

ARTICLE

<https://doi.org/10.1038/s41467-019-09645-5>

OPEN

Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL

Alessandro Pastore^{1,10}, Federico Gaiti^{2,3,4,10}, Sydney X. Lu¹, Ryan M. Brand^{2,3,4}, Scott Kulm³, Ronan Chaligne^{2,3,4}, Hongcang Gu⁵, Kevin Y. Huang^{2,3,4}, Elena K. Stamenova⁵, Wendy Béguelin³, Yanwen Jiang³, Rafael C. Schulman^{2,3,4}, Kyu-Tae Kim^{2,3,4}, Alicia Alonso³, John N. Allan³, Richard R. Furman³, Andreas Gnirke⁵, Catherine J. Wu^{5,6}, Ari M. Melnick³, Alexander Meissner^{5,7}, Bradley E. Bernstein^{5,8}, Omar Abdel-Wahab^{1,9,10} & Dan A. Landau^{2,3,4}

Cancer evolution is fueled by epigenetic as well as genetic diversity. In chronic lymphocytic leukemia (CLL), intra-tumoral DNA methylation (DNAm) heterogeneity empowers evolution. Here, to comprehensively study the epigenetic dimension of cancer evolution, we integrate DNAm analysis with histone modification mapping and single cell analyses of RNA expression and DNAm in 22 primary CLL and 13 healthy donor B lymphocyte samples. Our data reveal corrupted coherence across different layers of the CLL epigenome. This manifests in decreased mutual information across epigenetic modifications and gene expression attributed to cell-to-cell heterogeneity. Disrupted epigenetic-transcriptional coordination in CLL is also reflected in the dysregulation of the transcriptional output as a function of the combinatorial chromatin states, including incomplete Polycomb-mediated gene silencing. Notably, we observe unexpected co-mapping of typically mutually exclusive activating and repressing histone modifications, suggestive of intra-tumoral epigenetic diversity. Thus, CLL epigenetic diversification leads to decreased coordination across layers of epigenetic information, likely reflecting an admixture of cells with diverging cellular identities.

¹Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York 10065 NY, USA. ²New York Genome Center, New York 10013 NY, USA. ³Weill Cornell Medicine, New York 10021 NY, USA. ⁴Institute for Computational Biomedicine, Weill Cornell Medicine, New York 10021 NY, USA. ⁵Broad Institute of MIT and Harvard, Cambridge 02142 MA, USA. ⁶Dana-Farber Cancer Institute, Boston 02215 MA, USA. ⁷Max Planck Institute for Molecular Genetics, Berlin 14195, Germany. ⁸Department of Pathology, Massachusetts General Hospital, Boston 02114 MA, USA. ⁹Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York 10065 NY, USA. ¹⁰These authors contributed equally: Alessandro Pastore, Federico Gaiti. Correspondence and requests for materials should be addressed to O.A.-W. (email: abdelwao@mskcc.org) or to D.A.L. (email: dlandau@nygenome.org)

Cancer growth, progression, and relapse are the result of an evolutionary process fueled by intra-tumoral diversity^{1–3}. Chronic lymphocytic leukemia (CLL)—a common B cell malignancy—serves as a highly informative model for cancer evolution as it undergoes substantial genetic diversification⁴ and evolution with therapy⁵.

In addition to genetic changes, the CLL epigenome is an important disease-defining feature linked to its cell-of-origin and is predictive of outcome^{6–8}. In fact, the stable propagation of the ancestral epigenome allowed the use of DNA methylation (DNAm) patterns to precisely retrace the initially transformed cell-of-origin from which different CLLs emerge⁸. In addition to the largely stably inherited epigenome, we have previously shown that growing CLL populations also undergo ongoing somatic DNAm changes akin to the process of genetic diversification through ongoing mutations, leading to high intra-leukemic epigenetic heterogeneity, greater clonal evolution, and adverse outcome⁹, as has been shown for other malignancies¹⁰.

However, DNAm constitutes only a single layer of the epigenetic information encoding cell identity. Given the importance of histone modifications to lineage plasticity in cancer^{11,12}, we reasoned that intra-leukemic epigenetic heterogeneity may extend to histone modifications, likely promoting lineage plasticity by enabling permissive chromatin states. To address this question, we complemented bulk reduced representation bisulfite sequencing (RRBS) analysis with a chromatin immunoprecipitation sequencing (ChIP-seq) compendium of histone post-translational modifications and gene expression, together with joint DNAm and transcriptome single cell analysis in a cohort of 22 primary CLL and 13 healthy B lymphocytes samples. Our integrative analysis revealed a markedly decreased coordination between different layers of the CLL epigenome, whereby ongoing epigenetic diversification leads to an admixture of cells with diverging epigenetic identities, thus providing a novel perspective into the epigenetic dimension of cancer evolution.

Results

Super-enhancer and associated DNAm alteration in CLL. To comprehensively study the epigenetic landscape of evolving CLL and its relationship to intra-leukemic diversity, we generated genome-wide maps of histone marks with non-overlapping regulatory functions (H3K4me3, H3K27ac, and H3K27me3) and transcriptome sequencing (bulk RNA-seq) in a cohort of 20 primary *IGHV* mutated and unmutated CLL (corresponding to the major known disease subtypes¹³; $n = 14$ and $n = 6$, respectively), as well as 12 healthy B lymphocytes samples (CD19/CD23/IgD-positive CD27-negative tonsillar naïve B cells [NBCs; CD19⁺CD23⁺CD27⁻IgD⁺], $n = 2$; peripheral blood NBCs [CD19⁺CD23⁺CD27⁻IgD⁺], $n = 4$; CD19/CD23/CD27-positive IgD-negative tonsillar germinal center B cells [GCBs; CD19⁺CD23⁺CD27⁺IgD⁻], $n = 2$; peripheral blood memory B cells [GCBs; CD19⁺CD23⁺CD27⁺IgD⁻], $n = 3$; CD20-positive tonsillar B cells [CD20⁺], $n = 1$; Supplementary Fig. 1a, b).

Analysis of H3K27ac, a histone modification known to be a marker of active gene regulatory regions¹⁴, revealed core enhancer and super-enhancer (as defined in ref. 14; see Methods; Supplementary Fig. 1c) reprogramming in CLL. A total of 297 super-enhancers were differentially regulated in CLL compared with normal B cells (absolute \log_2 [H3K27ac fold-change] >2 and Wald test BH-FDR <0.01; see Methods), with increased H3K27ac in proximity to genes critical for lymphocyte proliferation and differentiation, including *BCL2*, *LEF1*, and *CTLA4*^{15–17} (Fig. 1a–c; Supplementary Fig. 1d) and involved in pathways previously reported to play key roles in CLL (e.g., B cell receptor, NF- κ B and MAPK inflammatory signaling pathways³; Fig. 1d). As ChIP-seq

experiments are prone to technical variation, we further demonstrated the reproducibility of H3K27ac derangements in CLL by analyzing additional CLL and normal B cell samples from the Blueprint Initiative¹⁸ (Supplementary Data 1), showing high pairwise correlations across our cohort and the Blueprint initiative samples at super-enhancers (Supplementary Fig. 1e). Fewer differences in the super-enhancer landscape were observed between the two major known CLL subtypes (*IGHV* mutated and unmutated; $n = 27$ super-enhancers differentially regulated; Supplementary Fig. 1f; see Methods), and with chromosome 13q deletion (del(13)q; $n = 25$ super-enhancers differentially regulated; Supplementary Fig. 1g; see Methods), consistent with previous studies showing more subtle chromatin differences between CLL subtypes¹⁹. In line with prior studies that profiled epigenomic features of a large CLL cohort and discrete normal B cell subtypes along the differentiation program^{8,19}, this extensive chromatin rewiring at super-enhancers is mediated by specific transcription factors, as evidenced by enrichment of their motifs in activated super-enhancers, including *NFAT*, a deregulated gene with functional and therapeutic potential in CLL⁸, and *TCFL2*, a downstream target of the WNT pathway overexpressed in CLL²⁰ (Fig. 1e, f; Supplementary Data 2, 3).

DNAm changes at enhancers and super-enhancers impact their transcriptional activity²¹. Therefore, to assess the relationship between DNAm and enhancers, we profiled bulk DNAm of normal B cell populations (peripheral blood naïve B cells [CD19⁺CD23⁺CD27⁻IgD⁺], $n = 3$; peripheral blood memory B cells [CD19⁺CD23⁺CD27⁺IgD⁻], $n = 2$) and CLL patient samples (*IGHV* unmutated, $n = 2$; *IGHV* mutated, $n = 3$) using a targeted bisulfite sequencing capture assay, which preferentially evaluates dynamic CpGs at gene-regulatory elements²² (Supplementary Fig. 1a, b; Supplementary Fig. 2a, b; Supplementary Data 4, 5). Consistent with prior reports^{7,9,23}, we observed a global decrease in DNAm in CLL compared with normal B samples (Supplementary Fig. 2c, left; Supplementary Fig. 2d), with a focal increase in methylation of CpG islands (CGI; Supplementary Fig. 2c, right).

In addition, we identified 41,057 differentially methylated regions (DMRs; absolute change in DNAm >0.3 and Fisher's exact test FDR <0.05²²; see Methods) between CLL and normal B samples, most of which were hypomethylated in CLL (Supplementary Fig. 2e; Supplementary Data 6–8). Interestingly, hypomethylation preferentially affected H3K27ac-enriched regions, including super-enhancers (Fisher's exact test $P < 0.0001$; Fig. 2a; Supplementary Fig. 2e, f). This extensive focal hypomethylation at super-enhancers was observed in proximity to genes involved in pathways previously reported to play key roles in CLL (e.g., B cell receptor activation, Notch signaling, and cell proliferation³; Supplementary Fig. 2g; Supplementary Data 9). Additionally, CLL-specific super-enhancers showed a strong decrease in DNAm compared to normal B samples (Mann–Whitney U -test, $P < 0.0001$; Fig. 2b, c), as illustrated for the *BCL2* gene locus (Fig. 2d). In contrast, super-enhancers that become inactive in CLL did not gain DNAm compared to normal B samples (Mann–Whitney U -test, $P > 0.05$; Fig. 2b–d), supporting the concept that DNAm is slow to accumulate with CLL progression^{19,24}.

