Published in final edited form as:

Cancer Cell. 2018 December 10; 34(6): 982–995.e7. doi:10.1016/j.ccell.2018.11.001.

Enhancer architecture and essential core regulatory circuitry of chronic lymphocytic leukemia

Christopher J. Ott[1,2,3,4,8,9,10], **Alexander J. Federation**[1,5,8], **Logan S. Schwartz**[1,2], **Siddha Kasar**[1], **Josephine L. Klitgaard**[1], **Romina Lenci**[1], **Qiyuan Li**[1,6], **Matthew Lawlor**[1,2], **Stacey M. Fernandes**[1], **Amanda Souza**[1], **Donald Polaski**[1], **Deepti Gadi**[1], **Matthew L. Freedman**[1], **Jennifer R. Brown**[1,3,9], and **James E. Bradner**[1,3,4,7,9]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, 02115, USA

[2]Massachusetts General Hospital Cancer Center, Boston, MA, 02129, USA

[3]Department of Medicine, Harvard Medical School, Boston, MA, 02115, USA

[4]Broad Institute of MIT & Harvard, Cambridge, MA, 02142, USA

[5]Altius Institute for Biomedical Sciences, Seattle, WA, 98121, USA

[6]Medical School, Xiamen University, Xiamen, 361102, China

[7]Novartis Institutes for Biomedical Research, Cambridge, MA, 02139, USA

[8]These authors contributed equally

[9]Co-corresponding authors

[10]Lead contact

SUMMARY

Enhancer profiling is a powerful approach for discovering *cis*-regulatory elements that define the core transcriptional regulatory circuits of normal and malignant cells. Gene control through enhancer activity is often dominated by a subset of lineage-specific transcription factors. By integrating measures of chromatin accessibility and enrichment for H3K27 acetylation, we have generated regulatory landscapes of chronic lymphocytic leukemia (CLL) samples and

Correspondence: Christopher J. Ott, PhD, Massachusetts General Hospital Cancer Center, Harvard Medical School, 149 13th St., Charlestown, MA 02129, USA, christopher.ott@mgh.harvard.edu; James E. Bradner, MD, Novartis Institutes for Biomedical Research, 250 Massachusetts Ave., Cambridge, MA 02139, USA, james.bradner@novartis.com; Jennifer R. Brown, MD, PhD, Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston, MA 02115, USA, jennifer_brown@dfci.harvard.edu.

AUTHOR CONTRIBUTIONS

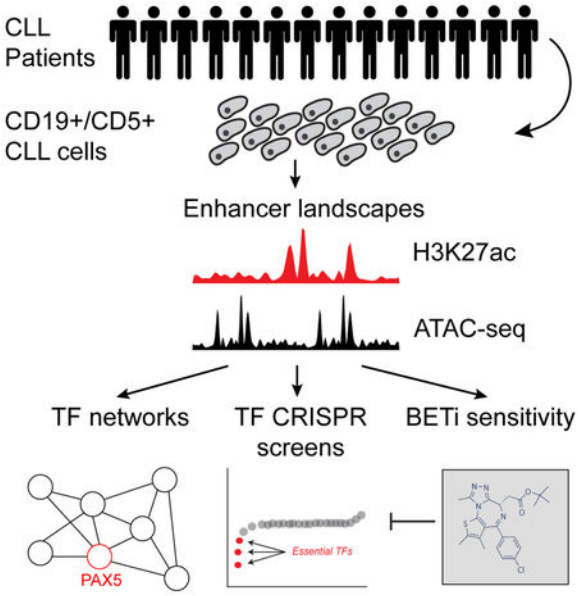
Conceptualization, C.J.O, A.J.F, J.R.B., and J.E.B.; Methodology, C.J.O., A.J.F., J.R.B., and J.E.B.; Software and formal analysis, C.J.O, A.J.F., S.K., M.L., L.S., and Q.L.; Investigation, C.J.O, A.J.F, S.K., J.L.K, R.L., M.L., D.G., L.S., and Q.L.; Resources, S.M.F and J.R.B.; Writing – Original Draft, C.J.O.; Writing – Review & Editing, C.J.O, A.J.F., M.L.F., J.R.B, and J.E.B.; Funding acquisition, C.J.O, M.L.F, J.R.B, and J.E.B.; Supervision, C.J.O, M.L.F, J.R.B., and J.E.B.

DECLARATION OF INTERESTS

J.E.B is now an executive and shareholder of Novartis AG, and has been a founder and shareholder of SHAPE (acquired by Medivir), Acetylon (acquired by Celgene), Tensha (acquired by Roche), Syros, Regenacy and C4 Therapeutics.

representative cell lines. With super enhancer-based modeling of regulatory circuits and assessments of transcription factor dependencies, we discover that the essential super enhancer factor PAX5 dominates CLL regulatory nodes and is essential for CLL cell survival. Targeting enhancer signaling via BET bromodomain inhibition disrupts super enhancer-dependent gene expression with selective effects on CLL core regulatory circuitry, conferring potent anti-tumor activity.

Graphical Abstract



In Brief

Analyzing integrative enhancer profiles and transcription factor dependencies, Ott et al. construct enhancer-based core regulatory circuits of chronic lymphocytic leukemia (CLL) and reveal a dominant and essential role for PAX5. BET inhibition disrupts CLL super enhancer networks and suppresses CLL growth.

INTRODUCTION

Measurements of epigenomic features, such as open chromatin regions and histone modifications, are useful indicators of genomic regulatory elements important for maintenance of cellular transcriptional states. This includes enhancers, non-coding *cis*-regulatory elements that are bound by multiple transcription factors (TFs) to control cell-type-specific gene expression through direct association with gene promoters (Spitz and Furlong, 2012). In-depth analyses of enhancer landscapes have begun to elucidate the mechanisms of cell-type-specific gene expression, including the dynamics of transcription factor binding and chromatin organization during tissue differentiation, and their contribution to the etiology and progression of disease (Corces et al., 2016; Roadmap Epigenomics Consortium et al., 2015).

We and others have reported that when assessed genome-wide, enhancer magnitude displays an asymmetrical distribution with distinct clusters of individual elements defining a subset of super enhancers (SEs) (Chapuy et al., 2013; Lovén et al., 2013). All cell types display asymmetrical patterns of enhancer magnitude, with up to 40% of enhancer-associated factors associated with SEs (Whyte et al., 2013). SE domains are defined using a rank-ordering algorithm that scores read density within regions of clustered peaks as measured by enhancer-bound factors or associated histone marks. When rank-ordered, SEs are discovered as the minority of clustered regions (typically ~3-5%) found to be asymmetrically larger than most 'typical' domains. SEs are enriched in disease-associated SNPs including those that predispose to cancer (Oldridge et al., 2015); maintain expression of lineage-defining transcriptional programs in stem cells (Adam et al., 2015); are sites of focal amplification and somatic alteration in cancer (Mansour et al., 2014; Zhang et al., 2015); and enforce expression of previously unrecognized tumor dependencies (Chapuy et al., 2013). Identification of SEs can also be used to define the TF core regulatory circuitry (CRC) of a given cell type (Saint-André et al., 2016). These core TFs self-regulate via inward binding to SEs within their own genic locus, while also regulating a coordinated set of TFs via SE binding, together forming an interconnected auto-regulatory loop. A well-established CRC exists in embryonic stem cells (ESC), where the core TFs essential for ESC identity OCT4, SOX2, and NANOG regulate themselves and each other through SEs (Boyer et al., 2005; Whyte et al., 2013). Identification of CRCs in tumors such as medulloblastoma can reveal clues pertaining to cell of origin and putative gene regulatory drivers of the oncogenic state (Lin et al., 2016). While targeting TFs directly with therapeutics remains a challenge, SEs can be hypersensitive to drugs that directly target enhancer-associated factors such as the bromodomain and extraterminal domain (BET) proteins and transcriptional kinases, offering a potential pharmacologic means to target cancer cell CRCs (Kwiatkowski et al., 2014; Lovén et al., 2013).

Here, we use a combination of genome-wide H3K27 acetylation (H3K27ac) and chromatin accessibility assays to define the global SE architecture and circuitry of the CLL epigenome. CLL is an incurable B-cell malignancy that is the most common hematologic malignancy in adults in the western hemisphere (Chiorazzi et al., 2005). CLL is clinically and molecularly heterogeneous, with 13q deletion the most recurrent abnormality (present in >50% of CLL patients). Other genetic and molecular events, including different recurrent copy number alterations (Brown et al., 2012; Edelmann et al., 2012; Grubor et al., 2009; Pfeifer et al., 2007), somatic coding and non-coding mutations (Landau et al., 2013; Landau et al., 2015; Puente et al., 2015; Wang et al., 2011), gene expression changes (Klein et al., 2001; Mansouri et al., 2012), and aberrant DNA methylation (Kulis et al., 2015; Landau et al., 2014; Oakes et al., 2016) tend to occur in less than 20% of patients each.

The genetic heterogeneity of CLL thus presents limited opportunities for mutation-targeted therapy. However recent development of therapeutic agents that target B cell signaling pathways, including CD20-directed antibodies and small-molecule antagonists of B cell receptor pathways (BTK inhibition with ibrutinib, PI3Kδ inhibition with idelalisib) reveal that effective treatments may be realized by targeting essential dependencies that define CLL cell state. Here, we have undertaken the complete characterization of the active *cis*-regulatory landscape of the CLL epigenome and explore SE analysis to enable the

identification of more universal drivers that underlie tumor cell specification, including CLL CRCs.

RESULTS

Histone acetylation and transposase hypersensitivity reveal active CLL enhancers

To map the enhancer landscape of the CLL genome, we employed ChIP-seq for the enhancer histone mark H3K27ac and assessed open chromatin regions by the assay for transposase accessible chromatin (ATAC-seq) (Buenrostro et al., 2013). We assembled a cohort of 23 primary B cell samples purified from the peripheral blood of treatment-naive individuals diagnosed with CLL. This cohort represented the most prevalent known disease subtypes, including immunoglobulin heavy chain (*IGHV*)-mutated and unmutated, and ZAP70-positive and negative disease (Table S1). Each sample was confirmed by flow cytometric analysis to be $CD19^{+}/CD5^{+}$ double-positive with a purity of >93%. We found that >99% of the total H3K27ac domains and ATAC hypersensitive sites (HSs) were discovered after measuring 16 samples (Figure S1A,B), indicating that this dataset comprehensively reflects the total enhancer elements of the CLL epigenome.

Assessment of enhancer magnitude genome-wide revealed characteristic, genome-wide enhancer asymmetry, with a minority of enhancer domains enriched with exceptionally high levels of H3K27ac (SEs) (Figure 1A and Table S2). These SE domains are found near genes known or surmised to play a prominent role specifically in CLL pathobiology and/or broadly across tumor types, including *CXCR4, CD74, PAX5, CD5, KRAS,* and *BCL2.* Large H3K27ac SE domains make up <10% of all total domains across the genome for each sample (Figure S1C). However, total H3K27ac enrichment at SEs represents on average 42% of total H3K27ac signal throughout the genome (Figure S1D), emphasizing the dominant nature these enhancer domains play in the transcriptional regulatory circuitry of CLL. At the *BCL2* locus, a gene commonly transcriptionally upregulated in CLL and targeted by the recently FDA-approved agent venetoclax (Roberts et al., 2015), we observed a broad H3K27ac SE domain with concomitant open chromatin at the promoter, throughout intronic regions, and 3’ to the gene (Figure 1B).

H3K27ac and ATAC-seq HSs are highly correlated, suggesting the bulk of accessible chromatin occurs in regions associated with transcriptionally active chromatin domains, with ATAC-seq read density occurring within localized regions of H3K27ac depletion (Figure S1E,F). These open chromatin regions within discrete H3K27ac domains of SEs reveals the *cis*-regulatory DNA available for TF binding within the SE and can be resolved to individual TFs through motif identification. The paired assays provide a high-resolution dataset for interrogating the active enhancer-based transcriptional architecture of CLL.

Interrogating the genomic location of CLL HSs, we found that a majority (>60%) occur in non-coding intronic or intergenic regions greater than 3 kb from transcription start sites (TSS) (Figure S1G,H), similar to prior observations made in normal primary human cell types and other studies of CLL open chromatin regions (Rendeiro et al., 2016). Further evidence that these HSs are indicative of sites of transcriptional activity, we observe that DNA within HSs is largely unmethylated in CLL patient samples, with greater than 90% of

CpGs within HSs displaying no cytosine methylation based on comparison to an annotated cohort of methylated CpGs observed in CLL (Landau et al., 2014) (Figure S1I). When enhancers are binned by decile, we observe significantly more ATAC-seq HSs in the largest H3K27ac domains, suggesting that these enhancer domains are associated with a greater number of *cis* operator elements (Figure S1J). When all CLL HSs from our cohort are compiled we find that approximately 30% are unique to a single sample, with ~60% of total peaks occurring in less than seven samples (Figures 1C and S1K).

