12-1-2020

Evolution of the Epigenetic Landscape in Childhood B Acute Lymphoblastic Leukemia and Its Role in Drug Resistance

Shella Saint Fleur-Lominy
Perlmutter Cancer Center, NYU Langone Health, New York, New YorkDepartment of Medicine, NYU Langone Health, New York, New York

Nikki A. Evensen
Perlmutter Cancer Center, NYU Langone Health, New York, New York

Teena Bhatla
Department of Pediatrics, Children's Hospital of New Jersey at NBI, RWJBarnabas Health, Newark, New Jersey

Gunjan Sethia
Perlmutter Cancer Center, NYU Langone Health, New York, New York

Sonali Narang
Perlmutter Cancer Center, NYU Langone Health, New York, New York

See next page for additional authors

Recommended Citation

Saint Fleur-Lominy S, Evensen NA, Bhatla T, Sethia G, Narang S, Choi JH, Ma X, Yang JJ, Kelly S, Raetz E, Harvey RC, Willman C, Loh ML, Hunger SP, Brown PA, Getz KM, Meydan C, Mason CE, Tsirigos A, Carroll WL. Evolution of the Epigenetic Landscape in Childhood B Acute Lymphoblastic Leukemia and Its Role in Drug Resistance. Cancer Res. 2020 Dec 1;80(23):5189-5202. doi: 10.1158/0008-5472.CAN-20-1145. Epub 2020 Oct 16. PMID: 33067268; PMCID: PMC8647946.

HHS Public Access

Author manuscript

Cancer Res. Author manuscript; available in PMC 2021 December 06.

Published in final edited form as:

Cancer Res. 2020 December 01; 80(23): 5189–5202. doi:10.1158/0008-5472.CAN-20-1145.

Evolution of the epigenetic landscape in childhood B acute lymphoblastic leukemia and its role in drug resistance

Shella Saint Fleur-Lominy[1,2,+], **Nikki A. Evensen**[1,+], **Teena Bhatla**[3,+], **Gunjan Sethia**[1], **Sonali Narang**[1], **Jun H. Choi**[2], **Xiaotu Ma**[4], **Jun J. Yang**[4], **Stephen Kelly**[1], **Elizabeth Raetz**[1,5], **Richard C. Harvey**[6], **Cheryl Willman**[6], **Mignon L. Loh**[7], **Stephen P. Hunger**[8], **Patrick A. Brown**[9], **Kylie M. Getz**[10], **Cem Meydan**[10], **Christopher E. Mason**[10], **Aristotelis Tsirigos**[1,11,*], **William L. Carroll**[1,5,11,*,#]

[1]Perlmutter Cancer Center, NYU Langone Health New York, NY

[2]Department of Medicine, NYU Langone Health, New York NY

[3]Department of Pediatrics, Children's Hospital of New Jersey at NBI, RWJBarnabas Health, Newark, NJ

[4]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN

[5]Department of Pediatrics, NYU Health, New York NY

[6]University of New Mexico Comprehensive Cancer Center, Department of Pathology, University of New Mexico School of Medicine and Health Sciences Center, Albuquerque, NM

[7]Department of Pediatrics, UCSF Benioff Children's Hospital, San Francisco, CA

[8]Department of Pediatrics and the Center for Childhood Cancer Research, Children's Hospital of Philadelphia and The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

[9]The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Johns Hopkins University School of Medicine, Baltimore, Maryland.

[10]Department of Physiology and Biophysics and Institute for Computational Biomedicine and Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY 10065, USA.

[11]Department of Pathology, NYU Langone Health, New York, NY

Abstract

Although B cell acute lymphoblastic leukemia (ALL) is the most common malignancy in children and while highly curable, it remains a leading cause of cancer-related mortality. The outgrowth of tumor subclones carrying mutations in genes responsible for resistance to therapy has led to a Darwinian model of clonal selection. Previous work has indicated that alterations in the

[#] Contact: William L. Carroll MD, Perlmutter Cancer Center, Smilow 1211, 522 First Avenue New York NY 10016, Phone: 212-263-9247 Fax: 212-263-9190, william.carroll@nyumc.org. [*]Co-Corresponding Authors.
[+]These authors contributed equally to this work.

The authors declare no potential conflicts of interest.

epigenome might contribute to clonal selection yet the extent to which the chromatin state is altered under the selective pressures of therapy is unknown. To address this, we performed chromatin immunoprecipitation, gene expression analysis, and enhanced reduced representation bisulfite sequencing on a cohort of paired diagnosis and relapse samples from individual patients who all but one relapsed within 36 months of initial diagnosis. The chromatin state at diagnosis varied widely among patients: while the majority of peaks remained stable between diagnosis and relapse, yet a significant fraction were either lost or newly gained with some patients showing few differences and others showing massive changes of the epigenetic state. Evolution of the epigenome was associated with pathways previously linked to therapy resistance as well as novel candidate pathways through alterations in pyrimidine biosynthesis and downregulation of polycomb repressive complex 2 targets. Three novel, relapse-specific super-enhancers were shared by a majority of patients including one associated with S100A8, the top upregulated gene seen at relapse in childhood B-ALL. Overall, our results support a role of the epigenome in clonal evolution and uncover new candidate pathways associated with relapse.

Statement of Significane: This study suggests a major role for epigenetic mechanisms in driving clonal evolution in B ALL and identifies novel pathways associated with drug resistance.

Keywords

Epigenetics; leukemia; drug resistance; super enhancers

INTRODUCTION

The prognosis for children with acute lymphoblastic leukemia (ALL) has improved significantly (1,2). However, relapsed ALL remains a leading cause of cancer mortality in children and there is an urgent need to identify the biological basis of drug resistance (3-5). We and others have studied the clonal evolution of drug resistance leading to the discovery of relapse-specific or enriched somatic genetic alterations that result in resistance to one or more agents used in treatment (6-10).

Interestingly, two thirds of relapse-enriched mutations occur in genes encoding epigenetic regulators (11-13). We previously reported that leukemic blasts display a unique gene expression signature at relapse (8) and found that the histone deacetylase inhibitor (HDACi) vorinostat “reversed” the signature and restored chemosensitivity (10). Lastly, the relapsed genome is hypermethylated compared to diagnosis (8) and combination therapy with DNA methyltransferase and HDAC inhibitors works collaboratively with conventional chemotherapy to overcome drug resistance (10).

These studies implicate the epigenome in clonal evolution and treatment failure. Investigators have examined the regulatory epigenetic networks in hematological malignancies, (14-17) but mapping of genome-wide epigenetic state of primary leukemic blasts at diagnosis and relapse of B-ALL is lacking.

To understand the impact of epigenetic variation on clonal evolution we embarked on an unbiased genome-wide approach to map the location of key DNA methylation and histone marks in diagnosis/relapse pairs from bone marrow of children with B-ALL. Our

studies reveal great diversity in the chromatin state between individual patients and over the course of therapy. However, we observed a coalescence on particular epigenetic pathways associated with known mechanisms of drug resistance as well as previously unrecognized pathways that may drive relapse. These observations underscore the importance of epigenetic changes in mediating clonal evolution of relapsed disease.

MATERIALS AND METHODS

Samples and Sequencing Methods:

Cryopreserved paired diagnosis/relapse primary patient bone marrow or blood samples from individual patients who initially relapsed within 36 months of diagnosis (except for 1 who relapsed at 38 months) were obtained from the Children's Oncology group (COG) ALL biorepository. All patients (and/or parents) had written informed consent for the use of their tissue in research. Two cohorts were processed in chronological order. Cohort A patients (14 pairs) were treated on COG protocols AALL0331 and AALL0232 for NCI standard and high risk patients, respectively. Cohort B patients (20 pairs) were treated on COG protocols AALL0932 and AALL1131 for NCI standard and high risk patients, respectively. Details of all the samples are summarized in Supplementary Table S1. Cohort A samples were processed using conventional ChIPseq using antibodies targeting histone marks associated with transcriptional activation (H3K4me3, H3K9ac), enhancers (H3K27ac), and repression (H3K9me3, H3K27me3). Gene expression microarrays were already available in the TARGET database. Cohort B samples were processed using ChIPmentation, which uses significantly fewer cells (18), for 3 histone marks (H3K27ac, H3K27me3 and H3K9ac) that showed significant shifts across samples in preliminary data from cohort A. RNAseq, DNA enhanced reduced representation bisulfite sequencing (eRRBS) and whole genome sequencing (WGS) were performed according to published methods as discussed in Supplementary Information. Sequencing data are deposited in Gene Expression Omnibus (GSE156563).

Analytical Methods:

Differential histone modification and methylation analysis: Sequencing reads were analyzed according to standard approaches as described in Supplementary Information. Genome-wide annotations of peaks were performed using ChIPseeker R package (R version 3.4) (19). Epigenetic changes between paired diagnosis/relapse samples were analyzed using the R Bioconductor package DiffBind (version 3.6) (20) with a blocking factor. To avoid picking up insignificant changes due to background noise, we enriched for the most significant peaks by selecting those in the top 10^{th} percentile in normalized read density across patients. To determine statistical significance, the distribution of the fold-changes was compared against the distribution of fold-changes under permutation (permuting the diagnosis and relapse label of each paired samples). Next, we defined a gain of peak at the time of relapse as fold change of peak intensity >1.5 (log2FC>0.58) and a loss of peak at the time of relapse as fold change <0.66 (log2FC<-0.58), and these changes are referred as "dynamic" changes. Furthermore, we delineated gains and losses as "exclusive" changes in any given genomic region if no peak was called in one condition (diagnosis or relapse) while

a peak of great intensity (absolute log2FC>3) was noted in the other condition. Epigenetic changes at a false discovery rate (FDR) level of 20% are reported as significant.