Notably, we observed that hypomethylation at super-enhancers resulted preferentially in intermediate DNAm levels in CLL (Fisher's exact test $P < 0.0001$; Fig. 2e; Supplementary Fig. 2h). These data demonstrated that cancer-associated hypomethylation is not limited to previously described intermediately methylated blocks in heterochromatin and lamina associated domains^{25,26}, but may also involve regions of active chromatin.

Decreased epigenetic-transcriptional coordination in CLL. The observed intermediate bulk DNAm patterns at H3K27ac

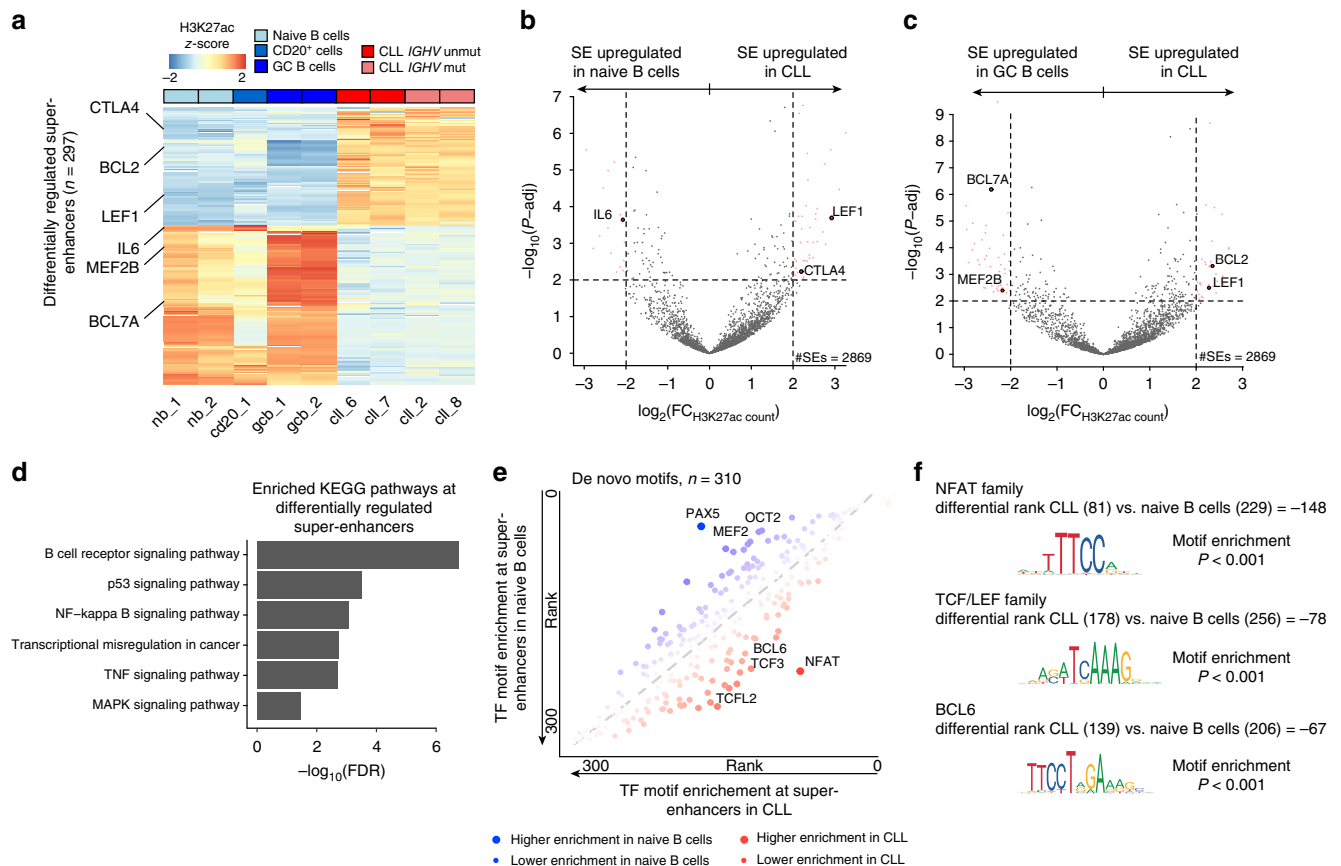


Fig. 1 Super-enhancer rewiring in CLL. **a** H3K27ac profiles of 297 differentially regulated super-enhancers (absolute $\log_2[\text{H3K27ac fold-change}] > 2$ and Wald test BH-FDR < 0.01) between CLL and normal B cells. Red indicates high H3K27ac level, blue low H3K27ac level. Gene assignment to super-enhancers based on proximity (average distance of 2775 bp). Genes critical for lymphocyte proliferation and differentiation are highlighted. **b** Differential H3K27ac at super-enhancers ($n = 2869$) between CLL and naïve B cells (same samples as in **a** were used in the analysis). Differentially regulated super-enhancers defined as having absolute $\log_2(\text{H3K27ac fold-change}) > 2$ and Wald test BH-FDR < 0.01 . **c** Same as **b** for differential H3K27ac at super-enhancers ($n = 2869$) between CLL and germinal center B cells. **d** KEGG pathways enriched at differentially regulated super-enhancers ($n = 297$; hypergeometric test BH-FDR < 0.01). **e** Comparison of rank in transcription factor de novo motif enrichment ($n = 310$, ranked by hypergeometric test P -value) between CLL (x-axis) and naïve B cells (y-axis) at super-enhancers. Critical TF motifs for lymphocyte proliferation and B cell differentiation are highlighted. LOESS regression line of observed ranked P -values is shown in dotted gray. Color: red—CLL biased; blue—naïve B cell biased, size adjusted based on residual. **f** Position weight matrices of selected de novo TF motifs significantly over-represented in CLL (standardized residuals > 1.65). For each motif, the differential ranked P -value between CLL and naïve B cells, motif enrichment hypergeometric test P -value, and the best match to JASPAR core database are shown

regulatory regions are reminiscent of our previous observation of intermediate DNAm in promoters stemming from stochastic DNAm intra-leukemic diversification during CLL evolution⁹. Therefore, to examine whether enhancer rewiring is also associated with disordered methylation leading to reduced coordination between DNAm and H3K27ac, we drew on a well-established metric in information theory—mutual information (MI)—which measures how much can be learned from one variable about another (see Methods). Consistent with disrupted coordination across these two layers of the CLL epigenome, we observed lower pairwise MI between bulk DNAm and H3K27ac in CLL samples (irrespective of their *IGHV* mutational status) compared with normal B cell samples at super-enhancer regions (Welch’s t -test, $P < 0.0001$; Fig. 3a; Supplementary Fig. 3a). This decrease in MI also corresponded to a weaker correlation between DNAm and H3K27ac in CLL samples compared with normal B cell samples at super-enhancers (linear regression R^2 of 0.558 normal B vs. 0.471 CLL samples, t -test $P < 0.0001$).

The decrease in MI was observed more broadly, including a 13% decrease in MI between DNAm at transcription start sites (TSSs) and gene expression, in CLL relative to normal B cells

(Supplementary Fig. 3b). This decrease in MI may result from greater intra-leukemic cell-to-cell heterogeneity that is not captured in bulk population sequencing assays. To directly test this hypothesis, we performed joint single-cell DNAm sequencing and whole transcriptome sequencing on additional normal B and CLL samples ($n = 96$ cells [1 sample], $n = 288$ cells [2 samples], respectively; Fig. 3b; Supplementary Fig. 3c). While MI was higher across samples in matched vs. scrambled single-cell DNAm and RNA-seq data (paired t -test, $P < 0.0001$), the matched single-cell MI increase was higher in CLL compared with normal B cells ($43 \pm 5.2\%$ vs. $29 \pm 4.8\%$, respectively; Mann–Whitney U -test, $P = 0.036$; Fig. 3c). These data suggest that, at least in part, the decreased epigenetic-transcriptional coordination observed in CLL is the result of cell-to-cell epigenetic diversification.

To more broadly examine the relationship between epigenetic states (i.e., combinatorial interactions of epigenetic marks) and transcriptional output, we modeled the combinatorial patterns of histone modifications (H3K4me3, H3K27ac, H3K27me3) and DNAm (based on bulk bisulfite sequencing), with or without gene expression (based on bulk RNA-seq), using a Dirichlet