Amidst the regulatory heterogeneity among all samples, we discovered a common set of HSs that occur in at least 13 samples (22% of all HSs discovered across cohort), suggesting a conserved core of CLL gene control determinants. The cumulative incidence profile of H3K27ac peaks is similar to that of ATAC HSs, with a majority of peaks observable in only a third or less of the samples (Figures 1C and S1L). This profile is slightly shifted however with coincident ATAC HSs/H3K27ac with almost half (49%) present in greater than 13 samples. When large enhancer domains such as SEs are assessed, we find that the degree of conservation among individual samples is greater than when total H3K27ac peaks are assessed with a majority (62.8% of SEs) present as an enhancer in at least 13 samples, including those at the loci of *KRAS, CD5, PAX5, CXCR4, BCL2,* and *CD74.* Together, these data suggest that these asymmetrically large domains dominate the conserved gene regulatory architecture of the CLL genome.

A distinct super enhancer signature in CLL

In order to discern a CLL-specific SE repertoire, we compared the CLL H3K27ac landscapes generated here with six H3K27ac enhancer landscapes derived from $CD19^+$ peripheral blood B cells harvested from normal donors (normal B cells, NBC). Three NBC samples were obtained as a part of this study, and three others have been previously generated by other groups (Bernstein et al., 2010; The FANTOM Consortium et al., 2014). Differential peak analysis revealed a total of 199 SEs enriched in the CLL cohort, while 230 SEs were NBC-enriched (Figure 2A). Proximal to NBC-specific enhancers included genes with evidence of lymphoma-suppressive functions such as *BACH2* and *BANK1* (Figure 2B). Enhancers whose signal was enriched in the CLL cohort included the *BCL2* SE and *CTLA4* (Figure 2C), the inhibitory T-cell checkpoint factor previously shown to be overexpressed in CLL cells (Kosmaczewska et al., 2005). Differential expression based on previously published mRNA-seq data comparing CLL with NBC cells is highly correlated with SE status: genes proximal to an SE 'lost' in CLL have lower mRNA expression, while those SEs 'gained' in CLL are associated with higher mRNA expression when compared to genes without significant changes in SE status (Figure 2D) (Landau et al., 2014). Thus, SE-localized transcription factor dysfunction is likely a contributing factor to the widespread gene expression changes observed in CLL. While we do not observe differentiation of genetically defined molecular *IGHV*-mutated (*IGHV*-M) and *IGHV*-unmutated (*IGHV*-U) CLL-subtypes based on these variable SEs, direct differential analysis of the CLL samples by *IGHV* status revealed several SEs specifically enriched in *IGHV*-M and *IGHV*-U, including at the *KLK2* locus (*IGHV*-U specific) and the *ZBTB20* locus (*IGHV*-M specific) (Figure S2), recently shown to display *IGHV*-M specific open chromatin (Rendeiro et al., 2016).

Transcription factor core regulatory circuitry of CLL

With the goal of understanding the critical TF interactions responsible for maintaining the CLL transcriptional program, we modeled transcriptional regulatory networks mediated by asymmetrically large enhancer domains. We utilized H3K27ac domains to define enhancer regions and ATAC-seq HSs to demarcate discrete *cis*-regulatory elements within these domains (Figure 3A) (Federation et al., 2018). From top-ranked enhancers, a subset was found in proximity to TF gene loci with acetylated promoters defined by ChIP-seq and these TFs defined the set of candidate nodes within the network. We confirmed physical TF gene promoter association with proximal SE by 4C-seq in the CLL MEC1 cell line at two TF loci, *IRF8* and *LEF1* (Figure S3A,B). Edges between nodes were established when a factor was found to have its DNA-binding sequence within an accessible regulatory region of another candidate TF. The total in-degree of a node is the sum of the number of TFs binding in its regulatory sequence, while the out-degree of a TF is defined as the number of other nodes it is regulating in the network. Importantly, using ATAC-defined HSs to demarcate TF binding regions dramatically reduces the search space for TF *cis*-regulatory sites when compared to previously used network modeling methods that rely on nucleosome-depleted regions ("valleys") to define these elements (Ziller et al., 2015), with substantial enrichment in known co-regulating TFs (Figures S3C,D). Network edges were compiled into a complete transcriptional network, with TFs being assigned to auto-regulatory cliques that in turn co-regulate an extended network or enhancer-associated genes. Scatterplots generated with calculated in-degree and out-degree for each sample reveal the most interconnected transcription factors (Figure 3B and S3E).

Among the most highly connected CLL TFs include known B cell lineage-defining TFs such as several IRF family members and PAX5, which displays a broad SE across the 5' end of its gene locus in both CLL and NBC (Figure S4). Other highly-connected TFs include FOXP1, RARA, ETS1, IRF2, and IRF8. From this analysis, we also discern enhancer-mediated auto-regulatory cliques for each sample (Figure 3C and Table S3). Each clique can be scored based on the total connectivity of its constitutive TFs, and include master B cell TFs such as PAX5, which in turn regulate genes associated with CLL pathobiology such as *BCL2, CXCR4,* and *CD83.* We calculated a clique enrichment score (the percentage of total cliques in which a TF is a constituent member) for each TF in each sample. Clustering samples based on TF clique enrichment scores separated CLL from NBC samples, revealing intrinsic CRC differences between cell types (Figure 3D). These differences are driven in part by hyper-connectivity of PAX5 in CLL, which displays a higher average clique enrichment in the CLL cohort compared to NBC samples.

Identification of essential components of the CLL CRC

To mechanistically assess CLL cell dependency on individual TFs, we utilized four human cell lines derived from CLL patients (MEC1, MEC2, OSU-CLL, CII). These CLL cell lines are derived from primary CLL samples and transformed by Epstein-Barr virus infection, which likely has a significant effect on transcription factor activity. When we compare TF clique enrichment scores between a combined meta-analysis of the primary CLL networks and each individual CLL cell line, several TFs including ETV6, IRF2, and PAX5 displayed comparable enrichments between cell lines and primary samples (Figure 4A). Some TFs,

including FOXP1 were highly enriched in primary CLL cliques yet were not present in cell line networks. Conversely *MYC,* a gene known to be deregulated upon EBV infection of B cells, displayed high enrichment specifically in cell line circuits (Lacy et al., 1987).

We then used a pooled CRISPR screening approach to systematically knock out CLL CRC TFs in the MEC1 and OSU-CLL cell lines (Figure S5A). Cells were stably transduced with Cas9, and a library of ~3500 gRNAs was transduced covering a set of 147 core CLL TFs. Notably, of the 147 TFs assessed, only a minority were significantly depleted in the screens (Figure 4B,C and Table S4). In MEC1 cells, targeting 34% of TFs significantly reduced cell proliferation over the course of the screen (p value < 0.05); in OSU-CLL cells, 46% reduced proliferation. Among CRC constituent TFs, *PAX5* knockout resulted in the most significant effect on cell proliferation in both cell lines (Figure 4C), revealing that this factor – essential for B cell development and differentiation – is the most essential and recurrent regulatory node within CLL TF networks. Notably, three other TFs that showed a high degree of essentiality in both cell lines - BATF, EBF1, and IRF4 - show little to no H3K27ac in primary CLL samples (Figure S5B).

In an orthogonal assay system in which gRNA vectors co-expressing GFP are transduced into Cas9-expressing CLL cell lines and GFP is tracked over a period of 30 days, gRNA expression that target Cas9 to the coding regions of *PAX5* depleted GFP-positive cells in a time-dependent manner (Figure 4D). As negative controls, we used a non-cut-inducing gRNA that targets the luciferase gene and two gRNAs that cut within non-coding regions adjacent to the locus (to control for potential gene copy-number-associated genome-cutting effects), and a positive control gRNA targeting the essential *RPS19* gene. Five *PAX5*-targeting gRNAs displayed similar dropout-out dynamics across 3 CLL cell lines tested, while negative control gRNAs had no effect. Four out of the five *PAX5*-targeted gRNAs reduced cell growth with remarkable consistency, typically with similar dynamics to the positive control *RPS19* essential gene knockout. On-target disruption of the *PAX5* locus with selected gRNAs was confirmed through the Tracking of Indels by Decomposition (TIDE) analysis and immunoblotting following select gRNA transduction and sorting for GFP-positive cells (Figure S5C,D). Notably, we observed a high correlation between results obtained from the pooled screen and GFP-depletion assays for a panel of CLL TFs, confirming a competitive growth disadvantage of CLL cells with *PAX5, MYC, IKZF1, IKZF3,* and *RARA* disruption, while CRISPR reagents targeting *NFATC1, NFATC2, PPARA, LEF1,* and *IRF8* had relatively little effect (Figure S5E).

We compared these results to five gRNAs targeting the genes of established therapeutic protein targets in CLL: *BCL2, BTK,* and *PI3KCD* (Figure S5F). These gRNAs target either the 5’ end of the genes, or their functional domains targeted by their respective drug inhibitors. While selective depletion of most gRNA-expressing cells was observed in the MEC1 cell line, the dynamics of drop-out was shallower, highlighting the unique growth disadvantage conferred by PAX5 loss in these cells.

BET bromodomain inhibition leads to disruption of CLL CRC expression

In a number of contexts, expression of genes driven by SEs can be perturbed to a selective degree by BET bromodomain inhibitors (BETi) (Gröschel et al., 2014; Lovén et al., 2013).

Thus, while no pharmacologic agent exists to specifically target PAX5 or most other members of the CLL CRCs, we explored the possibility that BETi might decrease CRC gene expression through enhancer disruption. BETi have apparent antiproliferative activity in preclinical models of several lymphoid malignances (Delmore et al., 2011; Ott et al., 2012). These studies and others have spurred ongoing clinical trials of this therapeutic class in lymphoid malignancies (NCT01713582, NCT01949883, NCT02308761).

To understand the transcriptional consequences of BETi in CLL cells, we treated four cell lines for six hours with the prototype BETi JQ1 and harvested mRNA for complete quantitative transcriptomic studies. In each cell line tested (MEC1, MEC2, OSU-CLL, CII), JQ1 treatment led to a significant decrease in SE-driven gene expression when compared to genes associated with typical enhancers (TE) (Figures 5A and S6A). This included a selective decrease of TF genes constituent within the CRC (Figure S6B). When JQ1 effects on CRC TFs are parsed further, we observe that effects on CRC mRNA levels cluster into TFs that are more highly affected and those that are relatively less affected. Among the selectively perturbed CRC members are TFs shown by our CRISPR screening to be essential for competitive cell proliferation, including *PAX5, MYC, IKZF1,* and *RARA* (Figure 5B). Downregulation of IKZF1, PAX5, and c-MYC protein levels with short-term JQ1 treatment was confirmed with immunoassays (Figures S6C,D,E). We also performed mRNA-seq on primary CLL cells and NBCs and found that global effects on the transcriptome with JQ1 treatment are highly correlated between the two cell types (Figure 5C). This included significant down regulation of CRC TF genes *PAX5* and *IKZF1.* Among all TFs expressed in the primary CLL samples, *PAX5* and *IKZF1* are among the most downregulated genes with short-term JQ1 treatment (Figure 5D).

BET bromodomain inhibition is an effective strategy to target CLL cell growth

To further explore whether BETi may be a therapeutic option for CLL, we treated a panel of available CLL cell lines with a broad dose range of JQ1 and other BETi. For comparative analysis, we included other agents used for targeted therapy in CLL (venetoclax, ibrutinib, idelalisib), and the chemotherapeutic agent fludarabine. Treatment effects of these agents clustered by target class, with all BET inhibitors clustering together (Figure 6A). BET bromodomain inhibitors were broadly more effective as antiproliferative agents across the cell lines tested when compared to the other drugs (Figure 6A,B).

JQ1 can selectively inhibit the chromatin association of BET family members, which include BRD2, BRD3, BRD4, and the testis-specific protein BRDT. As JQ1 treatment alone cannot independently assess differential dependency of CLL cells to BET protein activity, we used CRISPR targeting of each family member expressed to assess which BET factors may be responsible for promoting CLL cell proliferation in culture. We found that gRNAs targeting the *BRD3* locus had relatively little effect on cell proliferation; gRNAs targeting the *BRD2* locus generally had a moderate effect; and most gRNAs targeting *BRD4* had the most pronounced effect (Figure 6C).