Differential methylation was assessed using methylKit (21) leading to the identification of empirically based differentially methylated regions (22), which were CpGs with >25% differential methylation (up or down) and SLIM (Sliding Linear Model) adjusted p-values <0.01 (q-value). Q-value is the Sliding Linear Model-adjusted p-value from methylKit. Refseq was used to annotate genic regions and UCSC genome browser (http://genome.ucsc.edu) was used to annotate CpG islands.

Putative enhancer and super-enhancer analysis: To identify putative regulatory enhancers we generated chromatin state maps based on differential binding analysis of H3K27ac. We excluded regions +/−3 kb from TSS of RefSeq genes as well H3K4me3 peaks to eliminate promoters. Peaks with average normalized read intensity values across patients in the bottom 25th percentile were excluded to focus on highly enriched regions of H3K27ac. Typical enhancers were identified by executing the ROSE algorithm (23) using default parameters with a stitching distance of 12.5 and a TSS exclusion zone size of 0. Likewise, super-enhancers (SE) were identified as described previously (24), using the geometric inflection point to establish the cut-off. A simple linear model was fit to the ROSE output and the most significant genes closest to the SE were plotted with a FDR threshold of <0.2 and an absolute fold-change of >2.5.

Gene expression analysis: The TARGET database was used to segregate genes into up-regulated and down-regulated categories based on the log ratio of the fold change at relapse compared to diagnosis (Cohort A). RNAseq data was processed as described in Supplementary Information.

Correlation of gene expression with differential histone modification: Genes were paired on a one-to-one manner with promoter-associated (activation) or promoter and gene body-associated (repressive) histone peaks within individual samples and grouped based on types of histone changes seen between a diagnosis/relapse pair: dynamic gains or losses, exclusive gains or losses, and unchanged peaks. Box plots were generated to show the correlation of the different types of histone changes to the log2FC in gene expression. Wilcoxon test was performed to determine statistical significance.

Correlation of gene expression and DNA methylation changes: To correlate changes in gene expression with changes in promoter DNA methylation (Cohort B) we focused on those genes with robust changes in expression representing approximately 10% of the total number of transcripts. The normalized DESeq2 counts (cutoffs: low=≤4 and high=≥20) of protein coding genes were used to categorize genes based on the direction of expression changes from diagnosis to relapse: low to high (LH) and high to low (HL). The log2FC for each gene within the groups were calculated and used to generate a box plot. A box plot was also generated using the coherence values from the promoter-associated (+1kb TSS) methylation state of each gene within the two groups to show their correlation with log2FC in methylation (absolute log2FC>0.58). Wilcoxon test was performed to determine statistical significance.

Integration of gene expression data with ChIPseq and methylation data: The status for each specific histone mark and CpG methylation at the promoters of the up- and down-regulated genes (fold cut-off =1.5) for each sample pair were analyzed. For H3K27ac, changes in enhancers were also associated with the nearest genes. For each patient, the percentage of gains or losses in gene expression that directly correlated with changes in histone or DNA methylation status was determined. Bar graphs were generated to represent the percent correlation per patient, per mark. These were then combined to generate a stacked bar graph showing the overall contribution of epigenetic regulation of transcriptional output.

RESULTS

Extensive profiling of the chromatin state among matched diagnosis/relapse B-ALL pairs

We present data on two cohorts of matched diagnosis/relapse samples (Fig. 1a). Based on the number of live cells after thawing, immunoprecipitation was prioritized for H3K27ac and H3K27me3 marks. The mean number and size of peaks for each mark is enumerated in Fig. 1b and Supplementary Table S2. Although the total number of peaks varied widely among the samples (Fig. 1b), the numbers at diagnosis and relapse per sample were correlated with one another for all marks except H3K9me3, possibly due to the small number of samples analyzed (Fig. 1c, Supplemental Fig. S1). The width (genome occupancy) of peaks were greatest for H3K27me3 and H3K9me3 repressive marks compared to H3K4me3, H3K9ac and H3K27ac activation marks consistent with previous reports (25). These results indicate the regulatory chromatin network is quite diverse between samples from individual patients but that total peak number is relatively stable from diagnosis to relapse.

We also examined the distribution of histone peaks at diagnosis and relapse (Fig. 1d). We observed that 35-40% of peaks for activation marks (H3K4me3 and H3K9ac) were deposited within promoter regions at both time points, which is in contrast to 18% and 10% for repressive marks H3K9me3 and H3K27me3, respectively. Likewise, as expected, the majority of repressive marks (~50%) were in intergenic regions. As H3K27ac is a mark for promoters and enhancers, a significant number of peaks were also localized within gene bodies (~50%). Overall, no major differences were observed in the overall genomic region of occupancy between diagnosis and relapse.

Evolution of histone modifications from diagnosis to relapse

While the majority of histone marks were shared between diagnosis and relapse, a significant number of gains or losses were observed in all samples (Fig. 2a, Supplemental Table S3). Changes were most notable for H3K27me3 and H3K27ac (Fig. 2a). Overall, changes were evenly split between gains and losses as well as exclusive and dynamic changes. Examination of individual pairs revealed a striking degree of epigenetic diversity over time with some samples showing remarkable conservation and others revealing dramatic evolution from diagnosis to relapse (Supplemental Fig. S2a). This is not due to heterogeneity of samples (e.g. low blast percentage) as there was no significant correlation between difference in percent blasts and total changes (Supplemental Fig. S2b). As

expected, most samples contained relapsed-enriched mutations in epigenetic regulators, however no strict correlation between mutations in epigenetic regulators and epigenetic diversity over time was evident (Supplemental Table S4). Moreover, there appeared to be no correlation between the extent of changes between diagnosis and relapse and time to relapse. However, most of our samples were from children who relapsed within 36 months of diagnosis.

Interestingly, the majority of gains and losses were at promoter regions for all marks, except H3K9me3 (Fig. 2b vs. Fig. 1d) suggesting intergenic and gene body areas are more stable. To better understand changes in promoter states and the role of bivalent promoters, we focused on pairs of samples for which we had both H3K4me3 and H3K27me3 data (11 total) dividing promoters into repressed (H3K27me3), active (H3K4me3), bivalent (both) or absent (neither). Promoters defined as repressed, bivalent, or absent were more likely to demonstrate a shift compared to those defined as active. Bivalent promoters most often either remained bivalent or became active (Fig. 2c, Supplemental Fig. S3). For all the marks, the majority of the changes occurred at the promoters of protein coding genes (59-87.7%) although there were significant numbers of lncRNAs (9.4-34.2%) and miRNAs (1.97-7.7%) associated with the changed peaks (Fig. 2d, Supplemental Table S3). Unsupervised clustering of H3K9ac and H3K27ac promoter marks revealed that while the chromatin state was unique to each sample (e.g. the total number of peaks varied extensively from patient to patient) most pairs did not align next to one another suggestive of significant evolutionary drift over time for the majority of patients (21/31 for H3K27Ac; 12/22 for H3K9Ac) (Fig. 2e, Supplemental Fig. S4).

Promoter DNA hypermethylation at relapse

Analysis of DNA methylation state across different genomic regions revealed promoter associated CpGs were hypermethylated at relapse in the majority of the samples compared to diagnosis, which is in support of previous data from our lab and others (Fig. 3a). Unsupervised clustering showed that most pairs (12/17, 70%) clustered next to one another, indicating a highly personalized methylome for individual leukemia samples (note three relapse samples do not have associated diagnosis sample) (Fig. 3b). Five of 17 pairs, however, showed considerable divergence from diagnosis to relapse and these samples showed a greater median time to relapse (paired: 12.5 months, 95% CI=9.94-18.73; separated: 21 months, 95% CI=19.68-24.32, p=.0033) (Fig. 3c). Interestingly, only 7 out of 16 (44%) pairs clustered together based on H3K27ac (Fig. 2e), highlighting the importance of studying all layers of epigenetic regulation. GSEA identified polycomb repressive complex 2 (PRC2) targets as the one of the most significantly affected by DNA methylation changes with the majority of patients having hypermethylation of PRC2 complex target genes (Fig. 3d-e, Supplemental Table S5). The NOS and NANOG target pathways were shared only by those patients that had divergent methylomes from diagnosis to relapse, suggesting possible convergence over time (Fig. 3d, Supplemental Table S5).