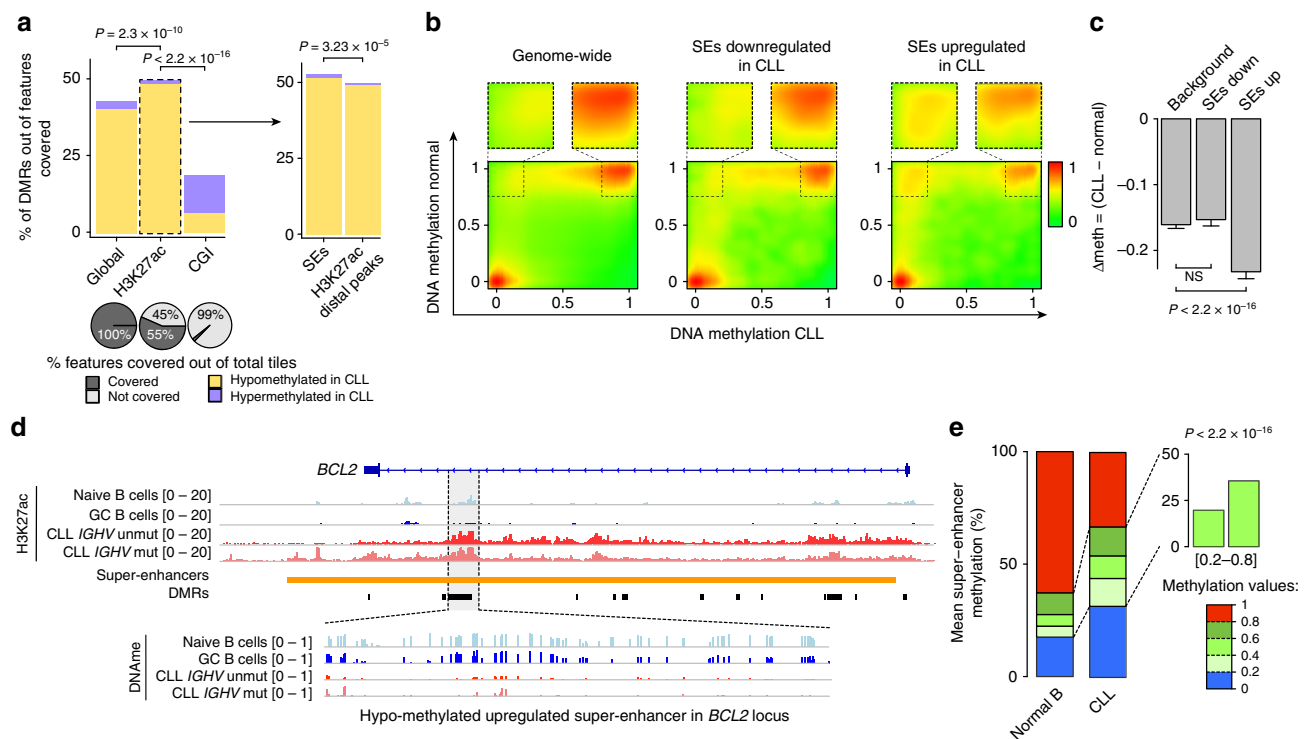


Fig. 2 DNA methylation alteration at super-enhancers in CLL. **a** Percentage of differentially methylated regions (DMRs) measured with targeted bisulfite sequencing capture assay in (i) Global: all covered 500bp-tiles in the genome; (ii) H3K27ac: union of all H3K27ac peaks; (iii) CGI: CpG islands; (iv) SE: union of all super-enhancers; (v) H3K27ac distal peaks: union of all H3K27ac peaks that did not overlap TSSs (± 2.5 kb). P -values are shown for two-sided Fisher's exact test. Bottom: percentage of distinct genomic features covered by the targeted bisulfite sequencing capture assay. **b** CpG methylation for CLL and normal B cells across all super-enhancers (left), super-enhancers upregulated in normal B cells (center), and super-enhancers upregulated in CLL (right), measured with targeted bisulfite sequencing capture assay. **c** Difference in mean CpG methylation between CLL and normal B cells for the three categories shown in **b**. Error bars represent 95% confidence interval. P -values are indicated for two-sided Mann-Whitney U -test. NS, not significant. **d** Epigenomic profiling of the *BCL2* locus in CLL compared with normal B cells. **e** The percentage of CpG methylation values at super-enhancers in CLL (no. of CpGs used = 468,303) and normal B cells (no. of CpGs used = 502,607), measured with targeted bisulfite sequencing capture assay. P -value is shown for two-sided Fisher's exact test for the intermediate category [0.2–0.8].

Process Mixture (DPM) approach, which allows learning de novo the number of combinatorial states. We observed a significantly higher number of states across CLL samples (in both *IGHV* mutated and unmutated samples), compared with normal B cells when adding RNA information into the DPM analysis, indicating that the transcriptional output of epigenetic states is less uniform in CLL (Mann–Whitney U -test, $P < 0.0001$; Fig. 3d; Supplementary Fig. 3d). Specifically, while H3K27me^{3hi}/H3K4me^{3low}/H3K27ac^{low}-marked genes ($n = 904$) were associated with uniform gene silencing in B cells, they were associated with variable expression in CLL (Mann–Whitney U -test, $P < 0.0001$; Fig. 3e), suggesting that disrupted Polycomb repression in CLL results in leaky silencing allowing partial reactivation of these genes. Notably, variable expression of H3K27me^{3hi}/H3K4me^{3low}/H3K27ac^{low}-marked genes preferentially affected genes related to the critical B-cell receptor (BCR) signaling pathway (Supplementary Fig. 3e). In addition, we observed enrichment of specific transcription factor binding motifs in H3K27me^{3hi}/H3K4me^{3low}/H3K27ac^{low}-marked regions, including *NFA17*⁶ and *MYB*, a proto-oncogene overexpressed in CLL²⁷ (Hypergeometric test $P < 0.0001$; Supplementary Fig. 3f). These data suggest that CLL epigenomes are associated with less uniform transcriptional outputs compared with normal B cell epigenomes.

Transcriptional variation in genes with similar epigenetic states may stem from cell-to-cell transcriptional heterogeneity. To test this, we computed gene expression information entropy, a measure of cell-to-cell gene expression heterogeneity⁹ in our

single-cell whole transcriptome data, and found that in CLL single cells ($n = 94$) H3K27me^{3hi}/H3K4me^{3low}/H3K27ac^{low}-marked genes were indeed associated with significantly higher intra-leukemic expression information entropy compared to normal B cells ($n = 84$), or compared to a set of genes with matched mean expression but not marked by H3K27me^{3hi}/H3K4me^{3low}/H3K27ac^{low} (Mann–Whitney U -test, $P = 0.0003$ and $P = 0.005$, respectively; Fig. 3f, g; Supplementary Fig. 3g). Our data therefore suggest a model in which H3K27me³-marked genes in CLL are incompletely silenced, resulting in greater cell-to-cell transcriptional heterogeneity.

Corrupted coherence across layers of the CLL epigenome. An alternative approach to assess the coordination between layers of the epigenome involves capturing their overlapping and mutually exclusive combinatorial patterns²⁸. We pursued this orthogonal approach by training a multivariate Hidden Markov Model (HMM) on CLL and normal B cells data based on three of the different histone modifications (H3K4me³, H3K27ac, H3K27me³), DNAm (based on bulk bisulfite sequencing), and gene expression information (based on bulk RNA-seq). We identified 12 distinct epigenetic states that fell into two broad categories. First, a category that correlated with active transcription including active promoters (“Active flanking TSS”, “TSS”), enhancers (“Enhancer”, “H3K4me³/H3K27ac”), and 5' and 3' boundaries of transcribed genes (“I–IV transcription”). Second, a category of genes with no or little detectable transcription, including bivalent or poised (“Bivalent/

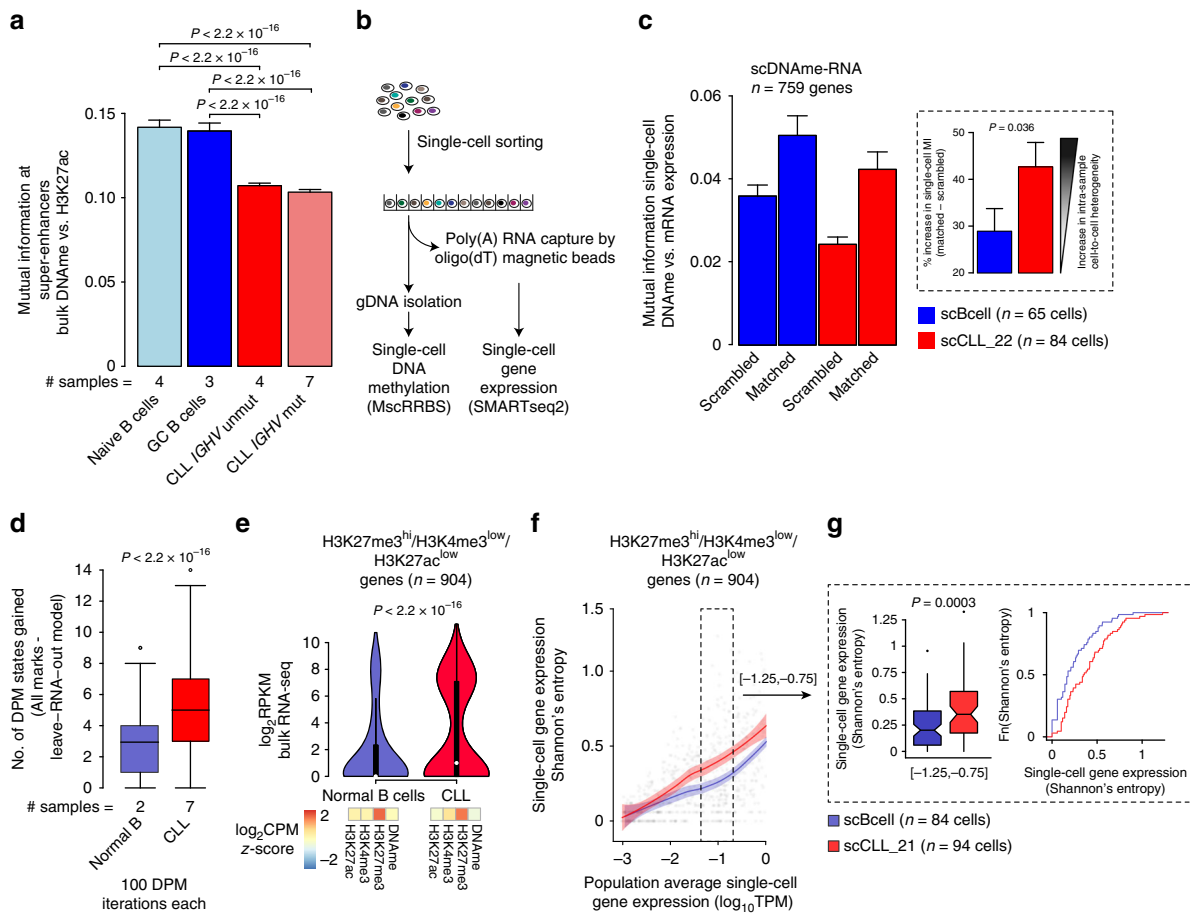


Fig. 3 Decreased epigenetic-transcriptional coordination in CLL. **a** Mutual information (MI) across super-enhancers ($n = 2869$) between bulk DNAm (measured with targeted bisulfite sequencing capture assay) and H3K27ac for CLL and normal B samples. P -values are indicated for two-sided Welch's t -test. **b** Schematic of the experimental procedures of the joint multiplexed single-cell RRBS and transcriptomics analysis. **c** MI between expression and promoter single-cell DNAm rate ($n = 759$ genes with sufficient RNA [expression seen in >5 cells] and DNAm >10 CpGs per promoter] information) in individual CLL cells ($n = 84$) and normal B cells ($n = 65$). The observed MI values were compared to values obtained by randomly permuting cell labels for the methylation values. Inset: percentage increase in MI when comparing matched vs. scrambled single-cell DNAm and RNA data in CLL and normal B cells. P -value is indicated for two-sided Mann-Whitney U -test. **d** Number of states gained when adding RNA data to the epigenomic mapping (H3K4m3, H3K27ac, H3K27me3, DNAm based on bulk bisulfite sequencing) in the DPM analysis for CLL and normal B samples. P -value is indicated for two-sided Mann-Whitney U -test. See also Supplementary Fig. 3d. **e** Expression levels (\log_2 [RPKM]) of 904 genes marked by H3K27me3^{hi}/H3K4me3^{low}/H3K27ac^{low} between normal B samples (blue) and CLL (red) samples ($n = 2$ and $n = 7$, respectively). P -value is shown for two-sided Mann-Whitney U -test. **f** Single-cell gene expression Shannon's information entropy in relation to the population average gene expression (based on single-cell whole transcriptome data) in CLL ($n = 94$) and normal B cells ($n = 84$). Local regression lines for the H3K27me3^{hi}/H3K4me3^{low}/H3K27ac^{low}-marked genes in CLL (red) and normal B cells (blue) are shown. **g** Single-cell gene expression Shannon's information entropy (left) for genes within a population average expression range of -1.25 to -0.75 \log_{10} (TPM) (to control for differences in this variable). P -values are shown for two-sided Mann-Whitney U -test. Cumulative distribution (right) showing the proportion of intermediate single-cell gene expression Shannon's information entropy values at H3K27me3^{hi}/H3K4me3^{low}/H3K27ac^{low}-marked genes. Throughout the figure: boxplot represents median and bottom and upper quartile; whiskers correspond to $1.5 \times$ IQR; error bars represent 95% confidence interval

Poised TSS”), repressed Polycomb (“PRC”), and mCpG-rich (“mCpG”) states (Fig. 4a).

