomes (chr1–22) were taken into account that were covered by all three normal B-cell samples (naïve or memory) and 90% of the CLL samples ($n = 18$ unmutated/n = 19 mutated CLL), respectively. For within-patient sample versus sample DMRs, all positions that were covered by both samples were taken into account. After DMR calling, all P values were adjusted for multiple testing using the R function p.adjust, and regions with an adjusted $P$ value < 0.05 were considered DMRs. As previously described, DMRs were separated into DMRs that overlap CpGs that are dynamic during B-cell differentiation (difference between all normal B cells > 0.25) and subsequently called B-cell-related DMRs, and those that do not overlap any dynamic position are called CLL DMRs. No DMRs (FDR < 0.05) were found when comparing the five matched MBL samples with the respective five CLL samples, when comparing the unmutated and mutated CLL, or when comparing the first pretreatment time points to the last pretreatment time points.

For heatmap visualization, methylation levels of DMRs were calculated as the mean methylation of all CpGs with a DMR for all samples and plotted using the heatmap function of the R package ComplexHeatmap (53). Row annotations were based on overlap with features; see Feature Annotations section above. The enrichment analysis of genes affected by DMRs was done using the online Web tool WebGestalt (54). An overrepresentation enrichment analysis (ORA) was calculated for the CLL DMRs with all DMRs as background for Panther pathways and the patient-specific DMRs by comparing recurrently hit genes among more than two patients with all DMR genes.

**Single-Cell Analysis.** For comparison of the RRBS single-cell experiments, only positions of the double-digest data (naïve and memory B cells) were considered that were also covered by the single-digest data (CLL).

**10x scRNA-seq.** The single-cell RNA was analyzed using the python toolkit “Scapan” with default parameters for clustering and UMAP generation (55). Gene expression profiles were generated using parameters for normalized gene expression representation for dotplot and heatmap representations.
curation, methodology. S.L. Slager: Resources, data curation. E. Braggio: Resources, data curation. T.D. Shanafelt: Resources, data curation. N.E. Kay: Resources, data curation. S.M. Fernandes: Resources, data curation, methodology. J.R. Brown: Resources, methodology. L. Wang: Resources, methodology. S. Li: Resources, methodology. K.J. Livak: Resources, methodology. D.S. Neuberg: Resources, methodology. S. Klages: Resources, methodology, project administration. B. Timmermann: Supervision, funding acquisition, project administration. T.J. Kipps: Supervision, funding acquisition. E. Campo: Supervision, funding acquisition, methodology, project administration. A. Gnirke: Conceptualization, resources, supervision, funding acquisition, investigation, methodology, writing–original draft, project administration, writing–review and editing. C.J. Wu: Conceptualization, resources, supervision, funding acquisition, visualization, methodology, writing–original draft, project administration, writing–review and editing. E. Braggio: Conceptualization, resources, supervision, funding acquisition, investigation, methodology, writing–original draft, project administration, writing–review and editing. A. Meissner: Conceptualization, resources, supervision, funding acquisition, visualization, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

The authors thank members of the Meissner lab, in particular Jocelyn Charlton, and both Erin M. Parry and Inaki Subero-Martin for critical reading of the article. The authors thank Jerome Ritz and the DFCI Pasquerruel Tissue Bank in Hematologic Malignancies for the prospective collection and processing of blood samples from healthy donors. This work was supported in part by the NCI (5P01CA081534-14). A. Biran was supported by the Lymphoma Research Foundation. C.K. Lemvigh acknowledges support from the Fishman Family Fund. M. Gruber received a Marie-Curie Outgoing Fellowship from the European Union (PIOF-2013-624924). T.D. Shanafelt and N.E. Kay were supported by the NIH (R01CA197120). E. Campo is supported by Instituto de Salud Carlos III, Madrid Spain (PIMP15/00007), “La Caixa” Foundation (grant CLEVELE-VRH17-00221), Health Research 2017, and European Research Council (ERC) BCLAtlantis - 810287 and is a Researcher of the “Institució Catalana de Recerca i Estudis Avançats” (ICREA) of the Generalitat of Catalunya. S.L. Slager and E. Braggio were supported by R01 CA235026. J.R. Brown is supported by NCI RO1 CA 213442 and the Rosenbach Fund for Lymphoma Research. S. Li is supported by the NCI Research Specialist Resources, methodology. Resources, data curation. J.R. Brown: Resources, data curation. N.E. Kay: Resources, data curation. A. Meissner: Conceptualization, resources, supervision, data curation, methodology. B. Timmermann: Supervision, funding acquisition, project administration. A. Gnirke: Conceptualization, resources, supervision, funding acquisition, investigation, methodology, writing–original draft, project administration, writing–review and editing. A. Meissner: Conceptualization, resources, supervision, funding acquisition, visualization, methodology, writing–original draft, project administration, writing–review and editing.

Received November 11, 2019; revised October 28, 2020; accepted November 18, 2020; published first December 3, 2020.

REFERENCES

The Altered DNA Methylation Landscape in CLL

Preneoplastic Alterations Define CLL DNA Methylome and Persist through Disease Progression and Therapy


Updated version
Access the most recent version of this article at:
doi: 10.1158/2643-3230.BCD-19-0058

Supplementary Material
Access the most recent supplemental material at:
http://bloodcancerdiscov.aacrjournals.org/content/suppl/2020/11/26/2643-3230.BCD-19-0058.DC1

Cited articles
This article cites 54 articles, 5 of which you can access for free at:
http://bloodcancerdiscov.aacrjournals.org/content/2/1/54.full#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
http://bloodcancerdiscov.aacrjournals.org/content/2/1/54.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link:
http://bloodcancerdiscov.aacrjournals.org/content/2/1/54.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.