Epigenetic changes correlate with alteration in gene expression

To understand the impact of epigenetic alterations on gene expression, we integrated already available gene expression microarray results from the TARGET database (cohort A) and

RNAseq (cohort B). As shown in Figure 4a, there was a positive correlation between differential H3K27ac peaks at gene promoters and changes in gene expression, which is consistent with previous reports (26). The correlation was more pronounced for exclusive changes compared to dynamic changes (Fig. 4a). H3K9ac differential promoter peaks also correlated with changes in gene expression, while H3K27me3 showed minimal negative correlation (Supplemental Fig. S5a). The lack of direct correlation of H3K4me3 with gene expression is not surprising, as a significant subset exists as bivalent promoters (range at: diagnosis 13.8-73.1%, median 37.9%; relapse 24.3-81.8%, median 33.5%) (Fig. 2d, Supplemental Fig. S3) (27,28). H3K9me3 did not correlate either; however, we had the least amount of samples for this mark.

Changes in DNA methylation at promoters in relapse were also correlated with alterations in gene expression. As expected, genes that were upregulated at relapse (low/high) had a loss of promoter CpG methylation while those that were downregulated at relapse (high/low) had a gain in promoter CpG methylation (Fig 4b). Based on our finding of hypermethylation of PRC2 targets, we assessed gene expression levels for the leading edge genes for each patient and found variability in the expected decrease in expression. In three of five patients the majority of the genes were indeed downregulated (absolute L2FC>0.32) while one patient showed many genes were paradoxically upregulated (Supplemental Fig. S5b). In comparison, the two patients who had hypomethylation of PRC2 targets had more upregulated genes (Supplemental Fig. S5b).

We also sought to understand the combined impact of alterations in chromatin and DNA methylation on changes in transcriptional output between diagnosis and relapse. Collectively, 20-30% of changes in gene expression correlated with changes in histone marks and an additional 5-10% correlated with DNA methylation differences (Supplemental Fig. S5c). However, these figures varied significantly across patients (Fig. 4c). When combining all histone and DNA methylation changes (cohort B), 23.6%-71.6% of the gains and 26.8%-61.7% of the losses in gene expression across patients were associated with expected changes in the epigenome. Collectively, these results further support our hypothesis that epigenetic reprogramming may have substantial impact on clonal evolution and it is highly variable among patients.

Convergent evolution of the epigenome correlates with pathways associated with drug resistance

To understand the extent to which the epigenome correlates with the downstream signaling state of leukemic blasts, we performed pathway analysis on each sample pair individually using differential gene expression or histone peaks at promoters for each mark and then determined the shared pathways. This analysis revealed known cancer driving pathways, such as p53, JAK/STAT, Ras, Wnt, and apoptosis for each of the histone marks, suggesting various epigenetic alterations converge on signaling networks previously shown to be associated with drug resistance (Supplemental Fig. S6) (9,29). These pathways were also shared by >20% of patients as assessed by differential gene expression (Supplemental Fig. S7). In fact, >50% of pathways impacted by differential H3K27ac and H3K9ac peaks were also identified using differential gene expression (Supplemental Fig. S7). We then

determined the extent to which shared pathways associated with H3K27ac promoter changes correlated with shared pathways associated with promoter changes in other histones, CpG methylation state, and gene expression. Pathway analysis of differential H3K27ac at promoters correlated best with pathways affected by gene expression and promoter changes of activating marks (H3K9ac, H3K4me3) and much less with the repressive marks (Fig. 5) consistent with their greater impact on gene expression. This analysis suggests that changes in different epigenetic marks likely facilitate evolution of similar transcriptional states and downstream pathways. For example epigenetic shifts are associated with changes in the expression of Ras, B cell activation, T cell activation, JAK/STAT (30), PI3K (31) and apoptotic programs all of which have been previously linked to acquisition of the drug resistant state.

We sought to determine the degree to which samples shared gain or loss of histone promoter marks at specific loci, which might reflect modulation of specific genes involved in tumor escape rather than convergence on downstream pathways. We did not find significantly shared differential H3K9ac promoter peaks among the whole cohort (FDR <0.05, abs L2FC≥0.58). However, when we separated patients into early relapse (≤18 months) vs. intermediate relapse (>18 months) we saw a significant number of shared differential peaks (4463) for intermediate relapse and none for early relapse (Fig. 6a). Likewise, when we examined H3K27ac promoter peaks we observed a significant number of shared differential promoters (238) among the intermediate relapses (FDR<0.05, abs L2FC≥0.58) (Fig. 6a). These results suggest that longer exposure to chemotherapy is more likely to result in modulation of specific genes that lead to later relapse. Pathway analysis (Fig. 6b) of these shared differential peaks among intermediate relapse patients revealed many of the same pathways noted above in analysis of the whole cohort and documented to be involved in drug escape. However, we also observed novel pathways such as pentose phosphate pathway, salvage pyrimidine ribonucleotides, and pyrimidine biosynthesis. In this regard, it is noteworthy that the maintenance phase of therapy begins about 6-9 months into therapy and is heavily dependent on methotrexate, which is an inhibitor of pyrimidine biosynthesis.

Identification of enhancers and super-enhancers associated with relapse

Overall, we identified 11317 unique enhancer regions across patients that were gained at relapse (fold-change cutoff=1.5), while only 405 regions were lost (fold-change cutoff=0.66). However, very few of these differential enhancers were shared among patients with early relapse (6) in contrast to patients with intermediate relapse where we identified 718 shared differential enhancer peaks (FDR<0.05, abs L2FC≥0.58) (Fig 6a). Pathway analysis on these shared enhancers again showed de novo pyrimidine synthesis as the top pathway as well as Wnt signaling (Fig. 6b), which has previously been associated with B-ALL drug resistance (9). Based on the *de novo* pyrimidine pathway having shared H3K9ac promoter and H3K27ac enhancer peaks, we compared histone peaks with RNA expression in genes that lead to identifying this pathway. As expected, we noted enrichment of H3K9ac promoter peaks for intermediate but not early relapse, but gene expression was increased at relapse compared to diagnosis in both early and intermediate relapse cases for all but one gene (Fig. 6c). The same trend was observed for 5/8 genes associated with the other H3K9ac associated pathways (Supplemental Fig. S8a). While the correlation was not as evident for

H3K27ac enhancer peaks, 3/6 genes showed a similar trend with enrichment for H3K27ac in intermediate relapse samples yet increased gene expression in both groups (Fig. 6c). This unexpected finding suggests that many factors may drive expression of key pathways early in therapy but more prolonged exposure to therapy might lead to epigenetic consolidation at specific loci. This pattern was less apparent for enhancer marks in Wnt signaling (Fig. S8b).

Across all patients, 3867 unique SEs were called with an average of 975 at diagnosis and 1111 at relapse per patient. Strikingly, we identified 3 gained SEs shared across a majority of patients overlapping *S100A8* (20/32), *RHAG* (17/32) and *HBG2/HBE1* (18/32) (Fig. 7a & c, Supplemental Fig. S9a & b). A gain/increase of SE correlated with an increase in expression of associated genes at relapse compared to diagnosis, whereas an unchanged/loss of the SE was associated with unchanged gene expression at relapse compared to diagnosis (Fig. 7b). This is consistent with previous data demonstrating significant upregulation of all three gene transcripts at relapse (8,32). These results linking relapse-specific activation of super-enhancers to the expression of their target genes across multiple patients provides a strong indication of their relevance in clonal evolution.

DISCUSSION

Relapsed leukemia is one of the leading causes of death in children with cancer. The major obstacle to effective therapy is tumor heterogeneity and the emergence of therapy resistant clones. Discoveries of relapse-enriched somatic variants have been enormously fruitful in elucidating pathways that drive tumor escape, but far fewer studies have focused on epigenetic alterations involved in the clonal evolution of drug resistant clones (29,33). This study provides global insights into the association of epigenetic alterations with individual genes and pathways known to drive chemoresistance in childhood B-ALL.

The importance of the epigenome in malignancy and drug resistance is supported by the discovery of somatic alterations that converge on epigenetic regulators. Mar et al showed that 57% of relapsed ALL patients harbored mutations in genes encoding epigenetic machinery (*SETD2*, *CREBBP*, *MSH6*, *KDM6A* and *MLL2*) with the majority representing a gain at relapse not seen at initial diagnosis (12). Likewise, Mullighan and colleagues observed that 18% of blasts from relapsed patients harbored deletions or somatic mutations of *CREBBP* (11). *CREBBP* missense mutations result in reduced expression of glucocorticoid-responsive genes. The availability of new agents targeting epigenetic machinery mandates study of how the epigenome responds to the selective pressures of therapy.

Our results indicate that the chromatin state varies considerably across patients, which is in line with the known genomic heterogeneity for various cancers including childhood leukemias. The fact that the majority of patient-specific patterns are conserved between diagnosis and relapse also indicates that the epigenetic state, like previous observations using transcriptome profiling, is highly individualized (34). However, we also show that considerable evolution of the epigenome can occur over time under the selective pressures of therapy. Some samples showed minimal differences from diagnosis to relapse while others showed massive reorganization of the chromatin state. Such variability did not correlate with

genetic subtype nor the presence of mutations in epigenetic regulators but we did not have detailed sequence information on all patients.