CLL overall showed high resemblance to normal B cells, and no significant differences in genomic coverage were observed between IGHV mutated and unmutated CLL (Supplementary Fig. 4a). Importantly, HMM analysis revealed a chromatin state simultaneously marked by H3K27ac and H3K27me3, modifications which are typically mutually exclusive, with a >2 -fold enrichment in CLL compared with normal B cells (Hypergeometric test $P < 0.0001$; Fig. 4a, b), and affecting ~ 1.6 M 200 bp genomic segments including non-first introns and distal regulatory elements (Supplementary Fig. 4b). We further validated this chromatin state by analyzing an additional 6 CLL and 7 normal B cell samples from the Blueprint Initiative¹⁸

and obtained high pairwise correlation (Spearman's rho correlation coefficient = 0.47) between H3K27ac and H3K27me3 marks at H3K27ac-H3K27me3 segments identified in our data (Supplementary Fig. 4c). Evaluation of these H3K27ac-H3K27me3 segments from CLL revealed that a notable fraction (46.7%) of these regions possessed repressive chromatin modifications in normal B cells (Fig. 4c), suggesting these are genomic regions that are subject to CLL-specific activation by gaining activating acetylation marks (H3K27ac). Gaining H3K27ac in the transition from a healthy to disease state may be associated with upregulation of neighboring genes. Consistent with this scenario, RNA gene expression was increased in proximity to regions that gain H3K27ac in CLL (Fig. 4d; Supplementary Fig. 4d). A gene set enrichment

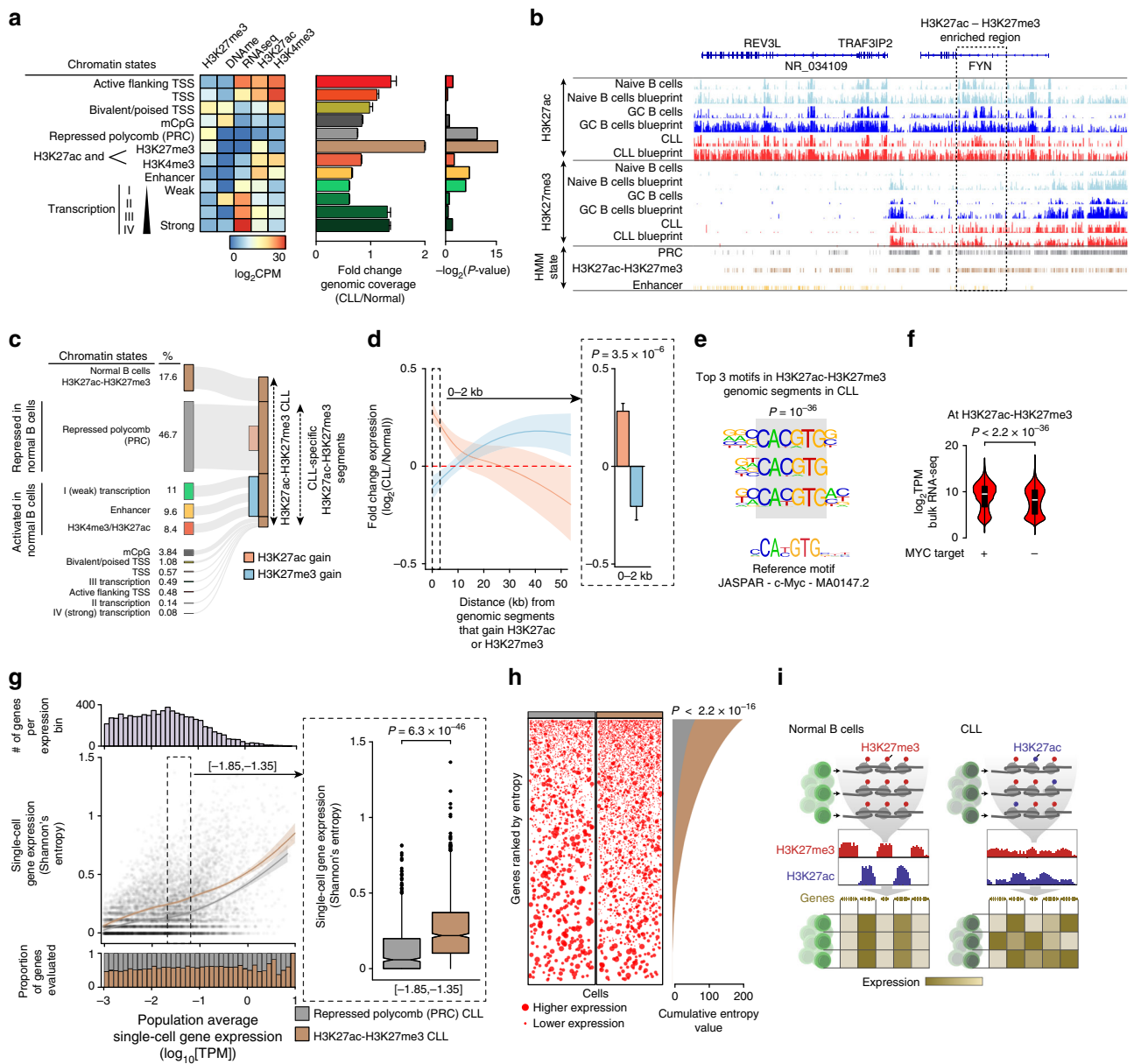


Fig. 4 Corrupted coherence across layers of CLL epigenome leads to cell-to-cell transcriptional heterogeneity. **a** Chromatin state definitions and enrichments for a 12-state Hidden Markov Model based on three histone marks (H3K4me3, H3K27ac, H3K27me3), DNase, and RNA information. *P*-values of a given HMM state between CLL and normal B cells are shown for two-sided hypergeometric test. **b** Epigenomic profiling of the *FYN* gene locus, demonstrating "H3K27ac-H3K27me3" state increase in CLL compared with normal B cells across our cohort and Blueprint initiative samples. **c** Sankey diagram showing that ~47% of the regions in a "H3K27ac-H3K27me3" state in CLL carried repressive chromatin modifications in B cells. **d** Fold-change gene expression between CLL and normal B cells in relation to genomic distance from regions that gain H3K27ac (orange; $n = 11,740$ genes) or H3K27me3 (blue; $n = 8867$ genes) in CLL. Mann-Whitney *U*-test. **e** Position weight matrices of the top three motifs over-represented in CLL "H3K27ac-H3K27me3" regions. Motif enrichment hypergeometric test *P*-value and the best reference motif match (JASPAR core database) are shown. **f** Expression levels ($\log_2[\text{TPM}]$) of *MYC* target genes (containing promoter *MYC* binding motif, as in analysis in **e**) compared with non-*MYC* target genes in "H3K27ac-H3K27me3" regions in CLL. Mann-Whitney *U*-test. **g** Single-cell ($n = 94$) gene expression Shannon's information entropy in relation to the population average gene expression in CLL (scCLL_21). Colored lines—local regression curves for genes in a 'H3K27ac-H3K27me3' (brown) or "Repressed Polycomb (PRC)" (gray) state. Inset: single-cell gene expression Shannon's information entropy for each of the two chromatin states comparing genes within a defined range of population average gene expression. Mann-Whitney *U*-test. **h** Gene expression magnitudes of genes in "Repressed Polycomb (PRC)" state and genes in "H3K27ac-H3K27me3" state. Cumulative distribution (right) showing the proportion of intermediate single-cell gene expression Shannon's information entropy values at these genes is also shown. Kolmogorov-Smirnov test. **i** Well-coordinated chromatin programs stabilize gene expression and cellular identities in normal B cells (left). On the contrary, intra-leukemic epigenetic diversity results in a permissive chromatin state in CLL cells (right), enhancing cell-to-cell transcriptional variation. Boxplots represent median and bottom and upper quartile; whiskers correspond to $1.5 \times \text{IQR}$; error bars represent 95% confidence interval

analysis of closest genes to these regions revealed enrichment in gene sets associated with stem cell identity^{29,30} (Hypergeometric test BH-FDR <0.05), linking regulatory chromatin variability to stem-like cell programs, according to the notion that epigenetic variability in cancer may lead to a drift toward a hybrid stem-somatic cell state^{9,31} (Supplementary Fig. 4e; Supplementary Data 10).

Epigenetic factors, such as aberrant regulation of H3K27 methylation¹² and sporadic TF activation³², have been recently implicated in promoting lineage plasticity in cancer. Thus, to identify which TFs may carry the potential to rewire CLL cells and promote lineage plasticity in CLL, we further mined the regions marked by H3K27ac-H3K27me3 for transcription factor motif enrichment and identified a significant enrichment of the proto-oncogene *MYC* motif, a TF associated with lineage plasticity and CLL transformation to aggressive large B cell lymphoma³³ (Hypergeometric test $P < 0.0001$; Fig. 4e). RNA gene expression of genes with a *MYC* binding motif at their promoters was increased compared with non-*MYC* target genes, in the regions marked by H3K27ac-H3K27me3 (median [IQR] of 9.44 [4.34] vs. 8.23 [5.17] \log_2 [TPM], respectively; Mann-Whitney U -test, $P < 0.0001$; Fig. 4f).