Significant effects on viability were also observed with JQ1 treatment of four primary CLL patient samples, with dose-dependent effects at both 24 and 48 hours (Figure 7A). JQ1 treatment also resulted in an increase in apoptotic cells in each of the four primary CLL

samples (Figure 7B and S7A,B), with similar effects observed in NBC (Figure S7C). We then used a modified *in vivo* model of CLL with a luciferase-expressing version of the MEC1 cell line engrafted into *Rag2*$^{-/-}\gamma_c^{-/-}$ mice by tail vein injection (Bertilaccio et al., 2010). This model closely recapitulates an aggressive form of disseminated CLL with rapid development of systemic disease. Daily treatment of mice with JQ1 (50 mg/kg by intraperitoneal injection) led to a significant decrease in tumor burden as measured with serial assessment of whole-body bioluminescence (Figure 7C,D). Importantly, JQ1 also led to an increase in overall survival compared to vehicle treated mice ($p = 0.0005$) (Figure 7E). Together, these data establish BET inhibition as promising for further exploration in translational model systems of CLL and potential clinical investigation.

DISCUSSION

Enhancers have long been recognized as facilitating augmented expression of genes in cancer, particularly in hematopoietic malignancies. Here we describe efforts to map and dissect enhancer loci in CLL, an incurable B cell disorder typified by heterogeneous genetic and DNA methylation profiles. Whole-exome sequencing studies have identified recurrent somatic mutations in coding regions of a number of genes including *TP53, SF3B1, NOTCH1,* and *MYD88,* as well as significant inter- and intra-tumoral heterogeneity (Landau et al., 2013; 2015). Whole-genome analysis of 150 CLL samples revealed disrupted gene regulatory sites in distal enhancers of the *PAX5* locus (Puente et al., 2015), and a very recent study from the same group profiled the epigenomic features of a large CLL cohort and panel of normal B cell subtypes, revealing dysregulated enhancers in CLL subtypes including *IGHV*-M and *IGHV*-U cases (Beekman et al., 2018). Another cohort of CLL samples has also recently been assessed by ATAC-seq, revealing subtype-specific open chromatin signatures of CLL (Rendeiro et al., 2016). By using complementary assay technologies – chromatin accessibility as a biophysical readout of the CLL epigenome, and H3K27ac histone immunoprecipitation as a biochemical signifier of enhancer activity – we have captured the active regulatory landscape of the CLL epigenome. Individually these assays are capable of identifying regulatory element domains. Analysis methods described here specifically report on the size distribution of these domains, systematically identifying SE and more typical enhancer domains, and use them to define the densely interconnected transcriptional networks of CLL. Uniquely, the combination of both assays can identify the precise TF-bound hypersensitive elements within large enhancer domains. This enables precise identification of enhancer-defined TF networks active in CLL. We do note that unlike genomic sequencing and DNA methylation analysis, the epigenomic assays included here do not resolve the intratumoral heterogeneity known to exist in CLL. However, emerging single cell epigenomic assays such as single cell ATAC-seq are quickly maturing, and may soon be used to appraise regulatory network heterogeneity within CLL and during treatment (Buenrostro et al., 2018).

This systematic identification of active CLL regulatory elements has led to several enabling observations. By constructing the CLL-specific TF regulatory network using enhancer profiles, we are able to identify the critical TF nodes that enforce the CLL epigenome. This, coupled with systematic knockout screening using robust CRISPR/Cas9 technology, enabled us to discern the critical, essential nodes of the CLL enhancer architecture. The PAX5-

mediated transcription circuit emerged as a singularly essential node in CLL. PAX5 is known to be essential for B cell development, and has been implicated in promoting lymphomagenesis by directly activating components of the B cell receptor signaling pathway (Cozma et al., 2007). Expression of *PAX5* along with other essential TFs can be perturbed with the BET bromodomain inhibitor JQ1. Treatment with JQ1 led to dramatic inhibition of cell proliferation and proved efficacious in xenografted mice with disseminated CLL tumors. Recently, a structurally dissimilar BET inhibitor PLX51107 was reported to have similar effects in pre-clinical models of CLL, with transcriptional disruption of B cell receptor signaling genes in proximity to BRD4-bound super enhancers including *PAX5* (Ozer et al., 2018). Thus, this class of pharmacologic agent may prove useful for CLL therapy, perhaps in combination with other agents targeting B cell signaling.

In conclusion, we demonstrate that *cis*-regulatory element discovery in primary cancer cells is a useful tool for uncovering tumor-specific hallmarks and active TF regulatory pathways. This analysis, coupled with large-scale functional genetic experiments, resolves the critical factors necessary for transcriptional networks. This type of epigenomic analysis complements other more well-established analyses of cancers like whole-exome and whole-genome sequencing and captures information about the essential active regulatory networks inaccessible by genetics-based approaches alone. We anticipate that these analyses will provide a framework for describing the epigenetic state of primary tumor samples and will be useful in identifying TF and other dependencies representing potential therapeutic targets.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Christopher J. Ott at christopher.ott@mgh.harvard.edu.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Primary samples—Peripheral blood samples from CLL patients were collected and viably frozen as part of an IRB-approved tissue banking protocol at the Dana-Farber Cancer Institute and all patients signed written informed consent prior to having the sample drawn. For samples with white blood cell count <25 K or absolute lymphocyte count <20 K, B cells were purified using the Easy Sep Human B cell Enrichment Kit (StemCell Technologies Inc.) according to the manufacturer's instructions. Following thawing, sample purity was confirmed by flow cytometry monitoring CD5/CD19 expression on a BD FACSAria II (antibodies: BV421 mouse anti-human CD5, #562646 and HIB19 APC mouse anti-human CD19, #555415, both from BD Pharmingen). Gender, age of diagnosis, and relevant genetic/immunophenotypic features of each sample are included in Supplemental Table 1. Normal B cell samples were harvested and purified from discarded material donated by normal healthy adults through the Kraft Family Blood Donor Center at the Dana-Farber Cancer Institute and Brigham and Women’s Hospital. Ten milliliters of peripheral blood was purified by Ficoll gradient, and B cells were enriched in the sample with a negative selection cocktail (RosetteSep, StemCell Technologies #15024).

Cell lines—MEC1, MEC2, and HG3 cells were obtained from DSMZ and cultured in RPMI with 10% FBS (Rosén et al., 2012; Stacchini et al., 1999). OSU-CLL cells were a kind gift from A. Johnson (The Ohio State University Cancer Center, Columbus) and cultured in RPMI with 10% FBS (Hertlein et al., 2013). CII cells were a kind gift of M. Kwok (Birmingham, UK) and A. Rosen (Linkoping, Sweden) and cultured in RPMI with 10% FBS (Karande et al., 1980). PCL12 cells were a kind gift from E. Ten Hacken (DFCI, Boston) and C. Scielzo (UniSR, Milan) and cultured in RPMI with 10% FBS (Agathangelidis et al., 2015). MDA-BM5 cells were a kind gift from E. Ten Hacken (DFCI, Boston) and I. McNiece (MD Anderson, Houston) and cultured in MEM Alpha with 20% FBS (Kellner et al., 2016). All cell lines were cultured at 37°C with 5% CO_2.

Animal studies—All animal studies were performed according to Dana-Farber Cancer Institute Institutional Animal Care and Use Committee-approved protocols. The CLL cell line MEC1 was engineered for *in vivo* imaging by transduction with VSV-G-pseudotyped lentivirus encoding firefly luciferase, mCherry, and puromycin-N-acetyltransferase. Five million cells were injected into 6-8 week old female $Rag2^{-/-}\gamma_c^{-/-}$ mice (The Jackson Laboratory, C;129S4-$Rag2^{tm1.1Flv}Il2rg^{tm1.1Flv}$/J) via the lateral tail vein. Disease burden was quantified using bioluminescent imaging. Eleven days following MEC1 cell injection, mice with established disease were randomly divided into two cohorts with equal mean bioluminescence and treated once daily with 50 mg/kg JQ1 or vehicle (10% (2-hydroxypropyl)-β-cyclodextrin (Sigma H107) in 0.9% sterile saline) via intraperitoneal injection. Mice were treated until presentation of moribund features, as determined when mice were observed with paralysis, emaciation or dyspnea at which point mice were immediately euthanized with CO_2.

METHOD DETAILS

ATAC-seq—ATAC-seq was performed with 50,000 viable cells as described with minor modifications (Buenrostro et al., 2013). Transposition reactions were performed for 1 hr at 37° C, followed by purification and sample barcoding by PCR. Samples were sequenced on an Illumina HiSeq 2000 or 2500 in paired-end mode with 100 × 100 cycles. ATAC-seq for MEC1, MEC2, OSU-CLL, and CII cells was performed as above after cells were grown in RPMI-1640 supplemented with 10% FCS. MEC1, MEC2, OSU-CLL and CII ATAC-seq libraries were sequenced on an Illumina NextSeq in paired-end mode with 75 × 75 cycles. Raw sequencing data was mapped to the hg19 build of the human genome with Bowtie2 v2.3.0 with default settings and the parameters –p 4 –k 1 (Langmead and Salzberg, 2012). Mapped reads were filtered to remove duplicate reads, those reads mapping to mitochondrial DNA and to regions in the ENCODE blacklist. Reads mapping to the forward strand were shifted forward 4 bp and reads mapping to the reverse strand were shifted backwards 5 bp. ZINBA v2.03.1 was then used to find peaks with extension=200, winSize=300 and offset=75 and default settings (Rashid et al., 2011).

Chromatin immunoprecipitation—Chromatin immunoprecipitation was performed with 3 million primary CLL cells per sample using an anti-H3K27ac specific antibody (Abcam #ab4729). ChIP procedure was performed as previously described (Chapuy et al., 2013). Samples were sequenced on an Illumina HiSeq 2000 or 2500, paired-end, 100 × 100

cycles. Cell lines were sequenced on an Illumina NextSeq in single-end mode 75 cycles. Raw sequencing data was mapped to the hg19 build of the human genome with Bowtie2 v2.3.0 with default settings and the parameters –p 4 –k 1. Mapped reads were filtered to remove duplicate reads, and to regions in the ENCODE blacklist. MACS v1.4 was used for peak identification with a p value cutoff of 1e-6 (Zhang et al., 2008). Comparisons of genomic loci were performed with BEDOPS v2.4.21 and mapping reads to genomic loci was performed with Bamliquidator v1.2.0 (Lin et al., 2016; Neph et al., 2012). Enhancer identification was carried out using ROSE2 (https://github.com/BradnerLab/pipeline) (Lin et al., 2016).

Super enhancer analysis—Super enhancers were defined using the Rank Ordering of Super Enhancers (ROSE2) algorithm (Lin et al., 2016). For all samples, the stitching distance was fixed at 12.5 kb to facilitate comparisons between samples. All other parameters used the default setting. Gene targets from the ROSE2 ENHANCER_TO_TOP_GENE.txt file was used for defining the target genes of super enhancers for subsequent analyses.

Characterization of genomic loci—Comparisons of genomic loci were performed with BEDOPS v2.4.21 (Neph et al., 2012). BED files containing ATAC-seq or ChIP-seq peaks for an indivudal sample were sorted with the sort-bed function. For downstream analyses, peaks were merged by disease state (CLL or NBC) using the bedops –merge function. Peak distribution analysis was performed with CEAS (Shin et al., 2009). The gene annotation sqlite3 file is provided by the Liu Lab in the software manual. The BED files used were the merged BEDs described above. Promoters were defined as those falling within 1kb of the annotated transcriptional start site. Additional annotations (intron, exon, UTR) are defined in the gene annotation file.

Sample saturation analysis—BED files for the peaks of interest in each sample were used, and the order by which to aggregate the peaks was randomly permutated. The number of peaks present in the first sample was counted, then merged with those in the second sample and counted again. This merged set was then combined with the third sample and counted, and this algirothm was continued until all samples were merged. Ten different permutations were considered and standard deviations of the number of peaks present after each iteration were plotted.