Another major goal of our study was to determine the extent to which epigenetic changes were associated with biological pathways previously shown to account for drug resistance as well as the identification of new pathways. In many respects epigenetic alterations might endow tumor subclones with greater flexibility to respond to the selective pressures of therapy. Overall, about one third of changes in gene expression directly correlated with epigenetic alterations with changes in H3K27ac having the best correlation as has been seen previously (26). We appreciate that this may be an underestimate since we focused on only the most robust changes in differential chromatin marks. However our results are consistent with previous work demonstrating that epigenetic changes may not have the expected correlation with expression indicating more subtle mechanisms of gene regulation (26,35-37). This is underscored by the fact that we showed correlation varied greatly among samples indicating that epigenetic reorganization may play a greater role in mediating changes in transcriptional output in some patients vs. others. Tumor heterogeneity is also known to impact correlation between epigenetic alterations and gene expression (14), therefore focusing exclusively on those chromatin and methylation changes that correlate with expression may miss important networks involved in clonal evolution. Epigenetic alterations reported here are disproportionately associated with pathways previously associated with relapse, such as the RAS/MAPK cascade (7,29).

Another interesting finding was DNA hypermethylation of PRC2 targets at relapse in a majority of patients. A recent pilot study evaluating epigenetic-targeted therapy in pediatric relapsed ALL found that the top pathway impacted by treatment was PRC2 targets. They observed reversal of promoter DNA hypermethylation and subsequent upregulation of gene expression for PRC2 target genes following treatment with decitabine and vorinostat (37). These data support a role for DNA methylation in transcriptional regulation of genes influenced by PRC2. There were five patients that had divergent evolution of their methylome between diagnosis and relapse, all of whom had intermediate relapses, possibly indicating convergent evolution. Of note, DNA hypermethylation of NOS and NANOG target genes, which have been implicated in maintenance of cancer stem cells and epigenetic reprogramming (38), was observed in only those patients that had divergent evolution.

Importantly our study also revealed novel pathways as candidate drivers of relapse. It was quite interesting to note that there were no shared individual differential H3K9ac promoter peaks for early relapse and only 6 shared enhancer marks. In contrast, patients with intermediate relapse showed many shared differential promoter and enhancer marks. This suggests that later relapse where selection takes place over an extended period of time might lead to outgrowth of clones with modulation of distinct loci. A particularly novel finding was the shared differentially histone peaks associated with genes involved in *de novo* pyrimidine biosynthesis. The extended pressure of methotrexate, an inhibitor of pyrimidine biosynthesis and a cornerstone of maintenance chemotherapy, might explain this observation. This interpretation is strengthened by the fact that previous observations also revealed upregulation of genes involved in nucleotide biosynthesis and folate metabolism in late but not early ALL relapse (8). When comparing the impact of histone peak changes at

specific loci with gene expression we came upon an unexpected finding. While we observed enrichment of H3K9ac promoter peaks in intermediate relapse, expression was increased at relapse for both early and intermediate relapse cases for five of the six genes in this pathway. Similar findings were also seen for genes in additional pathways identified by enrichment of H3K9ac promoter peaks in intermediate but not early relapse (pentose phosphate, salvage pyrimidine, heterotrimeric G-signaling). This indicates that myriad factors may drive resistance pathways in early relapse but the ongoing selective pressures of chemotherapy may lead to epigenetic "consolidation" of specific promoters possibly through outgrowth of subclones that drive evolution. Our analysis also indicated enrichment of H3K27ac ehancer peaks shared across patients for genes in the *de novo* pyrimidine biosynthesis pathway but here the correlation was less striking. Some of the disscrepency may be related to the fact that it is difficult to assign functional significance to H3K27ac peaks outside promoter regions and boxplots represents the most differential peak out of multiple peaks within the region of interest.

One of the most startling findings of our study was the discovery of three novel SE shared by a majority of patients at relapse. The shared SE at *S100A8* is particularly noteworthy as a meta-analysis of gene expression data from diagnosis/relapse pairs, including our prior work, indicated that *S100A8* was the top differentially expressed transcript and is part of the S100 calcium EF-hand superfamily (32). S100A8 can function as a complex with S100A9 (calprotectin) to buffer cytosolic calcium. Studies have associated elevated *S100A8* with poor outcome in AML as well as ALL (39,40). S100A8/S100A9 has been associated with glucocorticoid resistance in *KMT2A* rearranged infant ALL (41). Likewise, expression of *S100A8* and *S100A9* correlate with resistance to conventional agents and venetoclax in AML (42). Our findings along with previous work indicates that targeting S100A8 is likely to be a promising strategy to restore chemosensitivity.

We also identified a novel shared SE adjacent to the β-globin locus, composed of the *HBE*, *HBG1/2*, *HBD* and *HBB* genes, with upregulation of *HBG2*. We and others have noted upregulation of globin transcripts *HBG1/2*, *HBE* and *HBB* at relapse (8,32). Zheng et al. noted the induction of *HBB* specifically in circulating tumor cells from patients with breast, prostate and lung cancers (43). Increased intracellular reactive oxygen species (ROS) triggered the induction of *HBB* and depletion of *HBB* dramatically increased apoptosis following ROS exposure, a by-product of both chemo- and radiation therapy. Similarly, Park et al noted overexpression of HBE1 in radio-resistant colon cancer cell lines and correlated expression with decreased ROS in response to irradiation (44). Our observations raise the possibility for a similar role for members of the β–globin family in mediating chemoresistance in B-ALL.

The impact of the RHAG associated SE on clonal evolution is not obvious although both *RHAG* (encoding a glycoprotein that is part of the Rh complex) and *CRISP2* (encoding Cysteine Rich Protein 2, a cancer testis antigen) transcripts have been noted to be upregulated at relapse in multiple studies (8,32). These SEs and their neighboring genes are the focus of ongoing functional studies to validate their roles in B-ALL clonal evolution.

We recognize the limitations of conventional ChIPseq when quantitatively comparing histone levels between samples. Newer techniques are now available but when our project was initiated we were limited by cell numbers from banked samples (45). When examining shifts in the epigenetic state between diagnosis and relapse we were quite conservative to select only the most robust differences to avoid artifacts. It is also noteworthy that we found a strong correlation between peak number and genomic localization in the diagnosis and relapse sample of each patient. Finally, the correlation between epigenetic state and gene expression as seen by others validates our approach. Another potential confounding variable is sample purity since we did not separate blasts from non-leukemic hematopoietic cells for fear of depleting cell numbers. The majority of samples had >90% blasts and were separated by ficoll-paque techniques in COG cell bank and when there was a discrepancy we did not find a correlation between the magnitude of differences in blast percentage between diagnosis and relapse and the scale of epigenetic evolution.

In summary, we show that the chromatin state varies widely among samples but that substantial reorganization of the epigenome takes place under the selective pressures of therapy in the majority of patients. In spite of the differences among patients and lack of an overarching epigenomic signature, our data does suggest convergent evolution leads to shared genes/pathways that appear to mediate drug resistance. In addition to previously discovered pathways involved in drug resistance that are modulated by epigenetic dysregulation, our studies implicate changes in pyrimidine biosynthesis, suppression of PRC2 targets, and enhancer-driven upregulation of S100A8 as well as members of the β-globin and RHAG loci as potenetialy playing a major role in clonal evolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We gratefully acknowledge the funding received to complete this work. WLC is supported by the National Cancer Institute of Health (R01 CA140729-05), the Perlmutter Cancer Center Arline and Norman M. Feinberg Pilot Grant for Lymphoid Malignancies, and The Leukemia and Lymphoma Society Specialized Center for Research (7010-14). S.S. was supported by the Conquer Cancer Foundation of ASCO-YIA. T.B. was supported by the Pediatric Cancer Foundation. A.T. is supported by the American Cancer Society (RSG-15-189-01-RMC), St. Baldrick's foundation (581357) and NCI/NIH P01CA229086-01A1. We gratefully acknowledge the support of the Sohn Conference Foundation as well as the Genome Technology Center (GTC) for expert library preparation and sequencing, and the Applied Bioinformatics Laboratories (ABL) for providing bioinformatics support. We also acknowledge the assistance of the New York Genome Center for whole genome and enhanced reduced representation bisulfite sequencing. GTC and ABL are shared resources partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. This work has used computing resources at the NYU School of Medicine High Performance Computing Facility.