Notably, the observed co-mapping of H3K27ac and H3K27me3 to the same genomic locus in CLL may arise from cell-to-cell divergence in histone modification, rather than co-occurrence of these mutually exclusive marks in the same cells. Consistent with this hypothesis, we analyzed our single-cell whole-transcriptome data and observed that genes neighboring H3K27ac-H3K27me3 regions in CLL were associated with higher intra-leukemic expression information entropy in single cells compared with genes neighboring Polycomb repressed regions (Mann-Whitney U -test, $P < 0.0001$; Fig. 4g, h; Supplementary Fig. 4f, g). Collectively, these data suggest that CLL cell populations lose effective Polycomb repression of *MYC* targets, likely enabling an exploration of transcriptional stem-like cell programs in CLL evolution.

Discussion

While cancer evolution investigations have focused on genetic alterations, emerging data across cancer also highlighted the contribution of heritable epigenetic changes to cancer evolution^{11,12,32}. In this study, we provided an integrative analysis of the epigenetic landscape of CLL and its relationship to intra-leukemic epigenetic and transcriptional diversity.

We observed extensive chromatin rewiring at H3K27ac regulatory regions mediated by specific transcription factor families, in particular NFAT and TCF/LEF transcription factor families^{8,19,20}. Through targeted bisulfite sequencing capture assay, we further showed these regulatory regions to display the highest degree of change in DNAm. Notably, enhancer hypomethylation is preferentially associated with intermediate DNAm levels, likely reflecting intra-leukemic cell-to-cell heterogeneity^{9,10}. Thus, intermediately methylated regions in cancer may not be limited to heterochromatin as previously described^{25,26}, affecting also regions of regulatory chromatin.

Moreover, while normal B cells exhibit coordinated epigenetic-transcriptional regulation resulting in higher pairwise mutual information, CLL samples have a substantial decrease in DNAm-RNA mutual information. This finding is consistent with intra-leukemic heterogeneity decreasing the mutual information of these two variables when measured at the population level. To directly examine this scenario, we applied matched DNAm and mRNA single-cell information and found a greater increase in single-cell mutual information in CLL compared with normal B cells. This observation confirms that the relatively small

contribution of promoter DNAm to explaining transcriptional variation in bulk cancer studies⁹ results, at least in part, from intra-leukemic epigenetic diversity.

To further extend the evaluation of epigenetic co-ordination beyond two epigenetic layers, we modeled the combinatorial patterns of histone modifications, DNAm, and gene expression. Interestingly, we observed a dysregulation of the transcriptional output as a function of the combinatorial chromatin states. Specifically, while in normal B cells H3K27me3^{hi}/H3K4me3^{low}/H3K27ac^{low}-marked genes were generally associated with a uniform transcriptional output, in CLL these genes were associated with variable expression level. As H3K27me3 is typically deposited at gene promoters by Polycomb Repressive Complex 2 (PRC2) via its catalytic Ezh2/Ezh1 subunit³⁴, these results are consistent with CLL epigenetic landscape being marked by incomplete Polycomb complex-mediated gene silencing resulting in permissive chromatin states in a fraction of cells. Furthermore, as DNAm is important for appropriate retargeting of PRC2 and H3K27me3 histone modification across cell divisions³⁵, stochastic DNAm alterations during CLL evolution⁹ may lead to redistribution of the repressive activity of the PRC2 complex and the H3K27me3 mark, and cell-to-cell variation in the efficiency of PRC2 transcriptional silencing³⁶.

Lastly, we observed an unexpected co-occurrence of typically mutually exclusive activating (H3K27ac) and repressing (H3K27me3) histone modifications, closely associated with activation of stem-like programs and greater cell-to-cell transcriptional heterogeneity. Notably, the co-mapping of these typically mutually exclusive histone modifications was previously observed in the context of embryonic stem cell neural differentiation, reflecting cellular heterogeneity due to admixture of differentiated and undifferentiated cells³⁷. Thus, epigenetic diversification leads to corrupted coherence across the different layers of the epigenome in CLL, consistent with ongoing epigenetic diversification leading to an admixture of cells with diverging epigenetic identities (Fig. 4i).

While genetic heterogeneity plays a key role in cancer growth, progression, and evolution with therapy^{3,5,38,39}, epigenetic evolutionary routes are a major emerging theme across cancer, including prostate cancer, lung cancer and melanoma^{11,40}. Cancer cells can display profound non-genetically mediated transcriptional variability, which may enable adaptive changes such as therapeutic resistance, persistence or lineage plasticity. Notably, these states are efficiently propagated to progeny cells suggesting stable epigenetic encoding. Indeed, in CLL, non-genetic persistence as well as lineage transformation have been reported as potential routes of escape from therapeutic inhibition^{5,38,41}. Our data demonstrates that these adaptive capacities may be fueled by significant intra-tumoral epigenetic diversity resulting in permissive chromatin states across cells, leading to greater cell-to-cell transcriptional variation (Fig. 4i). Thus, intra-tumoral epigenetic diversity may permit leukemic cells to stochastically activate alternate gene regulatory programs, facilitating the emergence of novel cell states, ultimately fostering CLL's ability to efficiently explore the fitness landscape for superior evolutionary trajectories during tumorigenesis and in response to therapy.

Methods

Human subjects, sample collection, and genotyping. The study was approved by the local ethics committee and by the Institutional Review Board (IRB) and conducted in accordance to the Declaration of Helsinki protocol. Blood samples were collected in EDTA blood collection tubes (BD Biosciences) from patients and healthy adult volunteers enrolled on clinical research protocols at the Dana-Farber/Harvard Cancer Center (DF/HCC), Memorial Sloan Kettering Cancer Center (MSKCC), and New-York-Presbyterian/Weill Cornell Medical Center (NYP/WCMC). We note that the IRB does not permit collection of demographic information of healthy donors. Informed consent on DF/HCC, MSKCC and WCMC IRB-approved protocols for genomic sequencing of patient samples was obtained prior to the initiation of

sequencing studies. The diagnosis of CLL according to World Health Organization (WHO) criteria was confirmed in all cases by flow cytometry, or by lymph node or bone marrow biopsy. B cells from healthy donors and CLL patient samples were isolated from blood samples using Ficoll-Paque Plus (GE Healthcare) density gradient centrifugation and red blood cell lysis, followed by EasySep™ Human B Cell Enrichment Kit (STEMCELL Technologies, Vancouver, Canada) as per manufacturer recommendation. Cells were then cryopreserved in 50% FBS/40% RPMI/10% DMSO and stored in vapor-phase liquid nitrogen until the time of analysis. Tonsillar B cell populations were affinity-purified from de-identified human tonsillectomy specimens by magnetic cell separation⁴², cryopreserved in 50% FBS/40% RPMI/10% DMSO and stored in vapor-phase liquid nitrogen until the time of analysis. White blood cells (WBC) counts for the CLL patient samples used in our analyses were in a range of 50–394 K (median of 201 K), consistent with a purity of >90% based on previously published sequencing data³. Immunoglobulin heavy-chain variable (*IGHV*) homology (unmutated was defined as greater than or equal to 98% homology to the closest germline match) were determined⁴³. Cytogenetics were primarily evaluated by FISH analysis for the most common CLL abnormalities [del(13q), trisomy 12, del(11q), del(17p), del(6q), amp(2p)] (Supplementary Fig. 1b); if FISH was unavailable, genomic data were used.

Antibodies. Purified CD19⁺ naive B cells (CD19⁺CD23⁺CD27⁻IgD⁺) and germinal center memory B cells (CD19⁺CD23⁺CD27⁺IgD⁻) were sorted using PE/Cy7 anti-human CD27 (1:5 dilution; clone O323, Bio Legend) and FITC mouse anti-human IgD (1:5 dilution; clone IA6-2, BD Pharmingen) antibodies with a FACSAria II instrument (Becton Dickinson, Franklin Lakes, NJ). Tonsillar CD20⁺ cells were sorted as CD19⁺CD20⁺CD38⁺⁺. CD5⁺ normal B cells were not profiled due to their low frequency, and previous data⁷ showed minimal DNA methylation differences between CD5⁺ and CD5⁻ naive B cells. Antibodies used for ChIP include anti-H3K4me3 (1 mg for 50 mg of chromatin; 9751S Cell Signaling, Danvers, MA), anti-H3K27ac (2 mg for 25 mg of chromatin; ab4729 Abcam, Cambridge, United Kingdom), anti-H3K27me3 (2 mg for 25 mg of chromatin; 07–449 Millipore, Burlington, MA).

ChIP-seq and motif analysis. A minimum of 2 million purified human cells were used. Briefly, cells were fixed in a 1% methanol-free formaldehyde solution and then resuspended in sodium dodecyl sulfate (SDS) lysis buffer. Lysates were sonicated in an E220 focused-ultrasonicator (Covaris, Woburn, MA) to a desired fragment size distribution of 100–500 base pairs. ChIP assays were processed on a SX-8G IP-STAR Compact Automated System (Diagenode, Denville, NJ) using a direct ChIP protocol⁴⁴. Briefly, immunoprecipitation reactions were performed with the above-indicated antibodies, each on approximately 500,000 cells, and incubated overnight at 4 °C. The immune complex was collected with protein A/G agarose or magnetic beads and washed sequentially in the low salt wash buffer (20 mM Tris pH8, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA), the high salt wash buffer (20 mM Tris pH8, 500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA), the LiCl wash buffer (10 mM Tris pH8, 250 mM LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA) and TE. Chromatin was eluted with elution buffer (1% SDS, 0.1 M NaHCO₃), and then reverse cross-linked with 0.2 M NaCl at 65 °C for 4 h. DNA fragments were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). Barcoded immunoprecipitated DNA and input DNA were prepared using the NEBNext ChIP-seq Library Prep Master Mix Set for Illumina (#E6240, New England Biolabs, Ipswich, MA) and TruSeq Adapters (Illumina) according to the manufacturer's protocol on a SX-8G IP-STAR Compact Automated System (Diagenode). Phusion High-Fidelity DNA Polymerase (New England Biolabs) and TruSeq PCR Primers (Illumina, San Diego, CA) were used to amplify the libraries, which were then purified to remove adapter dimers using AMPure XP beads and multiplexed on the HiSeq 2000 (Illumina, San Diego, CA). Previously published CLL and normal B cells ChIP-seq datasets were downloaded from the Blueprint DCC portal (Blueprint; <http://dcc.blueprint-epigenome.eu/#/home/>).