Differential enhancer analysis—To identify variable SE domains enriched in either CLL or NBC we defined the union of all SEs discovered across the CLL and NBC cohorts, and used DESeq2 which can robustly estimate within-group variance in sample cohorts controlling for GC content of analyzed regions and technical variability in sample read depth within SEs (Love et al., 2014). Reads in all SE regions were quantified with the Rsubread Bioconductor R package using paired-end mode (Liao et al., 2013). Multi-overlapping read assignment was resolved by the length of overlap, and overlapping regions were resolved using chromVAR 'filterPeaks' function, removing all but the highest count region for each set of overlapping regions (Schep et al., 2017). GC content for each region was calculated using the Biostrings Bioconductor package (Pagès et al., 2018). EDASeq was used to

generate offsets first with 'EDASeq::withinLaneNormalization' and subsequently with 'EDASeq::betweenLaneNormalization' (Risso et al., 2011). Normalization factors were transformed such that geometric means of each row of the offset matrix are equal to 1. DESeq2 was then run with the formula '~ CONDITION', where CONDITION has levels 'CLL' and 'NBC'. Results for CLL versus NBC were extracted and log fold changes were shrunk using the 'apeglm' package. Differential SEs were defined as those with a Benjamini-Hochberg adjusted p value < 0.1. Heatmap was generated using the R package iheatmapr, with clustering performed by euclidean distance calculated with the base r stats::dist function (R Core Team, 2013; Schep and Kummerfeld, 2017).

Transcriptional regulatory networks—Network construction utilized the coltron python package (https://pypi.python.org/pypi/coltron), which is adapted from Saint-André et al. to include information from ATAC-seq and across multiple samples (Saint-André et al., 2016, Federation et al., 2018). Scripts used to perform network analysis is available at github.com/BradnerLab/CLL_TFnetworks_2018. To define the set of TF nodes for constructing the transcriptional regulatory network of each sample using paired ATAC-seq and H3K27ac ChIP-seq datasets, we assigned the top 1000 enhancers in the H3K27ac enhancer rank-ordered list to nearby highly acetylated gene promoters using the following strategy: Promoters overlapping the stitched enhancer region are given priority for assignment. If that condition is not met, then all genes with promoters overlapping an H3K27ac peak contained within 100 kb of the enhancer are assigned. If neither of those conditions find a suitable gene assignment, the enhancer is assigned to the closest promoter overlapping a peak of H3K27ac ChIP-seq.

Then, discrete ATAC-seq peaks within the large H3K27ac enhancer domains were identified. The underlying sequence from these peaks was extracted and FIMO v4.91 was used to search for binding sites of node TFs in these peaks with a p value of < 1e-4. TF position-weight matrices were taken from Transfac and Jaspar (Grant et al., 2011; Mathelier et al., 2013; Matys et al., 2006). This motif file is included in the distribution of the coltron package. When a motif of gene A was identified in an enhancer of gene B, an edge was drawn in the network between gene A and the gene B. To build a representative network of all primary CLL samples, the network construction algorithm described above was performed on merged enhancers from all primary CLL samples. Rank-ordering of enhancers for the consensus network was performed by summing the read density in each enhancer across all samples. The algorithm was run with 13 individual CLL samples (CLL-1 - CLL-13) and 3 NBC samples for which paired ATAC-seq and H3K27ac ChIP-seq datasets were generated, along with four CLL cell lines (MEC1, MEC2, OSU-CLL, CII).

In all networks, in-degree for TF_x is defined as the number of nodes with motifs found in any enhancer assigned to regulate TF_x. Out-degree for TF_x is defined as the number of nodes in the network which is regulated by an enhancer containing a motif for TF_x.

CRCs for each sample were defined as the complete auto-regulatory TF network (clique) within each sample that comprised TFs with the highest clique score. After the complete construction of the network, a clique is defined as a subnetwork with a size of at least 4 nodes where all nodes are connected to themselves and all other nodes within thet clique.

For each sample, the set of all cliques of size 4 or greater were defined. 'Clique enrichment' is defined for a specific TF as the percentage of all observed cliques within a sample in which that TF participates. 'Clique score', used to pick a top-ranked CRC, is defined as the average clique enrichment across an entire clique.

To generate extended networks, the same algorithm to identify TF binding within enhancers was utilized and applied to all of the top 1000 enhancers. ATAC-seq peaks within enhancers were identified, the underlying DNA sequence was extracted, then this sequence was searched for TF motifs for all node TFs within the network. These enhancers were then assigned using the TOP_GENE output from the ROSE2 software, which identifies the closest gene with an H3K27ac peak overlapping the promoter.

All heatmaps and clustering was performed using Morpheus software (https://software.broadinstitute.org/morpheus). To define consensus SEs across CLL samples, we adapted the ROSE2 algorithm used to define enhancer magnitude from ChIP-seq datasets. H3K27ac-enriched peaks across all samples were stitched, followed by remapping of the read density to stitched regions in each sample. Each sample was internally normalized, then the regions were averaged and ranked, producing the characteristic asymmetric distribution of H3K27ac domains. Using this method, we observe 571 consensus SEs in the CLL epigenome representing 3.7% of total enhancer domains (n = 15,310).

4C-seq—MEC1 cells (~10 million) were harvested from cell culture media, washed 2X in cold PBS, and crosslinked with 1% formaldehyde at room temperature for 10 min. Libraries for 4C sequencing were prepared as described (Splinter et al., 2012). Each 4C library uses a combination of two enzymes; a primary restriction enzyme and a secondary restriction enzyme. For this study we used NlaIII/primary (NEB #R0125L) – DpnII/secondary (NEB #R0543M). In brief, nuclei were isolated, and the crosslinked-chromatin was digested overnight (O/N) at 37°C with the primary restriction enzyme. Following this step, a ligation reaction with T4 DNA ligase (NEB #B0202S) was performed O/N, 16°C. Next, samples were incubated at 65°C to de-crosslink the chromatin. Next, the secondary restriction enzyme was used (O/N, 37°C). This step was followed by a ligation reaction using T4 DNA ligase (O/N, 16°C). For each library, a linear range of amplification was determined by performing PCR using a titration of four different DNA template amounts (100 ng, 50 ng, 25 ng and 12.5 ng). A 25 μl PCR (final volume) consisted of (all reagents listed are at final concentration): 1.75 U Expand Long Template Polymerase (Roche, #11759060001), 1X buffer 1 (supplied with Expand Long Template Polymerase), 30 picomoles of each primer (forward and reverse), 5 nanomoles of each dNTP and Milli-Q water to a total volume of 25 μl. PCR conditions were 2 min 94°C, 10 sec 94°C, 1 min 55°C, 3 min 68°C, 29 times repeat, 5 min 68°C and ∞ 4°C. The primer sequences are included in Supplemental Table 5. Once a linear range of amplification was determined, 16 independent reactions each using 200 ng of template was subjected to PCR (final volume of 50 μl). The final concentration for the 50 μl PCR reagents and PCR conditions were the same as described above. Next, all reactions were pooled and purified (High Pure PCR Product Purification Kit, Roche, #11732676001). For library preparation, 100 ng of the purified PCR product was used as input for the Rubicon ThruPLEX DNA-seq Kit, #R400407). All samples were barcoded and run on the Illumina NextSeq platform. We aimed to obtain 5 million reads per library. The 4C-seq data

was analyzed following the standard pipeline developed by Tanay and de Laat et al using the default parameters (van de Werken et al., 2012). The sequencing reads were first mapped to a set of reference sequences generated based on the primary and secondary restriction sites. Then, the reads that mapped to a certain locus were filtered and normalized to estimate a coverage profile, showing the enrichment of 4C reads on the genome with regard to the bait region.

Pooled CRISPR screening—Lentivirus was produced in 293FT cells by co-transfection of pMD2.G (Addgene #12259), psPAX2 (Addgene #12260), with spCas9-expressing plasmid (Addgene #73310). Viral supernatants were collected 60 hr after transfection, filtered through a 0.22 μm membrane and concentrated 20-fold with Lenti-X Concentrator. All cells lines were transduced by spinoculation at 2500 rpm for 1.5 hr at room temperature, and then incubated for 72 hr at 37 degrees/5% CO2. Cells were selected with blasticidin (10 ug/mL) until resistant, spCas9-expressing cells grew out. Cas9 expression was confirmed by immunoblotting. Cells were maintained in 10 ug/mL blasticidin throughout the screen. Guide RNA libraries were designed by manually curating optimized sgRNA sequences using the Broad Institute GPP CRIPSRko tool (http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design, (Doench et al., 2016); 20-40 gRNAs were designed to either knockout protein expression, or to create indels in functional domains such as DNA-binding domains. Transcription factor genes targeted in the library were chosen as the union of TFs that appear part of the enhancer-based core regulatory circuitry from the primary CLL analysis, MEC1 cells, MEC2 cells, and OSU-CLL cells combined. sgRNA sequences were purchased in pooled format (CustomArray, Inc), cloned by Gibson assembly in the LRG vector (Addgene #65656), and transformed into electrocompetent bacterial cells (Endura #60242-1, Lucigen). Pooled lentivirus was produced from the cloned library, and cells were transduced at approximately 5% efficiency (10 million cells total, >2800 cells per sgRNA). Cell numbers were maintained throughout passaging to maintain complexity, and cells were harvested 28 days post-transduction. Genomic DNA was extracted from samples with the QIAamp DNA Blood Maxi Kit (Qiagen, #51194), sgRNA sequences were amplified by two-step PCR (extraction and subsequent barcoding) as described (Shalem et al., 2014). Samples were sequenced on an Illumina NextSeq 500 with single-end 75 base pair reads. Gene-level analysis was performed by RSA normalizing to plasmid DNA reads obtained by sequencing the untransduced library (König et al., 2007).

Competitive CRISPR GFP depletion assays—Cells were engineered to express Cas9 by lentiviral transduction as described above. Co-expression of sgRNA and GFP was achieved by lentiviral transduction of the LRG plasmid cloned with individual sgRNA sequences (all sequences listed in Supplemental Table 5). Lentiviral transduction was performed in 12-well plates, one million cells per well, with spinoculation as described above. Media was replenished every two days, and flow cytometry to determine GFP^+ percentage was performed at indicated time points in triplicate in 96-well plates (Guava Easycyte HT, Millipore).

mRNA-seq—Cell lines MEC-1, MEC-2, OSU-CLL, and CII were seeded into 6-well plates at a density of 800,000 cells/mL, 2 mL per well. Twenty-four hr following seeding,

cells were treated with either vehicle (DMSO, 0.1% final concentration) or JQ1 (1 μM) in triplicate for a total of 6 hr. For primary cell assays, cells were seeded onto plates coated with IgM solution (10 μL/mL diluted in PBS; Jackson ImmunoResearch 109-006-129) and treated with JQ1 one hour after seeding. Cells were harvested following 6 hr of treament, RNA was isolated using miRNeasy RNA isolation kit (Qiagen), and ERCC spike-in controls (Invitrogen) were added in order to normalize expression values to input cell number per the manufacturer's protocol and as previously described (Loven et al., 2013). Library preparation was performed with the TruSeq Stranded mRNA kit (Illumina), and libraries were sequenced on a NextSeq 500 in paired-end mode, 35 × 35 cycles. Reads were aligned to hg19 using HISAT with default parameters, and transcript abundance was assessed using *cuffquant* and *cuffnorm* using default parameters (Kim et al., 2015; Trapnell et al., 2010). A custom R script (normalizeRNAseq.R) was used to assess fold change statistics and normalize to spike-in controls (github.com/BradnerLab/pipeline).

Immunoassays—For analysis of PAX5 levels following CRISPR gRNA expression, MEC1 cells stably expressing Cas9 were seeded at a density of 1×10^6 cells per mL in a total volume of 2 mL in a 12-well plate. 50 μL of concentrated lentivirus carrying gRNA-expressing LRG vector. After spinoculation, cells were expanded in culture for seven days then flow-sorted for GFP^+ cells. Cells (1×10^5) were used TIDE analysis (see below) and the remainder were lysed in RIPA buffer; lysates were run on 4-12% gradient SDS-PAGE gels for electrophoresis. Immunoblotting was performed with the following antibodies: PAX5 (Abcam, ab183575); alpha-tubulin (Sigma Aldrich, T5168). Blots were incubated with species-specific fluorophore-conjugated secondary antibodies and imaged on a LICOR Odyssey Clx imaging system. For analysis of IKZF1, c-MYC, and PAX5 levels following JQ1 treatment, CII cells were seeded onto 6-well plates at a density of 8×10^5. Twenty-four hr following seeding, cells were treated with either DMSO vehicle or 1 μM JQ1. Twenty-four hr following treatment, cells were harvested and lysed in RIPA buffer. Cell lysates were run on a ProteinSimple Wes capillary electrophoresis instrument using instrument default settings and manufacturer's standard protocol. Immunoassay was performed with the following primary antibodies: IKZF1 (Bethyl A303-516A); c-MYC (Santa Cruz N262); PAX5 (RD Systems MAB3487); β-actin (RD Systems MAB8929). Chemiluminescence was quantified by assessing the area under the curve, and total protein levels were normalized to β-actin controls.