REFERENCES

1. Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, et al. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2012;30:1663–9 [PubMed: 22412151]
2. Dores GM, Devesa SS, Curtis RE, Linet MS, Morton LM. Acute leukemia incidence and patient survival among children and adults in the United States, 2001-2007. Blood 2012;119:34–43 [PubMed: 22086414]

3. Freyer DR, Devidas M, La M, Carroll WL, Gaynon PS, Hunger SP, et al. Postrelapse survival in childhood acute lymphoblastic leukemia is independent of initial treatment intensity: a report from the Children's Oncology Group. Blood 2011;117:3010–5 [PubMed: 21193696]
4. Pulsipher MA, Langholz B, Wall DA, Schultz KR, Bunin N, Carroll W, et al. Risk factors and timing of relapse after allogeneic transplantation in pediatric ALL: for whom and when should interventions be tested? Bone marrow transplantation 2015;50:1173–9 [PubMed: 25961775]
5. Bhatla T, Jones CL, Meyer JA, Vitanza NA, Raetz EA, Carroll WL. The Biology of Relapsed Acute Lymphoblastic Leukemia: Opportunities for Therapeutic Interventions. J Pediatr Hematol Oncol 2014;36:413–8 [PubMed: 24942023]
6. van Galen JC, Kuiper RP, van Emst L, Levers M, Tijchon E, Scheijen B, et al. BTG1 regulates glucocorticoid receptor autoinduction in acute lymphoblastic leukemia. Blood 2010;115:4810–9 [PubMed: 20354172]
7. Irving J, Matheson E, Minto L, Blair H, Case M, Halsey C, et al. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. Blood 2014;124:3220–430
8. Hogan LE, Meyer JA, Yang J, Wang J, Wong N, Yang W, et al. Integrated genomic analysis of relapsed childhood acute lymphoblastic leukemia reveals therapeutic strategies. Blood 2011;118:5218–26 [PubMed: 21921043]
9. Dandekar S, Romanos-Sirakis E, Pais F, Bhatla T, Jones C, Bourgeois W, et al. Wnt inhibition leads to improved chemosensitivity in paediatric acute lymphoblastic leukaemia. British journal of haematology 2014;167:87–99 [PubMed: 24995804]
10. Bhatla T, Wang J, Morrison DJ, Raetz EA, Burke MJ, Brown P, et al. Epigenetic reprogramming reverses the relapse-specific gene expression signature and restores chemosensitivity in childhood B-lymphoblastic leukemia. Blood 2012;119:5201–10 [PubMed: 22496163]
11. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature 2011;471:235–9 [PubMed: 21390130]
12. Mar BG, Bullinger LB, McLean KM, Grauman PV, Harris MH, Stevenson K, et al. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. Nature communications 2014;5:3469
13. Jaffe JD, Wang Y, Chan HM, Zhang J, Huether R, Kryukov GV, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nature genetics 2013;45:1386–91 [PubMed: 24076604]
14. Pastore A, Gaiti F, Lu SX, Brand RM, Kulm S, Chaligne R, et al. Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL. Nature communications 2019;10:1874
15. Beekman R, Chapaprieta V, Russinol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, et al. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat Med 2018;24:868–80 [PubMed: 29785028]
16. Jin Y, Chen K, De Paepe A, Hellqvist E, Krstic AD, Metang L, et al. Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. Blood 2018;131:2138–50 [PubMed: 29519805]
17. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat Med 2016;22:792–9 [PubMed: 27322744]
18. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods 2015;12:963–5 [PubMed: 26280331]
19. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 2015;31:2382–3 [PubMed: 25765347]
20. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 2012;481:389–93 [PubMed: 22217937]

21. Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet 2012;8:e1002781 [PubMed: 22737091]
22. Li W, Cong Q, Kinch LN, Grishin NV Seq2Ref: a web server to facilitate functional interpretation. BMC Bioinformatics 2013;14
23. Hnisz D, Abraham B, Lee T, Lau A, Saint-Andre V, Sigova A, et al. Super-Enhancers in the Control of Cell Identity and Disease. Cell 2013;155:934–47 [PubMed: 24119843]
24. Whyte WO DA; Hnisz D; Rahl PB; Lee TI; Young RA. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell 2013;153:307–19 [PubMed: 23582322]
25. Kimura H Histone modifications for human epigenome analysis. J Hum Genet 2013;58:439–45 [PubMed: 23739122]
26. Jing D, Huang Y, Liu X, Sia KCS, Zhang JC, Tai X, et al. Lymphocyte-Specific Chromatin Accessibility Pre-determines Glucocorticoid Resistance in Acute Lymphoblastic Leukemia. Cancer Cell 2018;34:906–21 e8 [PubMed: 30537513]
27. Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. Genes Dev 2013;27:1318–38 [PubMed: 23788621]
28. Bernhart SH, Kretzmer H, Holdt LM, Juhling F, Ammerpohl O, Bergmann AK, et al. Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. Sci Rep 2016;6:37393 [PubMed: 27876760]
29. Jones CL, Gearheart CM, Fosmire S, Delgado-Martin C, Evensen NA, Bride K, et al. MAPK signaling cascades mediate distinct glucocorticoid resistance mechanisms in pediatric leukemia. Blood 2015;126:2202–12 [PubMed: 26324703]
30. Irving JA. Towards an understanding of the biology and targeted treatment of paediatric relapsed acute lymphoblastic leukaemia. British journal of haematology 2016;172:655–66 [PubMed: 26568036]
31. Shalapour S, Hof J, Kirschner-Schwabe R, Bastian L, Eckert C, Prada J, et al. High VLA-4 expression is associated with adverse outcome and distinct gene expression changes in childhood B-cell precursor acute lymphoblastic leukemia at first relapse. Haematologica 2011;96:1627–35 [PubMed: 21828124]
32. Chow YP, Alias H, Jamal R. Meta-analysis of gene expression in relapsed childhood B-acute lymphoblastic leukemia. BMC Cancer 2017;17:120 [PubMed: 28183295]
33. Meyer JA, Wang J, Hogan LE, Yang JJ, Dandekar S, Patel JP, et al. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nature genetics 2013;45:290–4 [PubMed: 23377183]
34. Li S, Garrett-Bakelman F, Perl AE, Luger SM, Zhang C, To BL, et al. Dynamic evolution of clonal epialleles revealed by methclone. Genome Biol 2014;15:472 [PubMed: 25260792]
35. Rauluseviciute I, Drablos F, Rye MB. DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. BMC Med Genomics 2020;13:6 [PubMed: 31914996]
36. Rendeiro AF, Schmidl C, Strefford JC, Walewska R, Davis Z, Farlik M, et al. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. Nature communications 2016;7:11938
37. Burke MJ, Kostadinov R, Sposto R, Gore L, Kelley SM, Rabik C, et al. Decitabine and Vorinostat with Chemotherapy in Relapsed Pediatric Acute Lymphoblastic Leukemia: A TACL Pilot Study. Clin Cancer Res 2020
38. Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, et al. Targeting cancer stem cell pathways for cancer therapy. Signal Transduct Target Ther 2020;5:8 [PubMed: 32296030]
39. Ouyang MF, Wang D, Liu YT, Xu LY, Zhao MY, Yin XC, et al. [Value of S100A8 in evaluating the prognosis of children with acute lymphoblastic leukemia]. Zhongguo Dang Dai Er Ke Za Zhi 2019;21:359–64 [PubMed: 31014429]
40. Yang XY, Zhang MY, Zhou Q, Wu SY, Zhao Y, Gu WY, et al. High expression of S100A8 gene is associated with drug resistance to etoposide and poor prognosis in acute myeloid leukemia through influencing the apoptosis pathway. Onco Targets Ther 2016;9:4887–99 [PubMed: 27540302]

41. Spijkers-Hagelstein JA, Schneider P, Hulleman E, de Boer J, Williams O, Pieters R, et al. Elevated S100A8/S100A9 expression causes glucocorticoid resistance in MLL-rearranged infant acute lymphoblastic leukemia. Leukemia 2012;26:1255–65 [PubMed: 22282267]

42. Karjalainen R, Liu M, Kumar A, He L, Malani D, Parsons A, et al. Elevated expression of S100A8 and S100A9 correlates with resistance to the BCL-2 inhibitor venetoclax in AML. Leukemia 2019;33:2548–53 [PubMed: 31175323]

43. Zheng Y, Miyamoto DT, Wittner BS, Sullivan JP, Aceto N, Jordan NV, et al. Expression of beta-globin by cancer cells promotes cell survival during blood-borne dissemination. Nature communications 2017;8:14344

44. Park SY, Lee SJ, Cho HJ, Kim JT, Yoon HR, Lee KH, et al. Epsilon-Globin HBE1 Enhances Radiotherapy Resistance by Down-Regulating BCL11A in Colorectal Cancer Cells. Cancers (Basel) 2019;11

45. van Galen P, Viny AD, Ram O, Ryan RJ, Cotton MJ, Donohue L, et al. A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes. Mol Cell 2016;61:170–80 [PubMed: 26687680]