ChIP-seq data were processed according to the ENCODE Histone ChIP-seq Data Standards and Processing Pipeline (<https://www.encodeproject.org/chip-seq/histone/>). Raw reads were mapped to the human genome GRCh37 assembly using Burrows-Wheeler Aligner⁴⁵ (BWA v0.7.17). Duplicate reads were removed using Picard (<https://broadinstitute.github.io/picard/>) and bigwig files were created for visualization. Peaks were identified with Macs2⁴⁶ (v2.0.10) with a *q*-value threshold of 0.01. De novo motif enrichment analyses were performed using Homer⁴⁷ against JASPAR CORE database (-cpg parameter was used for CpG% normalization). Peaks overlapping with Satellite repeat regions and Encode Blacklist were discarded. All ChIP-seq reads were normalized and displayed as read counts per million mapped reads. Super-enhancers in H3K27ac peaks were defined as in^{14,48}. First, for each sample, H3K27ac peaks without any overlap with known gene promoters (TSS ± 2.5 kb) were identified. Then, H3K27ac peaks within 12.5 kb of each other were concatenated and these regions were ranked by their total normalized H3K27ac signal. H3K27ac intensity was plotted against the corresponding concatenated enhancers rank. The cut-point between super-enhancers and enhancers was defined on the enrichment profile as the tangent with slope equal to 1. Enhancers on the right of the inflection point were defined as super-enhancers (see Supplementary Fig. 1c). We note that the number of super-enhancers identified in our CLL cohort (range [279–964]; median of 474 across

samples) is in line with a recent study that investigated enhancer architecture in a distinct CLL cohort⁴⁸. To identify variable super-enhancer domains enriched in either CLL or normal B cells, we defined the union of all super-enhancers discovered across the CLL and normal B cell cohorts. Differentially regulated super-enhancers in CLL compared with normal B cells (*n* = 297; see Fig. 1a–c and Supplementary Fig. 1d) were identified with DESeq2⁴⁹ as those with absolute log₂(H3K27ac fold-change) >2 and Benjamini-Hochberg adjusted *P*-value < 0.01. We note that few differences in the super-enhancer landscape were observed between the two major known CLL subtypes, even with less stringent fold-change and *P*-value thresholds (absolute log₂[H3K27ac fold-change] >1 and Benjamini-Hochberg adjusted *P*-value < 0.05; see Supplementary Fig. 1f, g), consistent with prior studies showing more subtle chromatin differences between CLL subtypes^{19,48}. For visualization purposes, count data matrices were transformed using variance stabilizing transformation (VST).

RNA-seq. RNA was extracted using Qiagen (Hilden, Germany) RNeasy columns according to the manufacturer's instructions. Subsequently, 500 ng of total RNA was used for polyA selection and TruSeq library preparation according to the instructions provided by Illumina (TruSeq RNA Sample Prep Kit v.2), with 8 cycles of PCR. Samples were barcoded and run on a HiSeq 4000 in a 125 bp paired-end mode, using the TruSeq SBS Kit v.3 (Illumina, San Diego, CA). An average of 75 million paired reads was generated per sample. Raw reads were mapped to the human genome GRCh37 using STAR (v2.5.2a) aligner⁵⁰. We used several QC metrics for the RNA-seq library, including intron-exon ratio, intragenic reads fraction, and GC bias. We quantified exon and gene expression using Salmon⁵¹ against the *Homo sapiens* transcriptome GRCh37.

Reduced representation bisulfite sequencing (RRBS). Genomic DNA from CLL samples and normal B cell samples were used to produce RRBS libraries. These were generated by digesting genomic DNA with MspI to enrich for CpG-rich fragments, and then were ligated to barcoded TruSeq adapters (Illumina, San Diego, CA) to allow immediate subsequent pooling. This was followed by bisulfite conversion and PCR. Libraries were sequenced and 29mers were aligned to the hg19 genome using MAQ version 0.6.6⁵². Reads were further filtered if: (i) the read did not align to an autosome, (ii) the read failed platform/vendor quality checks (SAMtools flag 0 × 200), and/or (iii) the read did not align to a MspI cut site. The methylation state of each CpG was determined by comparing bisulfite-treated reads aligning to that CpG with the genomic reference sequence. The methylation level was computed by dividing the number of observed methylated cytosines (which did not undergo bisulfite conversion) by the total number of reads aligned to that CpG. In addition, the number of CpG measurements on each read was noted.

Targeted bisulfite sequencing capture assay. Hybrid-selected sequencing libraries were prepared combining Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) with the NimbleGen SeqCap Epi Enrichment System (Roche NimbleGen), enabling lower input DNA quantities while maintaining library complexity. Briefly, pre-capture libraries were constructed following the “bisulfite-conversion first” library construction protocol of the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) with the following exceptions: (1) To minimize the off-target sequencing rate, we sheared the input DNA to ~200 bp fragments instead of shearing to ~400 bp fragments; (2) we doubled the PCR volume and used 8 PCR cycles for the pre-capture library amplification in 1× HiFi HotStart ReadyMix (Kapa Biosystems). SeqCap Epi hybridization reactions contained a total of 1 µg of a pool of 2–4 PCR-amplified pre-capture libraries, a total of 1 nmol of 2–4 index-specific blocking oligonucleotides, and the custom SeqCap probe pool designed for the targets listed in Supplementary Data 4. Hybrid-selected sequencing libraries were sequenced on an Illumina HiSeq 2500 instrument in fast mode together with a 10% spike-in of a non-indexed PhiX174 library to generate a median of ~48 million indexed 100-base purity-filtered paired reads per sample. Raw reads were aligned to the human genome (hg19) using bsMap v2.95⁵³ with the following parameters: `bsmap -s 16 -v 0.1 -S 1 -n 1 -q 20 -r 0`. Subsequently, we used Picard tools (<http://picard.sourceforge.net>) version 2.16.0 to further process and QC the aligned data files. Standard performance metrics for each library are available in Supplementary Data 5. We used MarkDuplicates with standard parameter settings to mark and remove likely PCR duplicates, CollectAlignmentMetrics to compute basic alignment statistics, and CalculateHsMetrics with Supplementary Data 4 to calculate all hybrid capture-related metrics, including the on-target rate (Supplementary Fig. 2a). To determine the methylation state of all CpGs captured and assess the bisulfite conversion rate, we used the mcall module in the MOABS⁵⁴ software suite with standard parameter settings. Then, we converted the resulting CpG level files to bigBed files for visualization in IGV⁵⁵, filtering out all CpGs that were covered with less than five reads. Analysis of targeted bisulfite sequencing capture assay data was conducted using the methylKit package⁵⁶ and a 500-bp tiling of the target capture set. Briefly, we imported the CpG level methylation call files from mcall into R using the methylKit function “methRead” and then computed the weighted methylation mean for each 500-bp tile using the function “getData”, weighting the methylation level of each CpG with its coverage. We then merged the tile level methylation information across all samples and retained only those tiles covered with more than 10 reads in 70% or more of all samples. To compute differentially

methylated tiles, we performed Fisher's exact test on pooled CLL vs. normal B samples for each tile. Subsequently, we corrected the resulting *P*-values using Benjamini–Hochberg correction and defined regions with a *Q*-value ≤ 0.05 and an absolute methylation difference ≥ 0.3 as differentially methylated. Finally, we merged differentially methylated tiles into larger differentially methylated regions (DMRs) if they were less than 400 bp apart²².

BEDTools v2.25.0⁵⁷ was used to calculate overlaps between differentially methylated regions with the different genomic features investigated, requiring a 50% minimal overlap fraction. Promoters were defined as 1 kb upstream and 1 kb downstream of hg19 RefGene gene transcription start sites (TSSs), unless stated otherwise. The set of CpG Islands (CGIs) were defined using biologically-verified CGIs⁵⁸. ChIP-seq peak sets were defined as above-described. For the pathway analysis in Supplementary Fig. 2g, we used GREAT version 3.0.0⁵⁹ to identify associated biological themes, using default association rule (i.e., basal plus extension: 5000 bp upstream, 1000 bp downstream; Hypergeometric test BH-FDR < 0.05).

Whole-exome DNA sequencing (WES). Genomic DNA from two *IGHV* mutated and two *IGHV* unmutated CLL patient samples were used to produce whole-exome libraries. Details of whole-exome library construction and analysis have been detailed elsewhere³. Briefly, output from Illumina software (Illumina, San Diego, CA) was processed by the Picard data processing pipeline to yield BAM files containing aligned reads with well-calibrated quality scores. We used the ABSOLUTE algorithm⁶⁰ to calculate tumor purity—the ratio of tumor cells to total cells in the sample—and obtained a very high degree of purity (median of 0.98; range [0.9–1]), consistent with a negligible contamination of non-malignant cells in our CLL samples.