TIDE assays—Genomic DNA from sorted PAX5 gRNA (gRNA-2 and Locus control gRNA-2)/Cas9-expressing MEC1 cells was harvested with DNeasy Blood & Tissue kit (Qiagen). PCR reactions were performed with primers listed in Supplemental Table 5. The forward primers were used to sequence each PCR amplicon, and TIDE analysis was performed using available software (tide.nki.nl) (Brinkman et al., 2014).

Cell viability assays—For cell line assays, cells were plated in 384-well plates at a seeding density of 5,000 cells/mL in RPMI media supplemented with 10% FBS. One day after seeding, cells were treated with compounds resuspended as DMSO stock solutions via pin-transfer robots (Janus Workstation, Perkin Elmer). Four days following treatment, 1/10 volume of AlamarBlue Cell Viability Reagent (Life Technologies) was added and incubated

with live cells for 24 hr. After incubation, fluorescence was read on an Envision multi-well plate reader (PerkinElmer). Compounds were obtained from commercial sources: venetoclax (Selleckchem, #S8048), fludarabine (Selleckchem, S1229), ibrutinib (Selleckchem, #S2680), idelalisib (Selleckchem, S2226). JQ1 used in assays was synthesized as previously described (Filippakopoulos et al., 2010). For primary cell assays, 96-well plates were coated with IgM solution (10 μL/mL diluted in PBS; Jackson ImmunoResearch 109-006-129) and primary CLL cells were seeded at 100,000 cells per well. Cell Titer Glo reagent (Promega) was added per the manufacturer's protocol at time points 0, 24, and 48 hr following treatment of either vehicle (DMSO, 0.1% final concentration) or JQ1.

Apoptosis assay—Cellular apoptosis was assessed with PE Annexin V apoptosis detection kit I by BD Pharmingen (#559763) after 48 hr of treatment with 1 μM JQ1 or DMSO. Briefly, cells were washed twice with cold PBS and incubated with 100 μL of binding buffer containing 5 μL of PE Annexin V and 5 μL of 7-AAD for 15 minutes at room temperature in the dark. Then 400 μL of binding buffer was added and flow cytometry was performed using BD LSR Fortessa instrument. The percentage of cells positive for PE-Annexin V and negative for 7-AAD was determined to be the apoptotic fraction.

QUANTIFICATION AND STATISTICAL ANALYSIS

T-tests were applied in Figures 2D, 3D, 5A, 7D, Suppl. Figures 2F, 13.

P values were generated for CRISPR screening results using 'redundant siRNA activity' (RSA) analysis (König et al., 2007) in Figure 4B,C.

Heatmaps clustered by euclidean distance were generated in Figures 2A, 3D, 5B, 6A. Software used and specifics of methods and cut-offs for each analysis are included in the Figure legends.

DATA SOFTWARE AND ACCESSIBILITY

All datasets (ChIP-seq, ATAC-seq, mRNA-seq) generated from cell lines have been deposited into the Gene Expression Omnibus (GEO) repository (accession #GSE119744). All sequencing data (ChIP-seq, ATAC-seq, mRNA-seq) generated from primary human subject samples have been deposited into a NCI dbGAP repository (Study ID #30987; accession #phs001704.v1.p1) with appropriate IRB-approved access restrictions. All custom software and scripts used in this study are available at github.com/BradnerLab.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank members of genome sequencing facilities at the Whitehead Institute of MIT, the Broad Institute, and the Dana-Farber Cancer Institute for facilitating this work; Charles Lin, Jaime Reyes, Jun Qi, and Gretchen Hankins for valuable input and reagents; Amy Johnson for the OSU-CLL cell line production of which was supported by NIH P30 CA016058 to the Ohio State University Cancer Center. C.J.O acknowledges support from the Leukemia Lymphoma Society and by an NCI Pathway to Independence Award (K99CA190861). A.J.F. acknowledges support from the NSF (2011125001), American Cancer Society (130965-PF-17-036-01-DMC) and from the Ashford

Foundation. M.L.F acknowledges support from the NCI (R01CA193910) and NIGMS (R01GM107427). J.R.B. acknowledges support from the American Cancer Society, the Leukemia Lymphoma Society, NIH (R01CA213442-01A1), and NIH PO1CA81534 to the CLL Research Consortium, as well as the Melton Family Fund for CLL Research and the Susan and Gary Rosenbach Fund for Lymphoma Research. This work was generously funded in part by an institutional grant award from the Broad Institute (to J.E.B.).

REFERENCES

Adam RC, Yang H, Rockowitz S, Larsen SB, Nikolova M, Oristian DS, Polak L, Kadaja M, Asare A, Zheng D, et al. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature 521, 366–370. [PubMed: 25799994]

Agathangelidis A, Scarfò L, Barbaglio F, Apollonio B, Bertilaccio MTS, Ranghetti P, Ponzoni M, Leone G, De Pascali V, Pecciarini L, et al. (2015). Establishment and Characterization of PCL12, a Novel $CD5^+$ Chronic Lymphocytic Leukaemia Cell Line. PLoS ONE 10, e0130195. [PubMed: 26110819]

Beekman R, Chapaprieta V, RussiñRocchirol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, Duran-Ferrer M, Kulis M, Serra F, Javierre BM, et al. (2018). The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nature Medicine 24, 868–880.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nature Biotechnology 28, 1045–1048.

Bertilaccio MTS, Scielzo C, Simonetti G, Ponzoni M, Apollonio B, Fazi C, Scarfò L, Rocchi M, Muzio M, Caligaris-Cappio F, et al. (2010). A novel $Rag2^{-/-}$ $\gamma_c{}^{-/-}$-xenograft model of human CLL. Blood 115, 1605–1609. [PubMed: 20018917]

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947–956. [PubMed: 16153702]

Brinkman EK, Chen T, Amendola M, and van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Research 42, e168. [PubMed: 25300484]

Brown JR, Hanna M, Tesar B, Werner L, Pochet N, Asara JM, Wang YE, Dal Cin P, Fernandes SM, Thompson C, et al. (2012). Integrative genomic analysis implicates gain of PIK3CA at 3q26 and MYC at 8q24 in chronic lymphocytic leukemia. Clin. Cancer Research 18, 3791–3802.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, and Greenleaf WJ (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods 10, 1213–1218. [PubMed: 24097267]

Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY, and Greenleaf WJ (2018). Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell 173, 1535–1548. [PubMed: 29706549]

Chapuy B, McKeown MR, Lin CY, Monti S, Roemer MGM, Qi J, Rahl PB, Sun HH, Yeda KT, Doench JG, et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell 24, 777–790. [PubMed: 24332044]

Chiorazzi N, Rai KR, and Ferrarini M (2005). Chronic lymphocytic leukemia. New England Journal of Medicine 352, 804–815. [PubMed: 15728813]

Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nature Genetics 48, 1193–1203. [PubMed: 27526324]

Cozma D, Yu D, Hodawadekar S, Azvolinsky A, Grande S, Tobias JW, Metzgar MH, Paterson J, Erikson J, Marafioti T, et al. (2007). B cell activator PAX5 promotes lymphomagenesis through stimulation of B cell receptor signaling. Journal of Clinical Investigation 117, 2602–2610. [PubMed: 17717600]

Delmore JE, Issa GC, Lemieux ME, Rahl PB, Shi J, Jacobs HM, Kastritis E, Gilpatrick T, Paranal RM, Qi J, et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell 146, 904–917. [PubMed: 21889194]

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology 34, 184–191.

Edelmann J, Holzmann K, Miller F, Winkler D, Bühler A, Zenz T, Bullinger L, Kühn MWM, Gerhardinger A, Bloehdorn J, et al. (2012). High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. Blood 120, 4783–4794. [PubMed: 23047824]

The FANTOM Consortium, Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461. [PubMed: 24670763]

Federation AJ, Polaski DR, Ott CJ, Fan A, Lin CY, and Bradner JE (2018) Identification of candidate master transcription factors within enhancer-centric transcriptional regulatory networks. *bioRxiv* 345413, doi.10.1101/345413.

Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Federov O, Morse EM, Keates T, Hickman TT, Felletar I, et al. (2010). Selective inhibition of BET bromodomains. Nature 468, 1067–1073. [PubMed: 20871596]

Grant CE, Bailey TL, and Noble WS (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018. [PubMed: 21330290]

Gröschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BAM, Erpelinck C, van der Velden VHJ, Havermans M, Avellino R, van Lom K, et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 157, 369–381. [PubMed: 24703711]

Grubor V, Krasnitz A, Troge JE, Meth JL, Lakshmi B, Kendall JT, Yamrom B, Alex G, Pai D, Navin N, et al. (2009). Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). Blood 113, 1294–1303. [PubMed: 18922857]

Hertlein E, Beckwith KA, Lozanski G, Chen TL, Towns WH, Johnson AJ, Lehman A, Ruppert AS, Bolon B, Andritsos L, et al. (2013). Characterization of a New Chronic Lymphocytic Leukemia Cell Line for Mechanistic In Vitro and In Vivo Studies Relevant to Disease. PLoS ONE 8, e76607–e76612. [PubMed: 24130782]

Karande A, Fialkow PJ, Nilsson K, Povey S, Klein G, Najfeld V, and Penfold G (1980). Establishment of a lymphoid cell line from leukemic cells of a patient with chronic lymphocytic leukemia. International Journal of Cancer 26, 551–556. [PubMed: 6263808]

Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, Lawrence MS, Kiezun A, Fernandes SM, Bahl S, et al. (2015). Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nature Communications 6, 8866.

Kellner J, Wierda W, Shpall E, Keating M, and McNiece I (2016). Isolation of a novel chronic lymphocytic leukemic (CLL) cell line and development of an in vivo mouse model of CLL. Leukemia Research 40, 54–59. [PubMed: 26601610]

Kim D, Langmead B, and Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12, 357–360. [PubMed: 25751142]

Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretti G, Husson H, Freedman A, Inghirami G, Cro L, Baldini L, et al. (2001). Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. Journal of Experimental Medicine 194, 1625–1638. [PubMed: 11733577]

König R, Chiang C-Y, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, et al. (2007). A probability-based approach for the analysis of large-scale RNAi screens. Nature Methods 4, 847–849. [PubMed: 17828270]

Kosmaczewska A, Ciszak L, Suwalska K, Wolowiec D, and Frydecka I (2005). 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. [PubMed: 15549146]

Kulis M, Merkel A, Heath S, Queirós AC, Schuyler RP, Castellano G, Beekman R, Raineri E, Esteve A, Clot G, et al. (2015). Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nature Genetics 47, 746–756. [PubMed: 26053498]

Kwiatkowski N, Zhang T, Rahl PB, Abraham BJ, Reddy J, Ficarro SB, Dastur A, Amzallag A, Ramaswamy S, Tesar B, et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. Nature 511, 616–620. [PubMed: 25043025]

Lacy J, Summers WP, Watson M, Glazer PM, and Summers WC (1987). Amplification and deregulation of MYC following Epstein-Barr virus infection of a human B-cell line. Proceedings of the National Academy of Sciences 84, 5838–5842.

Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152, 714–726. [PubMed: 23415222]

Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S, et al. (2014). Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell 26, 813–825. [PubMed: 25490447]

Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Böttcher S, et al. (2015). Mutations driving CLL and their evolution in progression and relapse. Nature 526, 525–530. [PubMed: 26466571]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359. [PubMed: 22388286]

Liao Y, Smyth GK, and Shi W (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote." Nucleic Acids Research 41, e108. [PubMed: 23558742]

Lin CY, Erkek S, Tong Y, Yin L, Federation AJ, Zapatka M, Haldipur P, Kawauchi D, Risch T, Warnatz H-J, et al. (2016). Active medulloblastoma enhancers reveal subgroup-specific cellular origins. Nature 530, 57–62. [PubMed: 26814967]

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. [PubMed: 25516281]

Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, and Young RA (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 153, 320–334. [PubMed: 23582323]

Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, Etchin J, Lawton L, Sallan SE, Silverman LB, et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science 346, 1373–1377. [PubMed: 25394790]

Mansouri L, Gunnarsson R, Sutton L-A, Ameur A, Hooper SD, Mayrhofer M, Juliusson G, Isaksson A, Gyllensten U, and Rosenquist R (2012). Next generation RNA-sequencing in prognostic subsets of chronic lymphocytic leukemia. American Journal of Hematology 87, 737–740. [PubMed: 22674506]

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, et al. (2013). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Research 42, D142–D147. [PubMed: 24194598]

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Research 34, D108–D110. [PubMed: 16381825]

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920. [PubMed: 22576172]

Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, Wang Q, Imbusch CD, Serva A, Koser SD, et al. (2016). DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nature Genetics 48, 253–264. [PubMed: 26780610]

Oldridge DA, Wood AC, Weichert-Leahey N, Crimmins I, Sussman R, Winter C, McDaniel LD, Diamond M, Hart LS, Zhu S, et al. (2015). Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature 528, 418–421. [PubMed: 26560027]

Ott CJ, Kopp N, Bird L, Paranal RM, Qi J, Bowman T, Rodig SJ, Kung AL, Bradner JE, and Weinstock DM (2012). BET bromodomain inhibition targets both c-Myc and IL7R in high-risk acute lymphoblastic leukemia. Blood 120, 2843–2852. [PubMed: 22904298]

Ozer HG, El-Gamal D, Powell B, Hing ZA, Blachly JS, Harrington B, Mitchell S, Grieselhuber NR, Williams K, Lai TH, et al. (2018) BRD4 profiling identifies critical chronic lymphocytic leukemia oncogenic circuits and reveals sensitivity to PLX51107, a novel structurally distinct BET inhibitor. Cancer Discovery 8, 458–477. [PubMed: 29386193]

Pagès H, Aboyoun P, Gentleman R, and DebRoy S (2018) Biostrings: efficient manipulation of biological strings. R package v2.48.0.

Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM, Fisch P, Timmer J, and Veelken H (2007). Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. Blood 109, 1202–1210. [PubMed: 17053054]

Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, Munar M, Rubio-Pérez C, Jares P, Aymerich M, et al. (2015). Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature 526, 519–524. [PubMed: 26200345]

R Core Team (2013) R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria ISBN 3–900051-07–0, http://www.R-project.org/.

Rashid NU, Giresi PG, Ibrahim JG, Sun W, and Lieb JD (2011) ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biology 12, R67. [PubMed: 21787385]

Rendeiro AF, Schmidl C, Strefford JC, Walewska R, Davis Z, Farlik M, Oscier D, and Bock C (2016). Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. Nature Communications 7, 1–12.

Risso D, Schwartz K, Sherlock G, and Dudoit S (2011) GC-content normalization for RNA-seq data. BMC Bioinformatics 12, 480. [PubMed: 22177264]

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ziller MJ, Amin V, Whitaker JW, Ward LD, Quon G, Pfenning A, Wang X, Claussnitzer M, et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. [PubMed: 25693563]

Roberts AW, Davids MS, Pagel JM, Kahl BS, Puvvada SD, Gerecitano JF, Kipps TJ, Anderson MA, Brown JR, Gressick L, et al. (2015). Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. New England Journal of Medicine 28, 311–322.

Rosén A, Bergh A-C, Gogok P, Evaldsson C, Myhrinder AL, Hellqvist E, Rasul A, Björkholm M, Jansson M, Mansouri L, et al. (2012). Lymphoblastoid cell line with B1 cell characteristics established from a chronic lymphocytic leukemia clone by in vitro EBV infection. Oncoimmunology 1, 18–27. [PubMed: 22720208]

Saint-André V, Federation AJ, Lin CY, Abraham BJ, Reddy J, Lee TI, Bradner JE, and Young RA (2016). Models of human core transcriptional regulatory circuitries. Genome Research 26, 385–396. [PubMed: 26843070]

Schep AN, and Kummerfeld SK (2017) iheatmapr: Interactive complex heatmaps in R. The Journal of Open Source Software 2, 359.

Schep AN, Wu B, Buenrostro JD, and Greenleaf WJ (2017) chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nature Methods 14, 975–978. [PubMed: 28825706]

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87. [PubMed: 24336571]

Shin H, Liu T, Manrai AK, Liu XS (2009) CEAS: *cis*-regulatory element annotation system. Bioinformatics 25, 2605–2606. [PubMed: 19689956]

Spitz F, and Furlong EEM (2012). Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics 13, 613–626.

Splinter E, de Wit E, van de Werken HJG, Klous P, and de Laat W (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58, 221–230. [PubMed: 22609568]

Stacchini A, Aragno M, Vallario A, Alfarano A, Circosta P, Gottardi D, Faldella A, Rege-Cambrin G, Thunberg U, Nilsson K, et al. (1999). MEC1 and MEC2: two new cell lines derived from B-chronic lymphocytic leukaemia in prolymphocytoid transformation. Leukemia Research 23, 127–136. [PubMed: 10071128]

Trapnell C, Williams B, Pertea G, Mortazavi A, Kwan G, van Baren J, Salzberg S, Wold B, and Pachter L (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching. Nature Biotechnology 28, 511–515.

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, et al. (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. New England Journal of Medicine 365, 2497–2506. [PubMed: 22150006]

van de Werken HJG, Landan G, Holwerda SJB, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BAM, et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nature Methods 9, 969–972. [PubMed: 22961246]

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, and Young RA (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319. [PubMed: 23582322]

Zhang X, Choi PS, Francis JM, Imielinski M, Watanabe H, Cherniack AD, and Meyerson M (2015). Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. Nature Genetics 48, 176–182. [PubMed: 26656844]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biology 9, R137. [PubMed: 18798982]

Ziller MJ, Edri R, Yaffe Y, Donaghey J, Pop R, Mallard W, Issner R, Gifford CA, Goren A, Xing J, et al. (2015). Dissecting neural differentiation regulatory networks through epigenetic footprinting. Nature 518, 355–359. [PubMed: 25533951]

HIGHLIGHTS

- Histone acetylation and chromatin accessibility reveals enhancer signatures of CLL
- Super enhancers mediate the CLL transcription factor core regulatory circuitry
- PAX5 is a core regulator of CLL super enhancers essential for CLL cell survival
- BET inhibition effectively disrupts CLL super enhancer circuits

SIGNIFICANCE

Although transcriptional dysregulation has long been observed in CLL, a comprehensive understanding of the enhancers and transcription factors driving this aberrant cell state has not be elucidated. Using integrated analysis of enhancer profiles across a cohort of primary CLL patient samples, we define large enhancer clusters - super enhancers - characteristic of the CLL epigenome. Coupled with high-resolution chromatin accessibility measurements, we construct the enhancer-based core regulatory circuits governed by a limited number of highly connected transcription factors including PAX5. Functional genetic screening identified PAX5 as a singularly essential, hyperconnected enhancer factor in CLL. We further show that CLL super enhancer networks can be pharmacologically perturbed by BET bromodomain inhibition, nominating a potential therapeutic option for this disease.

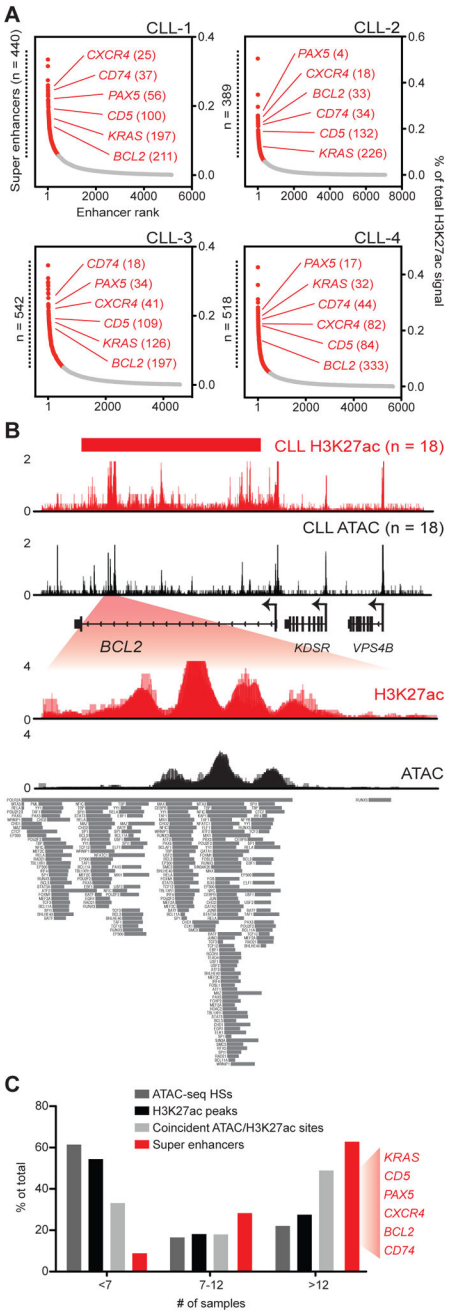


Figure 1: The super enhancer landscape of CLL.

A. Enhancer profiles of four representative CLL samples, using H3K27ac ChIP-seq signal. Enhancers are ranked ordered as a percentage of total signal. SEs are highlighted in red with ranks of selected SE-associated genes. SE identification for all samples are included in Table S2.

B. Overlaid H3K27ac and ATAC-seq profiles of a cohort of primary CLL B cell samples (n = 18 for each). Shown is the SE detected at the *BCL2* locus, with a highlighted region displaying chromatin accessibility within 'valleys' of H3K27ac-enriched regions. Bottom track shows TF binding sites as determined by ChIP-seq (ENCODE, genome.ucsc.edu/

encode). Top track hg19 coordinates chr18:60,748,958-61,129,600; bottom track coordinates chr18:60,825,744-60,832,442; y-axis values represent normalized read density.

C. Distribution analysis of shared ATAC-seq HSs and H3K27ac regions across the cohort. Regions that overlap in the listed number of samples are shown as a percentage of the total number of each feature identified. Recurrent SE-associated genes are highlighted. Features were determined to be recurrent between samples if at least one base pair overlapped between determined peaks.

See also Figure 1 and Tables S1 and S2

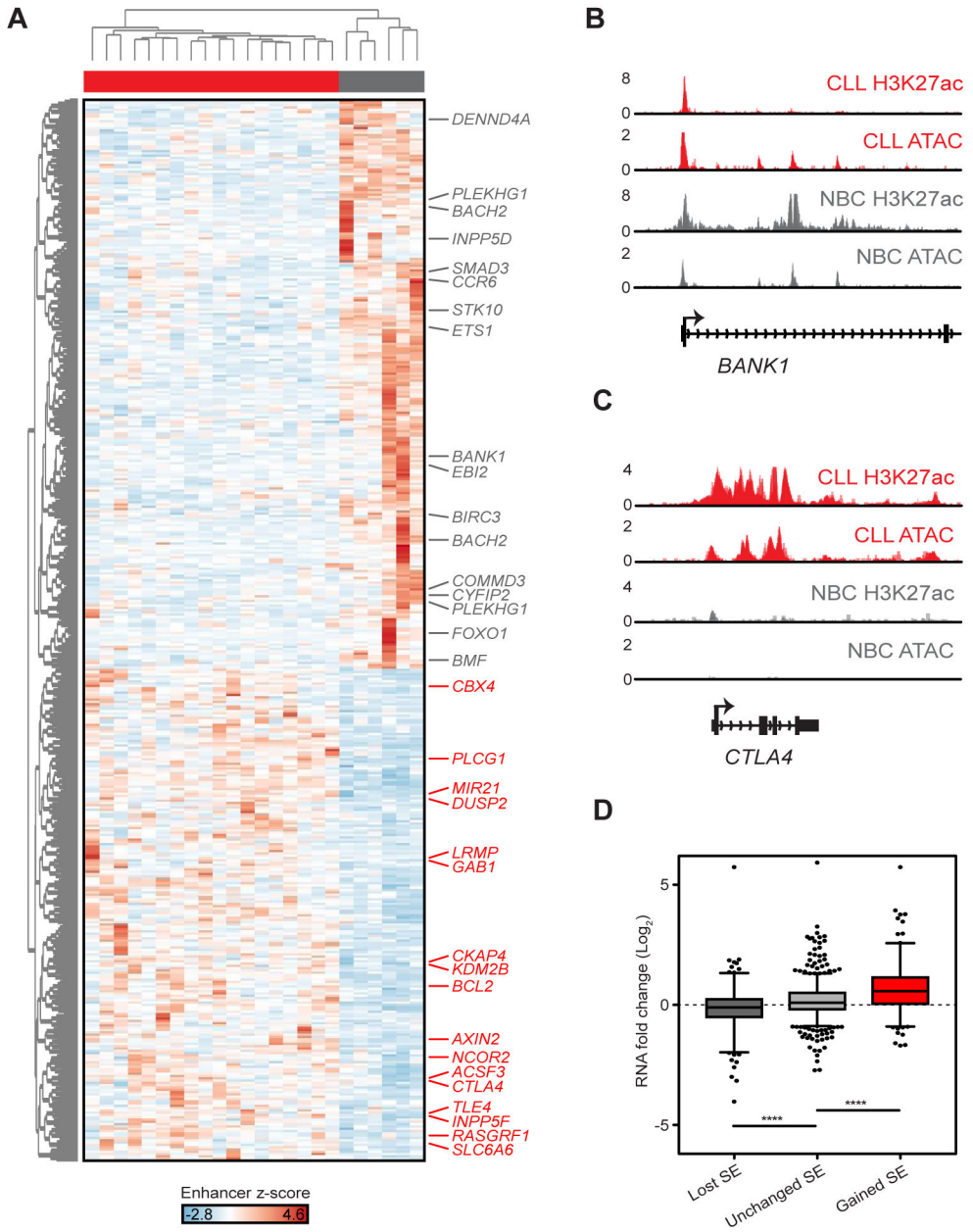


Figure 2: Distinct SE profiles of CLL

A. Differential H3K27ac enrichment between CD19^{+} NBC and CLL cells. Each column represents an SE found to have differential enrichment between the two cohorts (p value < 0.1). Data is presented as row-normalized z-scores, and clustering of samples and enhancers is performed by euclidean distance using iheatmapr.