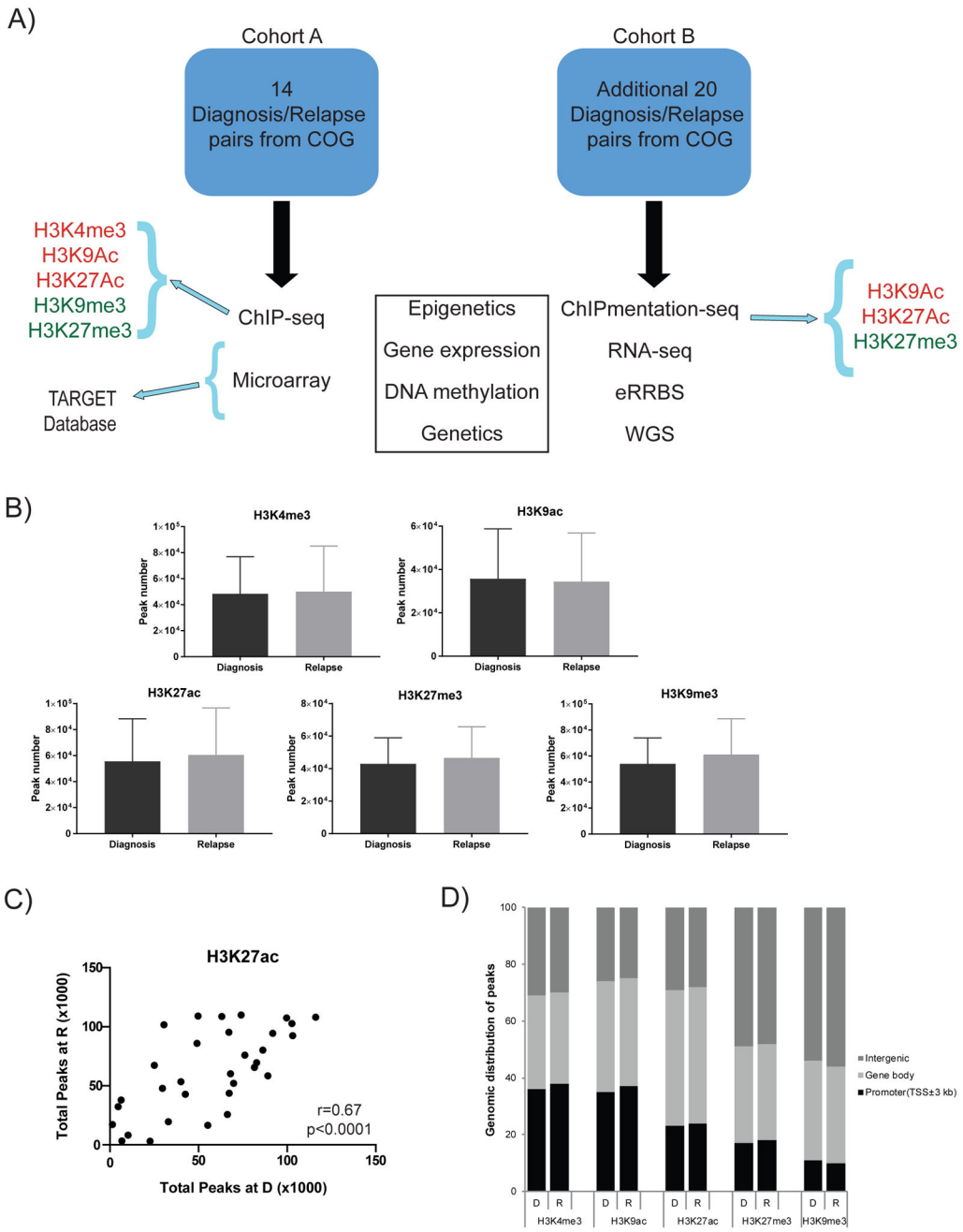

Figure 1: Mapping chromatin marks in leukemia samples.

A) Technical approaches in two cohorts of patient samples. **B)** Similarity in average number of peaks for each chromatin mark at diagnosis and relapse. **C)** Correlation between total number of H3K27ac peaks at diagnosis and relapse. Each dot represents an individual patient. Pearson correlation test performed to determine r and p-values. **D)** Bar graph representing the percentage of peak occupancy in the promoter regions (±3 kb up and down stream of TSS), gene body and intergenic regions for H3K9ac, H3K27ac, H3K4me3, H3K27me3, and H3K9me3 at diagnosis and relapse.

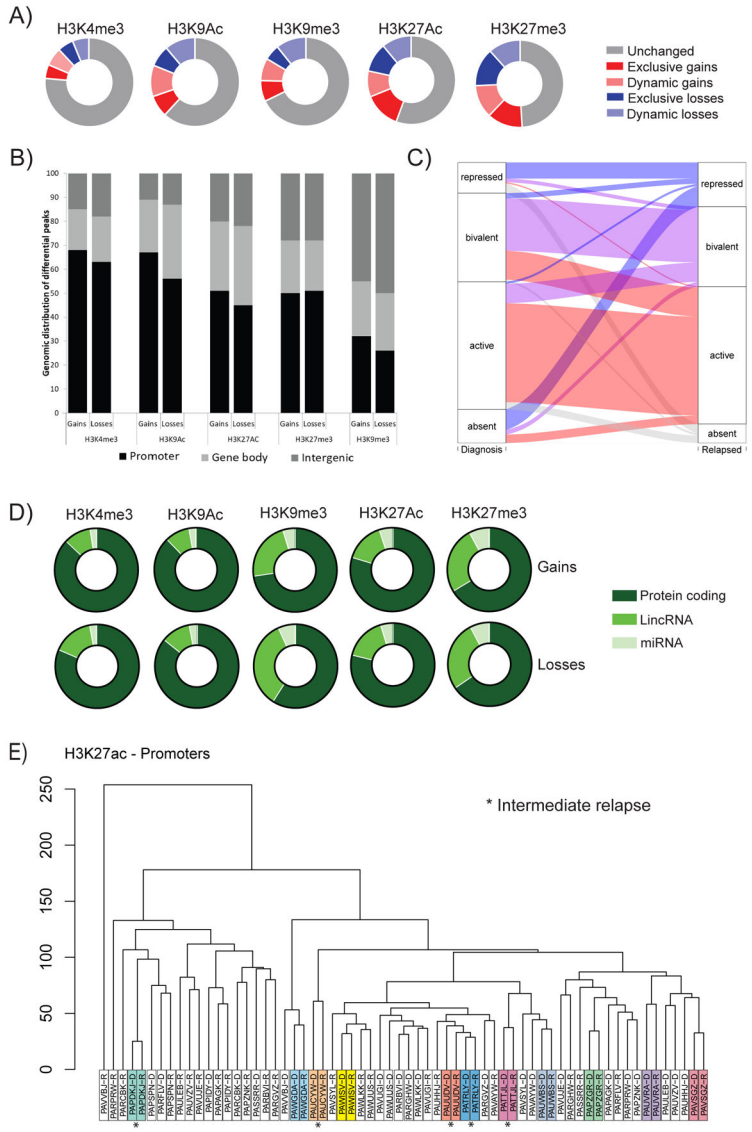

Figure 2: Evolution of the chromatin state between diagnosis and relapse.

A) Average percentage of shifts in histone peaks for five marks. Grey unchanged, red gains at relapse (dark exclusive, light dynamic) and blue losses at relapse (dark exclusive, light dynamic). **B)** Genomic distribution of gains and losses for the five histone marks studied. **C)** Average shifts in promoter categories (repressed=H3K27me3 only, active=H3K4me3 only, bivalent=K27me3+K4me3, absent=neither) between diagnosis (left) and relapse (right). **D)** Types of transcripts being regulated by promoter (+/− 3 kb TSS) gains or losses. **E)** Unsupervised clustering of all samples using the top 5,000 differential H3K27ac promoter peaks. Patients that aligned next to one another are colored and pairs with intermediate relapse (>18 months from diagnosis) are indicated by an asterisk.

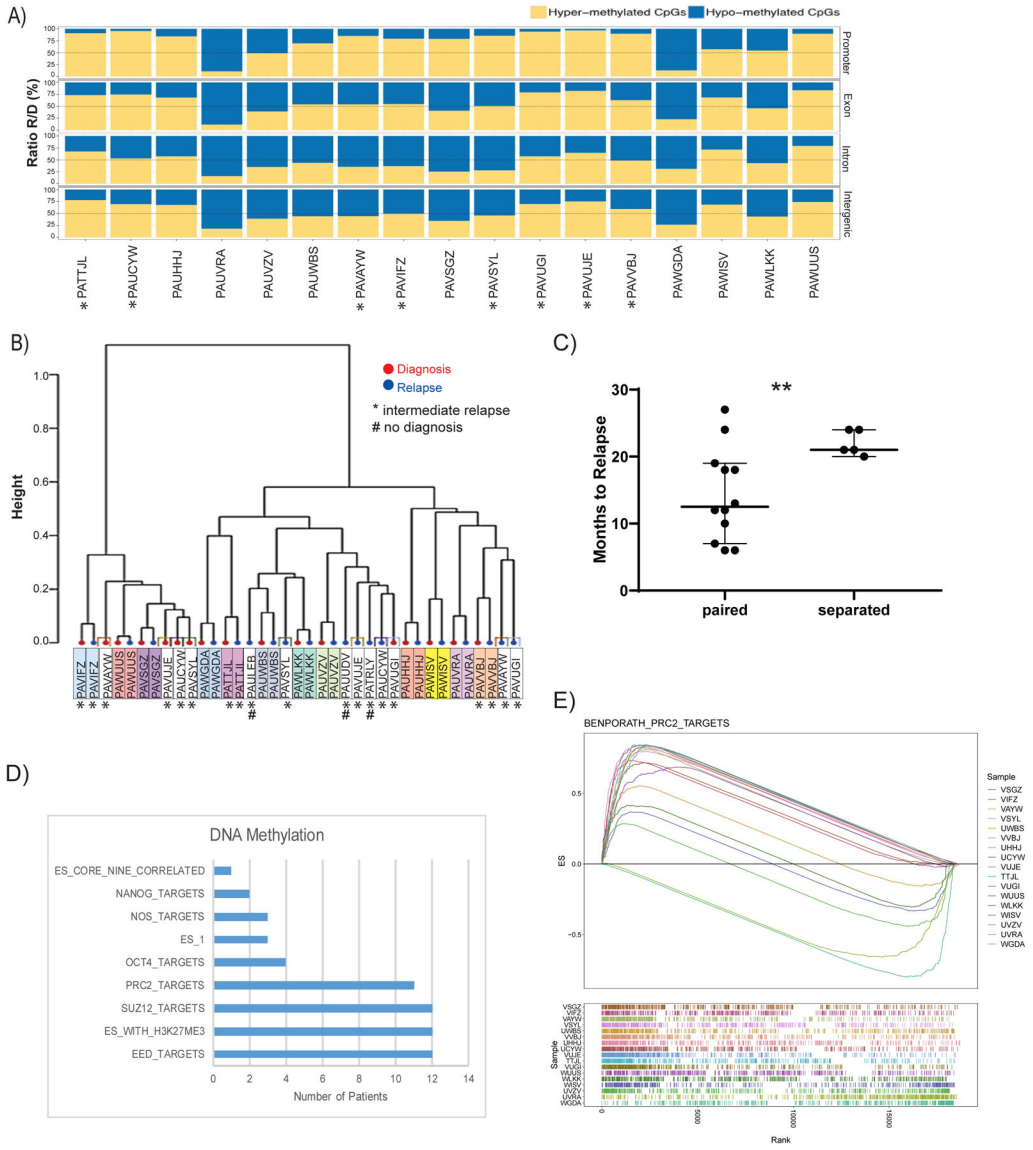

Figure 3: Changes in DNA methylation from diagnosis to relapse.