Multiplexed single-cell RRBS (MscRRBS) library construction. Single cell experiments were performed by sorting DAPI negative cells in 96-well plates in 3 μ L of 0.1 \times CutSmart buffer (New England Biolabs) per well using a BD Influx sorter (Becton Dickinson, Franklin Lakes, NJ). Nucleated CLL cells were gated and index-sorted as CD19⁺CD5⁺ cells, which in CLL patients are overwhelmingly malignant⁶¹ ($\geq 95\%$). Plates were stored at -80°C until further processing. The day of the experiment, cells were lysed for 2 h at 50°C in 1 \times CutSmart buffer supplemented with Proteinase K (0.2U, NEB) and Triton X-100 (0.3%, Sigma Aldrich) for a final volume of 5 μ L. Proteinase K was heat-inactivated for 30 min at 75°C . DNA was incubated with 10 units of the restriction enzyme Msp1 (Fermentas) in 6.5 μ L final volume reaction during 90 min at 37°C . Heat-inactivation was performed for 10 min at 70°C . Digested DNA was filled-in and A-tailed at the 3' sticky ends in 8.5 μ L final volume of 1 \times CutSmart with 2.5 units of Klenow fragment (Exo-, Fermentas). Reaction was supplemented with 1 mM dATP and 0.1 mM dCTP and 0.1 mM dGTP (NEB) and performed as follows in a thermocycler: 30°C for 25 min, 37°C for 25 min and heat-inactivation at 70°C for 10 min. Custom barcoded methylated adapters (0.1 μ M) were then ligated overnight at 16°C with the dA-tailed DNA fragments in the presence of 800 units of T4 DNA ligase (NEB) and 1 mM ATP (Roche) in a final volume of 11.5 μ L of 1 \times CutSmart buffer. T4 DNA ligase heat-inactivation was performed at 70°C for 15 min the next day. Genomic DNA from 24 individual cells were pooled together according to their barcodes, giving, for a 96-well plate, 4 pools of 24 cells. Pooled genomic DNA was cleaned-up and concentrated using 1.8 \times SPRI beads (Agencourt AMPure XP—Beckman Coulter). Each pool was then sodium bisulfite converted (Fast Epitect Bisulfite, Qiagen) following manufacture recommendations. To ensure full bisulfite conversion, two cycles of conversion were performed. The double-stranded DNA was first denatured 10 min at 98°C and then incubated for 20 min at 60°C . Hundred nanogram of dephosphorylated and sheared bacterial DNA was added as carrier to every pool prior to conversion. Converted DNA was then amplified using primers containing Illumina i7 and i5 index. Following Illumina pooling guidelines, a different i7 index was used for every 24-cell pool, allowing multiplexing of 96 cells for sequencing on one Illumina HiSeq lane. Library enrichment was done using KAPA HiFi Uracil + master mix (Kapa Biosystems) and the following PCR condition was used: 98°C for 45 s; 6 cycles of: 98°C for 20 s, 58°C for 30 s, 72°C for 1 min; followed by 12 cycles of: 98°C for 20 s, 65°C for 30 s, 72°C for 1 min. PCR was terminated by an incubation at 72°C for 5 min. Enriched libraries were cleaned-up and concentrated using 1.3X SPRI beads. DNA fragments between 200 bp and 1 kb were size-selected and recovered after resolving on a 3% NuSieve 3:1 agarose gel. Libraries molarity concentration calculation was obtained by measuring concentration of double stranded DNA (Qubit) and quantifying the average library size (bp) using an Agilent Bioanalyzer. Every 24-cells pool was mixed with the others pool in an equimolar ratio. All cells from a 96-well plate were sequenced as paired-end on HiSeq 2500 with 10% PhiX spike-in. Negative controls (empty wells with no cell) were used to control for non-specific amplification of the libraries.

MscRRBS read alignment. Each pool of 96 cells was first demultiplexed by Illumina i7 barcodes (Supplementary Data 11), resulting in four pools of 24 cells. Each pool of 24 cells was further demultiplexed by unique cell barcodes (Supplementary Data 12). Reads were assigned to a given cell if they matched 80% of the template adapters. Adapters and adapter reverse complements (6 bp) were trimmed from the raw sequence reads. After adapter removal, reads were trimmed from their 3' end for read quality by applying a 4 bp sliding window and removing bases until the mean base quality of the window had a Phred quality score greater than 15. Read

pairs with a read shorter than 36 bp after trimming were discarded. We aligned trimmed reads in single-end mode to the hg19 human genome assembly using Bismark⁶² (v.0.14.5; parameters: -multicore 4 -X 1000 --un --ambiguous) running on bowtie2-2.2.8 aligner⁶³. Bismark methylation extractor (--bedgraph --comprehensive) was used to determine the methylation state of each individual CpG. For downstream analyses, a site was considered methylated or unmethylated only if there was 90% agreement of the methylation state for all reads mapped to the site.

Joint MscRRBS and single-cell RNA-seq library construction. Single cells were sorted by flow cytometry into 2.5 μ L of RLT Plus buffer (Qiagen) supplemented with 1 U/ μ L of RNase Inhibitor (Lucigen). Sorted cells were immediately stored at -80°C . Genomic DNA (gDNA) and mRNA have been separated manually. A modified oligo-dT primer (5'-biotin-triethyleneglycol-AAGCAGTGGTATCAACGCA-GAGTACT30VN-3', where V is either A, C or G, and N is any base; IDT) was conjugated to streptavidin-coupled magnetic beads (Dynabeads, Life Technologies) according to the manufacturer's instructions. To capture polyadenylated mRNA, we added the conjugated beads (10 μ L) directly to the cell lysate and incubated them for 20 min at room temperature with mixing to prevent the beads from settling. The mRNA was then collected to the side of the well using a magnet, and the supernatant, containing the gDNA, was transferred to a fresh plate. Single-cell complementary DNA was amplified from the tubes containing the captured mRNA according to the Smart-seq2 protocol⁶⁴. After amplification and purification using 0.8 \times SPRI beads, 0.5 ng cDNA was used for Nextera Tagmentation and library construction. Library quality and quantity was respectively assessed using Agilent Bioanalyzer 2100 and Qubit, respectively. Genomic DNA present in the pooled supernatant and wash buffer from the mRNA isolation step was precipitated on 0.8 \times SPRI beads and eluted directly into the reaction mixtures for Msp1 (Fermentas) enzymatic reaction (10 μ L final reaction). MscRRBS protocol was then performed on the digested gDNA after the restriction enzyme digestion step.

Single-cell RNA-seq gene expression quantification. The sequenced read fragments were mapped against the hg19 human genome assembly using the 2pass default mode of STAR⁵⁰ (version 2.5.2a) with the annotation of GENCODE⁶⁵ (version 19). The number of read counts overlapping with annotated genes were quantified applying the "GeneCounts" option in the STAR alignment. We filtered out poor quality cells when the detected number of genes was below 500 or the fraction of mitochondrial gene counts was higher than 20%. To compare the cells in terms of their transcriptional differences, we normalized the read counts by scaling for the total number of counts per cell. To assess potential confounding effect due to cell cycle phase, we classified CLL cells into cell cycle phases using AUCcell method implemented in the SCENIC analytical toolkit⁶⁶. Briefly, AUCcell uses the area under the curve (AUC) to identify cells with active gene sets, by calculating the proportion of genes in each input gene set that is enriched within the expressed genes for each cell. Each cell is assigned an AUC score for each gene set. Cells expressing many genes from the gene set will have higher AUC score than cells expressing fewer. Last, the highest AUC threshold is used to consider a gene set "active" in a given cell. The Molecular Signature Database⁶⁷ (MSigDB; <http://www.broad.mit.edu/gsea/>) C2 curated gene sets "BIOCARTA_G2_PATHWAY" and "BIOCARTA_G1_PATHWAY" were used as input gene sets for this analysis. We observed that the vast majority of cells are classified as being non-cycling cells ($n = 240$; 99.6%), with a negligible number of cells being in either G2/M ($n = 1$; 0.4%; AUC > 0.27) or G1 phase ($n = 0$; 0%; AUC > 0.18), consistent with the majority of CLL cells being in a resting non-cycling state⁶⁸.

Single-cell DNAm-gene expression MI analysis. To begin, cells with fewer than 500 detected genes or a proportion of mitochondrial or ribosomal reads above 20% were removed from the analysis for quality control. Constitutively highly expressed mitochondrial genes and genes encoding ribosomal proteins across all cells were then removed. Then, cells in the bottom 10th percentile of total read counts for a given sample were discarded, and each of the remaining cells was probabilistically down-sampled to match the number of reads at this cutoff. Subsequently, genes with reads detected in less than five cells were removed from the analysis. At single-cell resolution, a gene's promoter methylation rate was represented by the proportion of methylated CpGs in the region 2500 base pairs upstream and downstream of the transcription start site. Genes with less than 10 CpG observations in the promoter region for a given cell were removed. We then computed the mutual information between promoter methylation rate and gene expression for each cell using a threshold of zero, implying that any detected methylation or expression for a given gene was treated as having a value of 1 for that cell, and 0 otherwise. For a gene to be included in the final analysis, it was required to have at least 10 cells with sufficient CpG data for a methylation call (10 CpG observations) as well as greater than 10% non-zero expression across all cells to mitigate the impact of dropout. The approach was validated by a non-parametric permutation test, in which we randomly permuted the cell methylation values for each gene while holding the corresponding expression vector constant (such that RNA and DNAm are no longer linked at the single-cell level) and computed an unmatched version of the mutual information. This was repeated as many times as cells were available for a given gene, and the final unmatched mutual information value provided corresponds to the median of the result for each of these permutations. We note that the analysis for Fig. 3c was

performed with downsampling to create a balanced dataset by matching the number of genes between CLL and normal B cells ($n = 759$ genes).

Gene set enrichment analysis. Gene set enrichment analysis was performed using GSEA software, and Molecular Signature Database⁶⁷ (MSigDB; <http://www.broad.mit.edu/gsea/>). Specifically, we used the C2 curated gene sets and Benjamini-Hochberg FDR adjusted P -value cut-off of 0.05.

Chromatin hidden Markov model (HMM). Chromatin states across the genome were defined using EpicSeg⁶⁹, which is based on a multivariate HMM, using H3K4me3, H3K27ac, H3K27me3, whole cell extract, RNA-seq and DNase (based on bulk RRBS) datasets as input. ChIP-seq reads were shifted in the 5′–3′ direction by 100 bp. Reads counts were computed in 200 bp non-overlapping bins. Normalized raw counts were then modeled with an HMM assuming that the hidden state vector followed a negative binomial distribution. We trained several HMM models in parallel mode with the number of states ranging from 5 states to 25 states and chose a 12-state model as the best model that captures all the key interactions between the epigenetic marks and cover all possible genomic elements (promoter, enhancer, gene body) that we expected to resolve given the selection of datasets we used (H3K4me3, H3K27ac, H3K27me3, RNA-seq, and DNase). Genomic regions were then annotated with the state with the maximum posterior probability in the 200 bp bin. State enrichment in different genomic features was calculated dividing the percentage of nucleotides occupied by a state in a particular genomic feature by the percentage of nucleotides that this genomic feature represents in the entire genome.

Chromatin Dirichlet process Gaussian mixture model. Infinite mixture model with the Dirichlet process was used to model the normalized signal count matrix and to derive a segmentation of the chromatin tracks. The scikit-learn Python library (sklearn.mixture.BayesianGaussianMixture v0.19) was used to generate an independent model for each sample. Cross-validation for each sample was performed training on a random 1/10 of the genome, applying the cross-validation model to the sample and repeating this procedure 100 times. Subsequently, a leave-one-out procedure was implemented to assess the contribution of each chromatin and transcriptome track independently. Unsupervised hierarchical clustering of state emission was performed to identify unique states.