B. H3K27ac and ATAC-seq profiles of the *BANK1* locus (chr18: 60,825,744 - 60,832,442) in CLL and NBC samples.

C. H3K27ac and ATAC-seq profiles of the *CTLA4* locus (chr2: 204,724,952 - 204,748,658) in CLL and NBC samples.

D. Gene expression changes as measured by mRNA-seq of CLL Lost SE, Unchanged SE, and Gained SE; ****, p < 0.0001 (unpaired, two-tailed t test with Welch's correction,

GraphPad Prism). Boxes represent the 25th percentile, median and 75th percentile. Tails represent the 5th and 95th percentiles, all other data are represented as individual points. See also Figure S2.

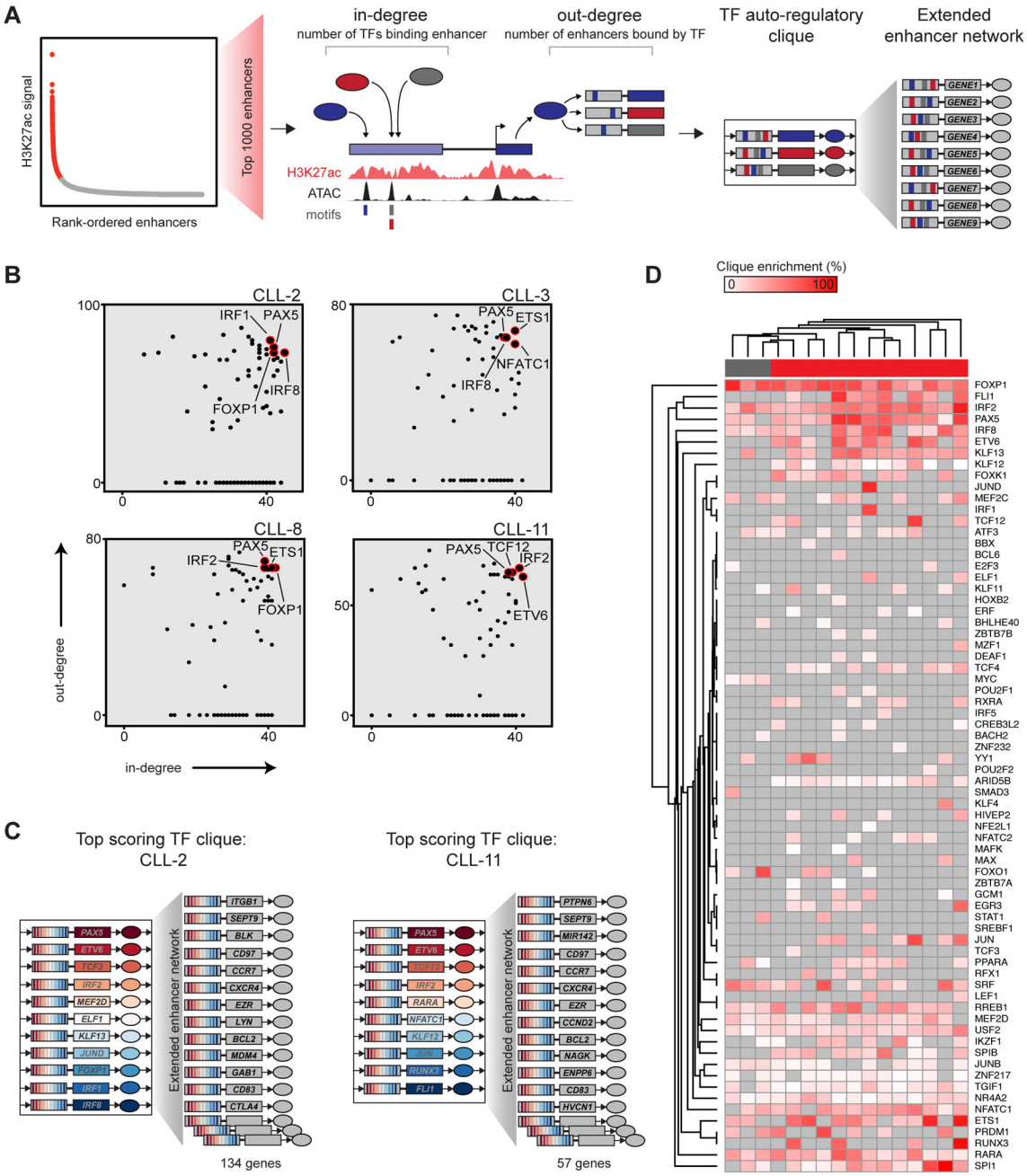


Figure 3: TF connectivity analysis defines core gene regulatory circuitry of CLL.

A. Schematic of enhancer-based CRC analysis. For each sample, the top 1000 enhancers as determined by ROSE2 are extracted. For every TF associated with a top enhancer, in-degrees are assessed by motif analysis within HSs of the enhancer; out-degrees are assessed for each TF associated with a top enhancer by determining all other bound enhancers at TF gene loci. Node connections between TFs are used to discern auto-regulatory cliques that in turn regulate a broader enhancer network.

B. Degree plots of four representative CLL samples, with several highly connected TFs highlighted. Plots for all CLL, NBC, and cell line samples are provided in Figure S3E.

C. The top scoring TF cliques (as defined by the cumulative connectivity of constituent TFs) for two primary CLL samples (CLL-2 and CLL-11), with the extended enhancer circuitry shown on the right. Each TF is associated with an enhancer with a predicted binding site of each other clique within an HSs of that enhancer. Enhancers within the extended network also contain predicted sites for each clique TF (a total of 134 genes for CLL-2 clique, 57 genes for CLL-11 clique; 13 extended network genes listed for each).

D. A heatmap of clique enrichment scores for the union of all TFs associated with top enhancers across all NBC and CLL samples with paired H3K27ac and ATAC-seq datasets. Clique enrichment scores are calculated by the percentage of cliques within each sample of which that TF is a constituent. Gray boxes are used when a TF is not associated with any clique in that given sample. TFs and samples are clustered by euclidean distance using Morpheus software to reveal the most consistent highly connected TFs.

See also Figures S3 and S4 and Table S3.

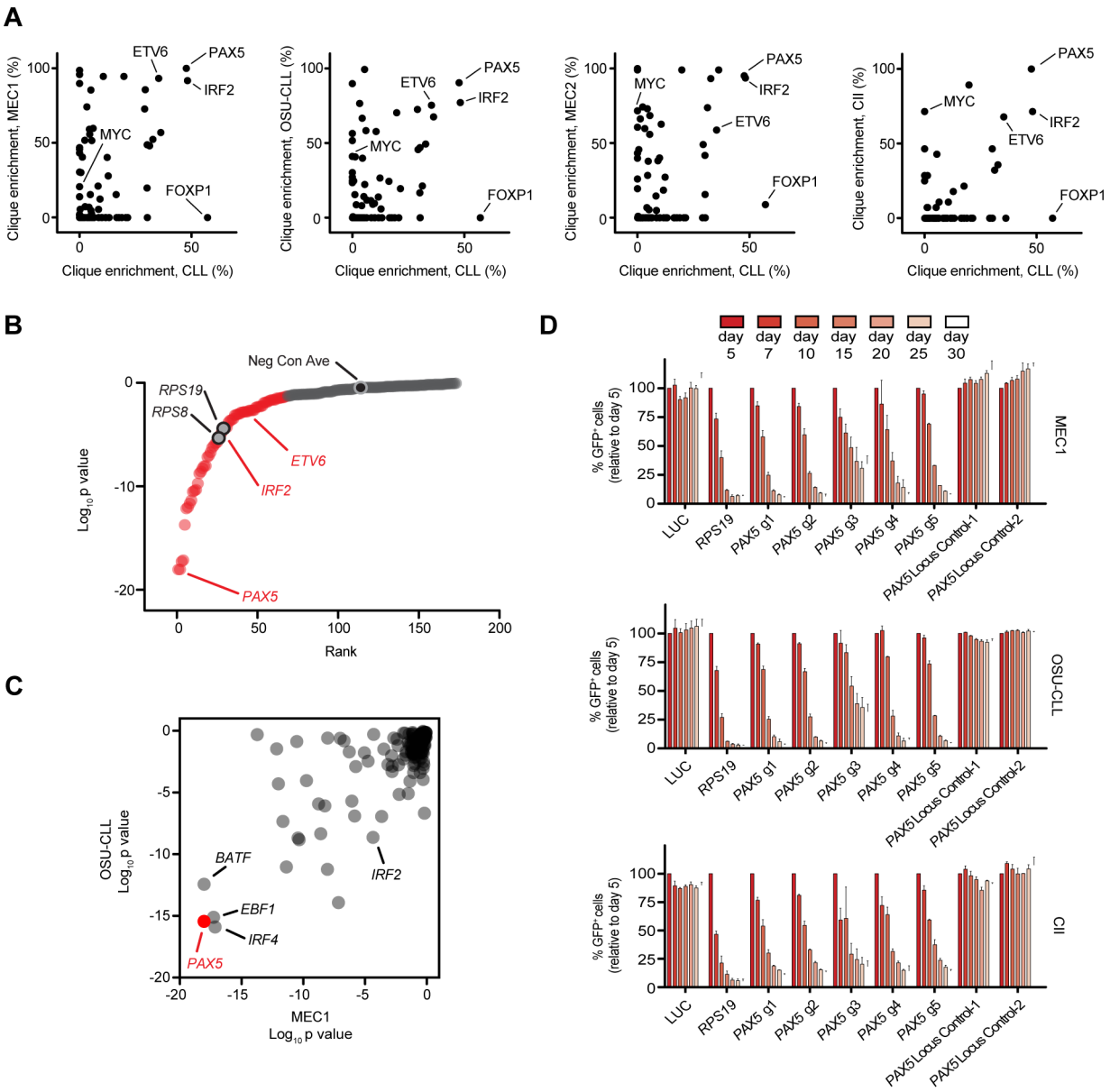


Figure 4: Essential TFs of CLL.

A. TF clique enrichment score correlation between individual cell lines and primary CLL cells. A cumulative analysis was performed on the cohort of primary CLL samples to derive an average clique enrichment value across the cohort (see Methods). Highlighted are highly correlated CRC TFs (PAX5, ETV6, IRF2), and selected cell line-enriched TFs (MYC) and primary cell-enriched TFs (FOXP1).

B. TF CRISPR screening results in MEC1 cells. Gray circles represent p value dependency scores for *RPS8* and *RPS19,* two positive control essential genes; black circle represents the average score for all negative control gRNAs (n = 100) included in the library. All other circles represent the average gRNA p value dependency scores for each TF gene. Genes with statistically significant dependency scores ($p < 0.05$) are in red, non-essential genes are in

gray. P values determined by RSA analysis (König et al., 2007). Neg Con Ave represents the average p value of 100 non-targeting control gRNAs included in the library.
C. Correlation of CRC TF dependency scores between the MEC1 and OSU-CLL screens.
D. Negative selection competition assay of three CLL cell lines stably expressing Cas9 and transduced with lentiviral constructs expressing gRNAs targeting the *PAX5* locus. LUC, control gRNA targeting the luciferase gene. Data are represented as mean ± SD, n = 3.
See also Figure S5 and Table S4.