A) Location of DNA methylation changes (promoter, exon, intron and intergenic) shows that most promoters are hyper-methylated at relapse. Blue indicates loss of methylation (hypo-methylation) at relapse while yellow indicates gain (hyper-methylation). **B)** Unsupervised clustering of promoter methylation marks among all samples. Diagnosis samples labeled in red, relapse in blue. # represent relapse samples without associated diagnosis sample. Patients that aligned next to one another are colored and pairs with intermediate relapse (>18 months from diagnosis) are indicated by an asterisk. **C)** Time to relapse (from diagnosis) for samples that aligned next to one another on unsupervised clustering analysis (paired) vs. those that were separated indicating that those samples with greater differences in methylation between diagnosis and relapse were associated with greater time to relapse. Unpaired t test with Welch's correction was performed to determine p-value. **D)** GSEA per patient using median difference between each within promoter regions (TSS+−1 kb). Shared

Benporath significant (q<0.01) pathways are shown. **E)** GSEA for all patients indicates that PRC2 targets are selectively hypermethylated at relapse in 11 out of 17 pairs.

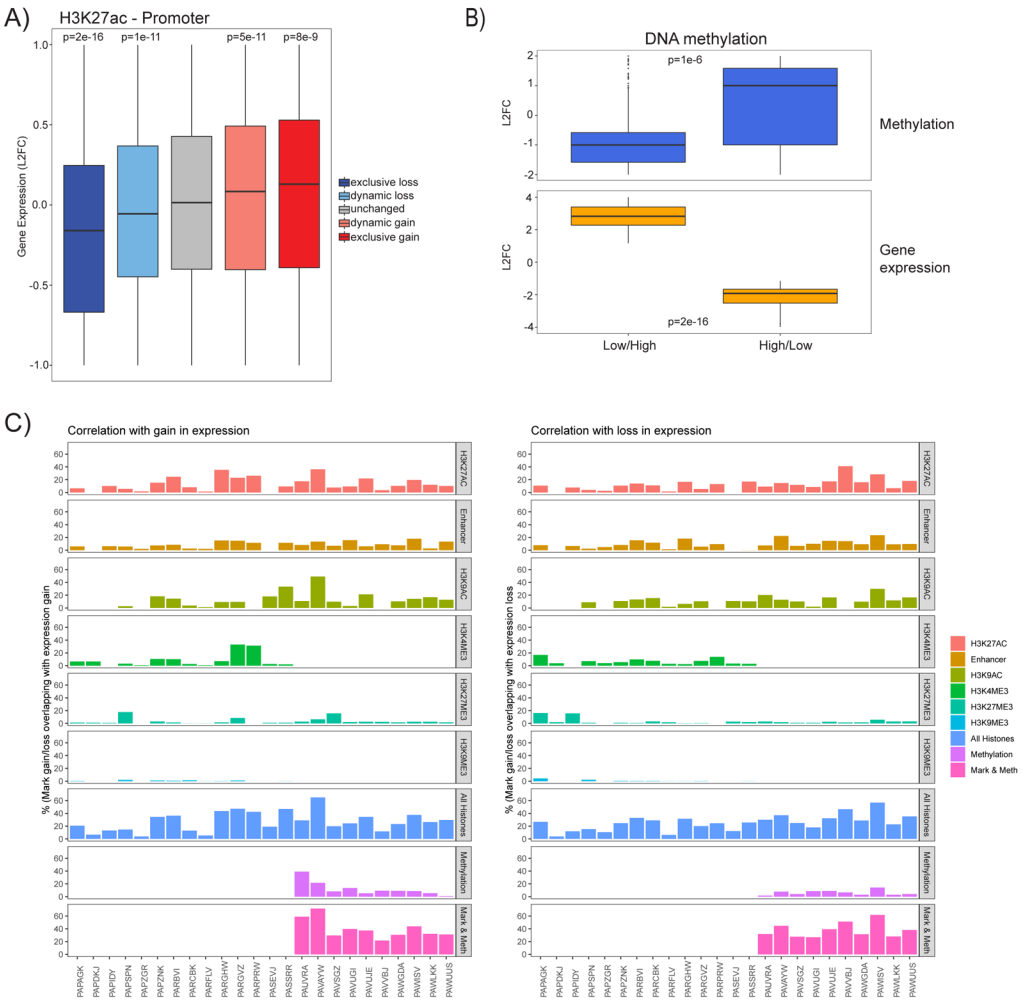

Figure 4: Correlation of epigenetic changes with gene expression.

A) Correlation of Log2 fold changes in gene expression with losses or gains of H3K27ac promotor peaks (blue=loss, grey=no change, and red=gain of mark (light or dark shading indicates dynamic or exclusive changes respectively)). **B)** Log2 fold changes in DNA methylation (top panel) and gene expression (bottom panel) for promoters whose gene expression are increased at relapse (Low/High) (left) vs. those genes whose expression decreases at relapse (High/Low) (right). Wilcox test was performed to determine significance. **C) Left panel**. Contribution of gains of individual activation marks (H3K27ac, H3K9ac, H3K4me3) and losses of repressive marks (H3k27me3, H3K9me3), all histones, loss of promoter DNA methylation and combined histone/methylation changes to increases in gene expression at relapse compared to diagnosis. **Right panel**. Contribution of loss of individual activation marks, gain of repressive marks, all histones, gain of promoter DNA methylation and combined histone/methylation to loss of gene expression at relapse compared to diagnosis.

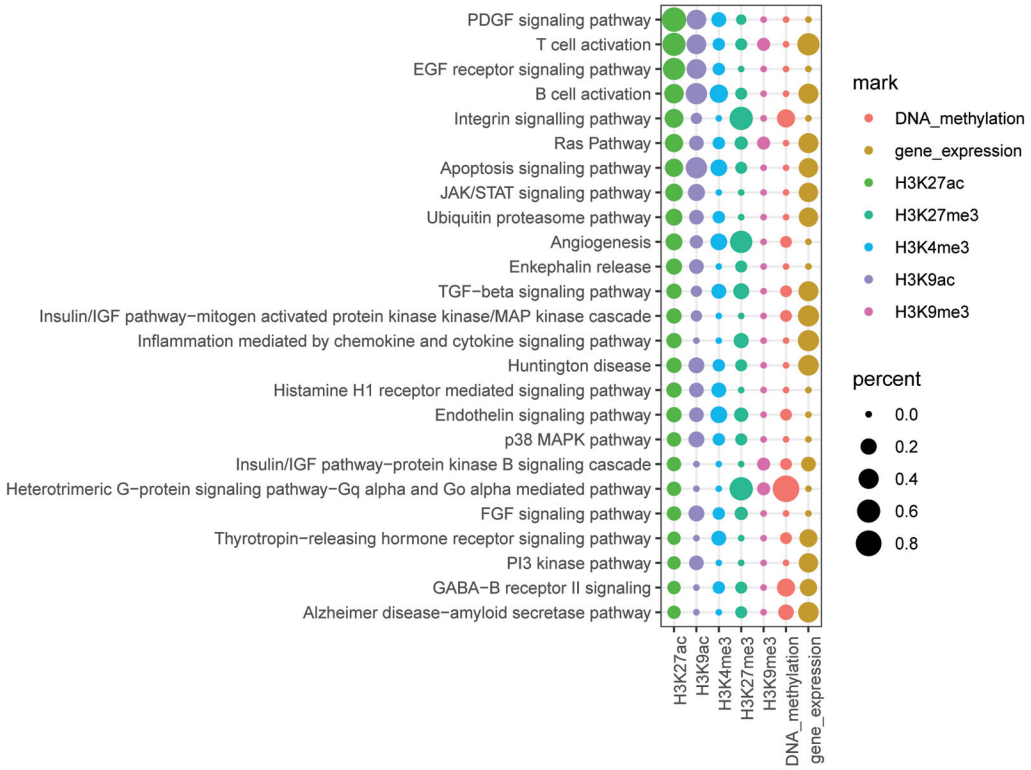

Figure 5: Pathway analysis of alterations in promoter-based epigenetic marks.