Single-cell entropy analysis. To test for significance of association of chromatin state status with expression heterogeneity in Figs. 3f and 4g, single cell RNA-seq read counts observed in each cell were normalized by the effective library size and transcript length, and the fraction of positive cells (fpc) was calculated per gene (a cell is defined as positive if > 0 reads aligned to the gene). Subsequently, Shannon's information entropy (ent) was calculated for each gene as followed:

$$\text{ent} = [-1 \times (\text{fpc} \times \log_2(\text{fpc}) + (1 - \text{fpc}) \times \log_2(1 - \text{fpc}))] \quad (1)$$

The association with chromatin state status was tested using a generalized additive model (implemented by gam R package). The following type of model was tested:

$$\text{ent} \sim s(\text{population average expression}) + \text{chromatin state status}$$

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

Statistical methods. Statistical analysis was performed with Python 2.7.13 and R version 3.4.2. Categorical variables were compared using the Fisher's Exact test. Continuous variables were compared using the Mann-Whitney U -test, Welch's t -test, paired t -test, non-parametric permutation test or Kolmogorov-Smirnov test as appropriate. P -values were adjusted for multiple comparisons by Benjamini-Hochberg FDR procedure, as appropriate. All P -values are two-sided and considered significant at the 0.05 level unless otherwise noted.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

ChIP-seq, RNA-seq, and DNase datasets have been deposited to the NCBI Gene Expression Omnibus⁷⁰ (GEO) under accession number GSE119103. MscRRBS and single-cell Smart-seq2 datasets have been deposited to the NCBI GEO under accession number GSE109085. The dbGaP accession number for the whole-exome sequencing data reported in this paper is phs000435.v2.p1. H3K27me3 ChIP-seq data for primary human tonsillar naive B cells and tonsillar germinal center B cells were downloaded from NCBI GEO under accession number GSE45982⁵⁰. Previously published CLL and normal B cell ChIP-seq and RNA-seq datasets were downloaded from the Blueprint DCC portal under accession number EGAC00001000135.

Received: 4 October 2018 Accepted: 22 March 2019

Published online: 23 April 2019

References

- Puente, X. S. et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
- Schuh, A. et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120**, 4191–4196 (2012).
- Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
- Fabbri, G. et al. Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. *J. Exp. Med.* **210**, 2273–2288 (2013).
- Burger, J. A. et al. Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to btk inhibition. *Nat. Commun.* **7**, 11589 (2016).
- Oakes, C. C. et al. Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov.* **4**, 348–361 (2014).
- Kulis, M. et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 1236–1242 (2012).
- Oakes, C. C. et al. DNA methylation dynamics during b cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.* **48**, 253–264 (2016).
- Landau, D. A. et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* **26**, 813–825 (2014).
- Landan, G. et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**, 1207–1214 (2012).
- Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
- Liau, B. B. et al. Adaptive chromatin remodeling drives glioblastoma stem cell plasticity and drug tolerance. *Cell Stem Cell* **20**, 233–246 (2017).
- Damle, R. N. et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94**, 1840–1847 (1999).
- Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2014).
- Wang, L. et al. Somatic mutation as a mechanism of wnt/ β -catenin pathway activation in CLL. *Blood* **124**, 1089–1098 (2014).
- Robertson, L. E. et al. Bcl-2 expression in chronic lymphocytic leukemia and its correlation with the induction of apoptosis and clinical outcome. *Leukemia* **10**, 456–459 (1996).
- Kosmaczewska, A. et al. CtlA-4 overexpression in cd19+/cd5+ cells correlates with the level of cell cycle regulators and disease progression in b-CLL patients. *Leukemia* **19**, 301–304 (2004).
- Adams, D. et al. Blueprint to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
- Beekman, R. et al. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* **24**, 868–880 (2018).
- Gutierrez, A. J. et al. Lef-1 is a pro-survival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. *Blood* **116**, 2975–2983 (2010).
- Qu, Y. et al. Cancer specific changes in DNA methylation reveal aberrant silencing and activation of enhancers in leukemia. *Blood* **129**, 13–25 (2016).
- Ziller, M. J. et al. Targeted bisulfite sequencing of the dynamic DNA methylome. *Epigenet. Chromatin* **9**, 55 (2016).
- Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome — biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).
- Kulis, M. et al. Whole-genome fingerprint of the DNA methylome during human b cell differentiation. *Nat. Genet.* **47**, 746–756 (2015).
- Berman, B. P. et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44**, 40–46 (2012).
- Timp, W. et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med.* **6**, 61 (2014).
- Vargova, K. et al. MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. *Blood* **117**, 3816–3825 (2011).
- Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Lim, E. et al. Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* **12**, R21 (2010).
- Wong, D. J. et al. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333–344 (2008).
- Timp, W. & Feinberg, A. P. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat. Rev. Cancer* **13**, 497–510 (2013).
- Mu, P. et al. Sox2 promotes lineage plasticity and antiandrogen resistance in tp53- and rb1-deficient prostate cancer. *Science* **355**, 84–88 (2017).

33. Delgado, M. D. & León, J. Myc roles in hematopoiesis and leukemia. *Genes Cancer* **1**, 605–616 (2010).
34. Hoffmann, A. et al. Switch-like roles for Polycomb proteins from neurodevelopment to neurodegeneration. *Epigenomes* **1**, 21 (2017).
35. Reddington, J. P. et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).
36. Gal-Yam, E. N. et al. Frequent switching of Polycomb repressive marks and DNA hypomethylation in the PC3 prostate cancer cell line. *PNAS USA* **105**, 12979–12984 (2008).
37. Rehimi, R. et al. Epigenomics-based identification of major cell identity regulators within heterogeneous cell populations. *Cell Rep.* **17**, 3062–3076 (2016).
38. Landau, D. A. et al. The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. *Nat. Commun.* **8**, 2185 (2017).
39. Landau, D. A. et al. Mutations driving clonal evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
40. Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
41. Byrd, J. C. et al. Ibrutinib in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.* **369**, 1278–1279 (2013).
42. Béguelin, W. et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* **23**, 677–692 (2013).
43. Rassenti, L. Z. et al. Relative value of zap-70, cd38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood* **112**, 1923–1930 (2008).
44. O’Geen, H. et al. in *Epigenetics Protocols* (ed. Tollefsbol, T. O.) 265–286 (Humana Press, Totowa, NJ, 2011).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
46. Zhang, Y. et al. Model-based analysis of chip-seq (macs). *Genome Biol.* **9**, R137 (2008).
47. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol. Cell* **38**, 576–589 (2010).
48. Ott, C. et al. Enhancer architecture and essential core regulatory circuitry of chronic lymphocytic leukemia. *Cancer Cell* **6**, 982–995.e7 (2018).
49. Love, M. I. et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
50. Dobin, A. et al. Star: ultrafast universal rna-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
51. Patro, R. et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
52. Li, H. et al. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
53. Xi, Y. & Li, W. Bismap: whole genome bisulfite sequence mapping program. *BMC Bioinform.* **10**, 232 (2009).
54. Sun, D. et al. Moabs: model based analysis of bisulfite sequencing data. *Genome Biol.* **15**, R38 (2014).
55. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
56. Akalin, A. et al. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
57. Quinlan, A. R. & Hall, I. M. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
58. Illingworth, R. S. et al. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* **6**, e1001134 (2010).
59. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
60. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
61. Hayes, G. M. et al. Isolation of malignant B cells from patients with chronic lymphocytic leukemia (CLL) for analysis of cell proliferation: validation of a simplified method suitable for multi-center clinical studies. *Leuk. Res.* **34**, 809–815 (2010).
62. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
63. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
64. Picelli, S. et al. Full-length rna-seq from single cells using smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
65. Harrow, J. et al. Gencode: the reference human genome annotation for the encode project. *Genome Res.* **22**, 1760–1774 (2012).
66. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
67. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci.* **102**, 15545–15550 (2005).
68. Herndon, T. M. et al. Direct in vivo evidence for increased proliferation of CLL cells in lymph nodes compared to bone marrow and peripheral blood. *Leukemia* **31**, 1340–1347 (2017).
69. Mammana, A. & Chung, H.-R. Chromatin segmentation based on a probabilistic model for read counts explains a large portion of the epigenome. *Genome Biol.* **16**, 151 (2015).
70. Edgar, R. et al. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

Acknowledgements

We thank members of Landau and Abdel-Wahab labs for constructive discussion, and the Integrated Genomics Operation (IGO) at Memorial Sloan Kettering and Epigenomics Core Facility at Weill Cornell Medicine for technical help. A.P. is supported by Mildred-Scheel Postdoctoral Research Fellowship of the Deutsche Krebshilfe e.V.. S.X.L. is supported by an AACR fellowship. R.C. is supported by LRF and Marie Skłodowska-Curie fellowships. A.G. is supported by Broad Institute SPARC funding. C.J.W. and O.A.-W. are Scholars of the Leukemia and Lymphoma Society (LLS). D.A.L. is supported by the Burroughs Wellcome Fund Career Award for Medical Scientists, ASH Scholar Award, Pershing Square Sohn Prize for Young Investigators in Cancer Research, and NIH Director’s New Innovator Award (DP2-CA239065). This work was also supported by the Starr Foundation, the Max Planck Society, LLS Translational Research Program, National Cancer Institute (R01-CA229902), and Stand Up To Cancer Innovative Research Grant (SU2C-AACR-IRG-0616).

Author contributions

A.P., F.G., O.A.-W. and D.A.L. conceived and designed the project. S.X.L., R.C., H.G., E.K.S., W.B., Y.J., A.A., J.N.A., R.R.F., and A.G. performed patient selection and prepared samples for sequencing. R.C., H.G., A.G., and A.M. designed the joint single-cell RRBS and RNAseq protocol. A.P., F.G., R.M.B., S.K., K.Y.H., R.S., and K.T.K performed the computational genomics analyses. F.G., A.P., A.G., C.J.W., A.M.M., A.M., B.E.B., O.A.-W. and D.A.L. wrote the manuscript with comments and contributions from all authors.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-09645-5>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019