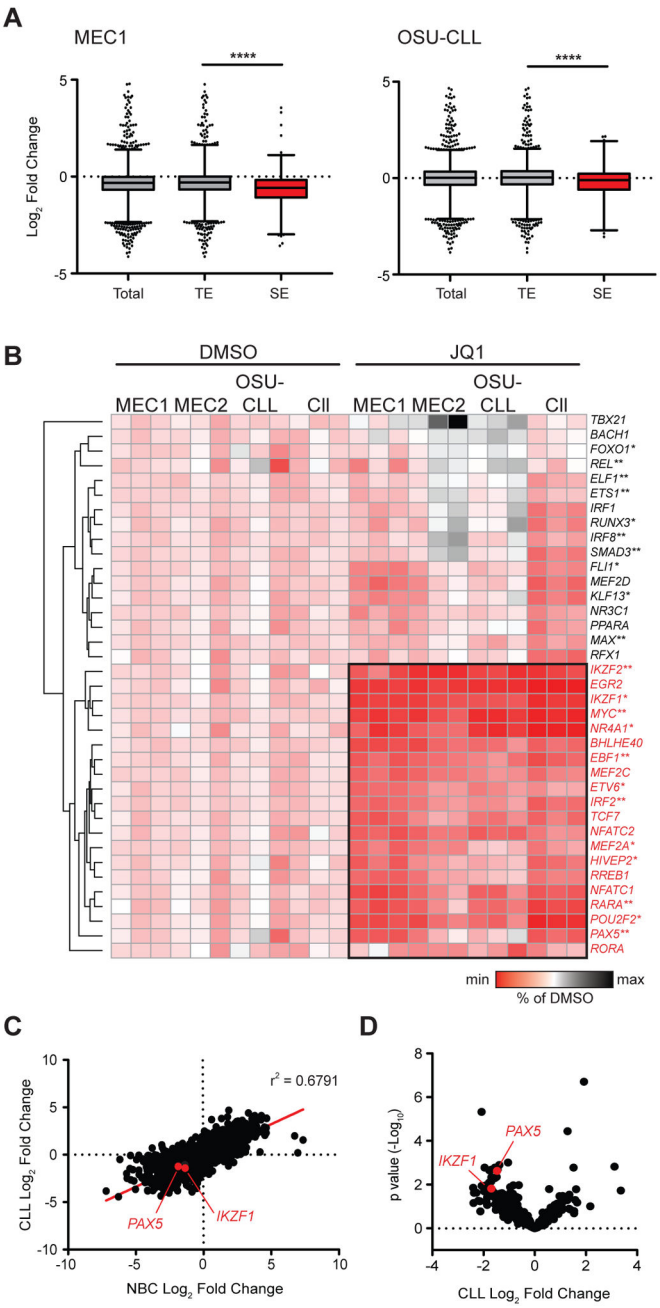


Figure 5: Effects of BET bromodomain inhibition on CLL CRC expression.

A. RNA-seq of two CLL cell lines treated with the BET bromodomain inhibitor JQ1 for 6 hr (1 μM). Change in mRNA transcripts was assessed compared to control samples, and plotted as total genes, TE-associated genes, and SE-associated genes; ****, $p < 0.0001$, unpaired, two-tailed t test with Welch's correction (GraphPad Prism). Boxes represent the 25th percentile, median and 75th percentile. Tails represent the 5th and 95th percentiles, all other data are represented as individual points.

B. JQ1 effects on mRNA levels of all CRC TFs as determined from MEC1, MEC2, OSU-CLL, and CII cell line enhancer profiling. TF FPKM values for each cell line were

normalized by row to the mean of vehicle-treated samples and displayed as percent of vehicle-treated mean. Effects on TFs are hierarchically clustered by euclidean distance using Morpheus software. The most effected TFs cluster together, including several required for CLL cell proliferation including *MYC* and *PAX5* (* = significant dependency in one cell line CRISPR screen; ** = significant dependency score in both screens).

C. Correlation between average gene expression changes observed in primary CLL cells compared to NBCs with 6 hr JQ1 treatment (1 μM). Genes shown restricted to those with $p < 0.05$. Four different primary CLL samples were analyzed in duplicate (n=8); two NBC samples were analyzed in duplicate (n=4).

D. Volcano plot of TF gene expression changes in primary CLL cells treated with either JQ1 or DMSO as vehicle control. TF genes shown restricted to those with average FPKM value > 1. P values determined by t tests using a custom R script.

See also Figure S6.

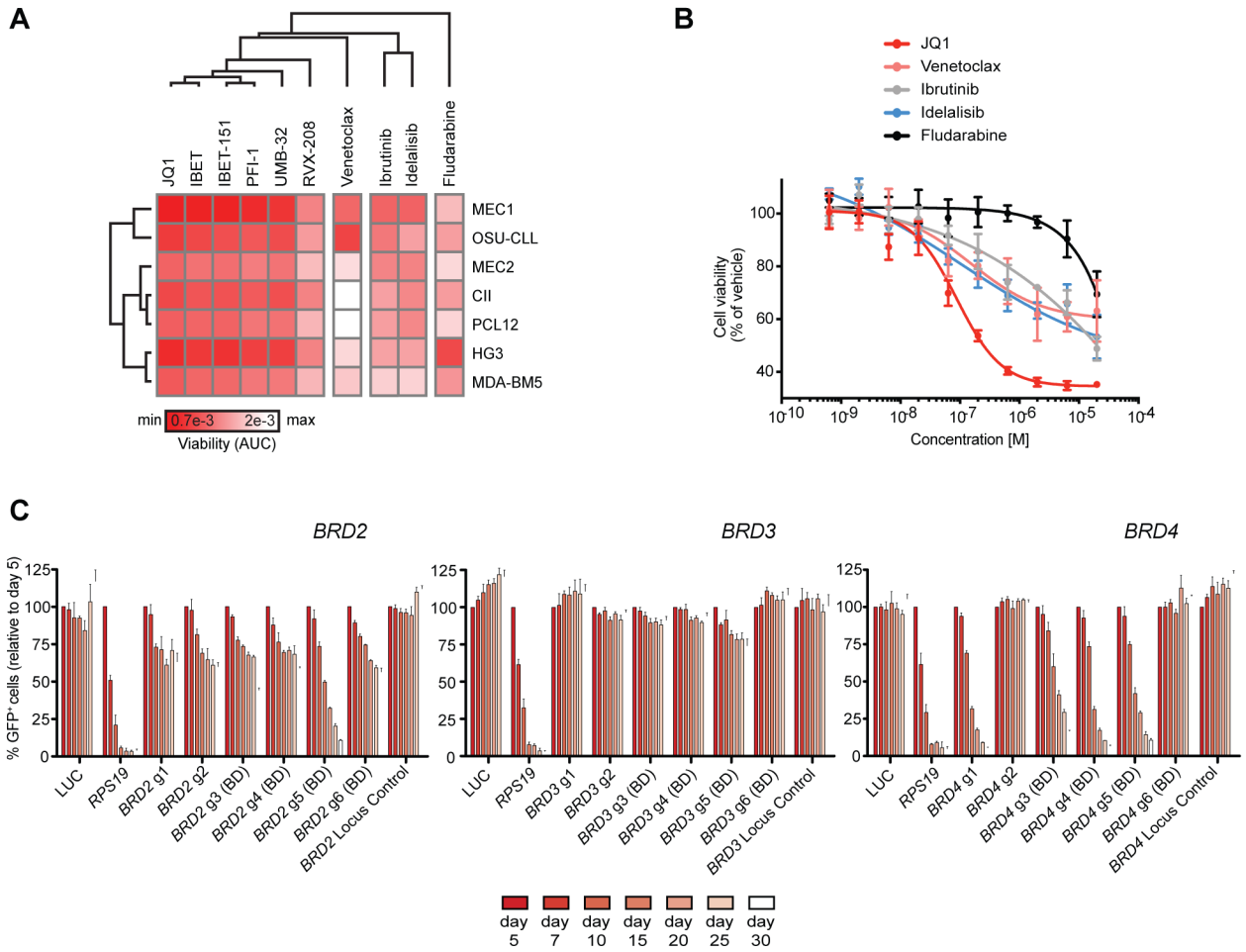


Figure 6: CLL dependency on BET bromodomain proteins.

A. Clustered heatmap of JQ1 effects on the viability of a panel of CLL cell lines. Also shown are effects of CLL targeted therapies and the chemotherapy agent fludarabine. Cell viability is measured by alamar blue reagent uptake 5 days following treatment (4 replicates at each dose). Drug response is summarized by average area-under-the-curve (AUC, GraphPad Prism). Clustering was performed by euclidean distance using Morpheus software.

B. Dose-response of JQ1 and other CLL drugs in the MEC1 cell line, normalized to vehicle-treated controls. Data are represented as mean ± SD, n = 4.

C. Negative competition CRISPR assay in the MEC1 cell line with gRNAs targeting three BET bromodomain proteins expressed in CLL: BRD2, BRD3, BRD4. Four gRNAs for each gene target exons coding for JQ1-binding bromodomains (BD). LUC, control gRNA targeting the luciferase gene. Data are represented as mean ± SD, n = 3.

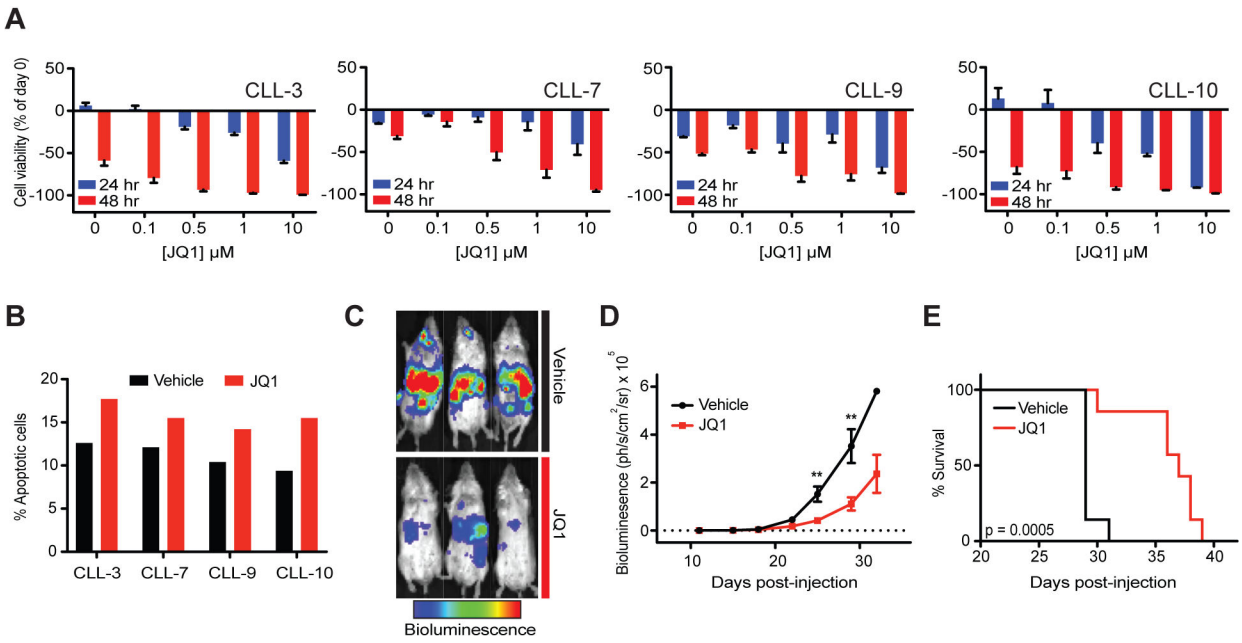


Figure 7: JQ1 effects on primary CLL cells and in vivo model of CLL.

A. Dose-response of JQ1 on four individual samples of primary CLL B cells, normalized to viability readings at day zero. Cellular viability as approximated by ATP content measurements at two time points (24 and 48 hr). Data are represented as mean ± SEM, n = 3.

B. Apoptotic cell populations as defined by Annexin V positive/7-AAD negative staining after 48 hr of JQ1 treatment, 1 μM.

C. Bioluminescence imaging of JQ1 and vehicle-treated mice at day 29 following MEC1 inoculation.

D. Bioluminescence quantification of JQ1- (n = 7) and vehicle-treated mice (n = 7). **, $p < 0.05$, unpaired, two-tailed t test (GraphPad Prism). Data are represented as mean ± SEM.

E. Kaplan-Meier plot of JQ1- and vehicle-treated mice (n=7). P value was generated by log-rank Mantel-Cox test (GraphPad Prism).

See also Figure S7.