Pathway analysis of changes in H3K27ac at promoters first column followed by the extent to which changes in other histone marks, DNA methylation, and gene expression correlated with the H3K27ac driven pathways. The size of each bubble reflects the extent to which changes in the promoter marks for genes in each pathway was shared across patients. Source of pathways was Panther.

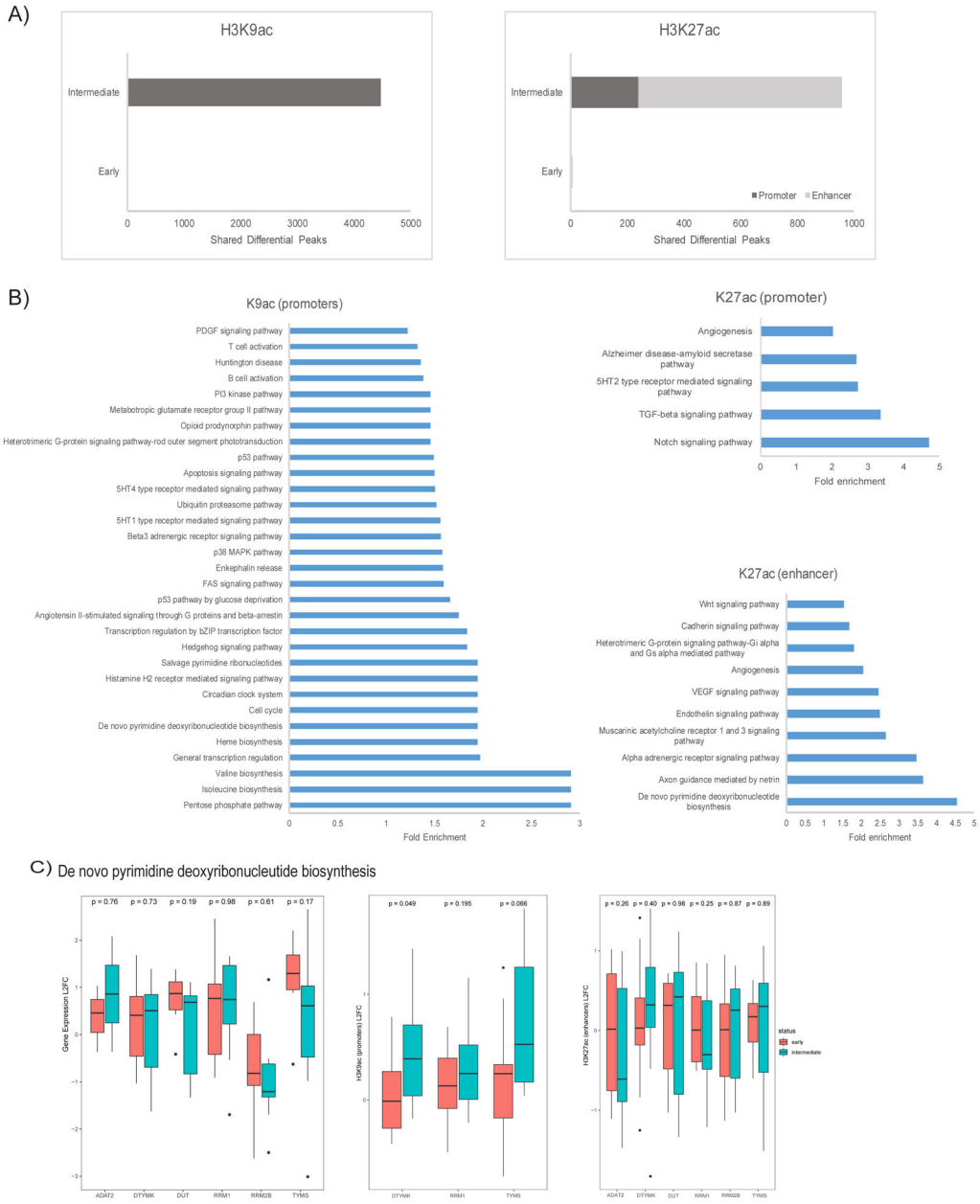

Figure 6: Convergent changes in promotor and enhancer marks shared across patients.

A) Left Panel. Number of differential H3K9ac promoter peaks shared by patients (FDR<0.05, absolute L2FC ≥0.58) for intermediate (>18 months; n=12) and early (≤18 months; n=10) relapse. **Right Panel**. Shared differential H3K27ac promoter (dark grey) and enhancer peaks (light grey) for patients who relapse intermediate (>18 months; n=18) vs. early (≤18 months; n=14) (FDR <0.05, absolute L2FC ≥0.58). **B) Left panel**. GREAT Panther pathway analysis of shared differential H3K9ac promoter peaks. **Right panel**. GREAT Panther pathway analysis of shared differential H3K27ac promoter (top right) and enhancer (bottom right) peaks. All significant pathways shown (hypergeometric raw p<0.05). **C)** The Log2 fold change (L2FC) for gene expression, H3K27ac, and H3K9ac across both early and intermediate patients is shown for those genes associated with shared

differential peaks that led to the identification of pathways shown in panel (B). H3K27ac peaks were restricted to enhancers (>+/− 3kb TSS) and H3K9ac peaks were restricted to promoters (<+/−3kb TSS). The peak with greatest absolute L2FC is shown as representative for each gene. Paired t test was performed to determine significance.

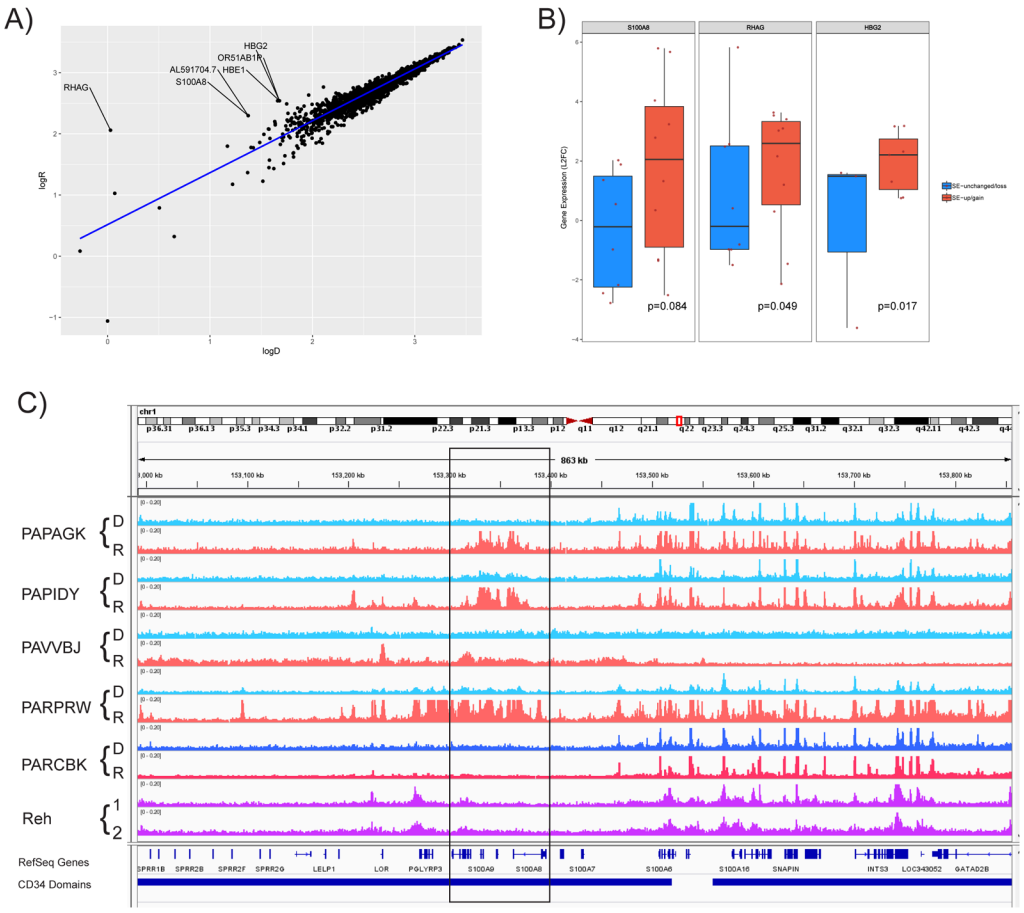

Figure 7: Shared relapse-specific super-enhancers.

A). Log fold peaks of SE at relapse (Y axis) vs. diagnosis (X axis). Three SE show significant increases at relapse (labeled according to neighboring genes (RHAG, S100A8 and HBG2). **B)** Correlation of gene expression for S100A8, RHAG and HBG2 (Y axis) and samples with unchanged or loss of SE (blue) or gain of SE (red). Wilcox test was performed to determine significance. **C)** IGV image showing the H3K27ac peak intensity at S100A8 locus (black box) at diagnosis (turquoise) and relapse (pink). Top four samples representative of SE gains and fifth sample without SE gain (diagnosis-blue, relapse-red) for comparison. Bottom two tracks show results from two ChIPseq experiments with the REH B-ALL cell line (purple).