cis-Regulatory Circuits Regulating NEK6 Kinase Overexpression in Transformed B Cells Are Super-Enhancer Independent

Graphical Abstract

Highlights

- **NEK6**, a mitotic kinase gene, is overexpressed in specific types of B cell lymphoma
- Super-enhancer is dispensable for elevated NEK6 expression in transformed B cells
- Genome engineering reveals regulatory elements for NEK6 cis-regulatory circuits
- Need for rigorous testing of predicted cis-regulatory circuits and super-enhancers

Authors

Yue Huang, Olivia I. Koues, Jiang-yang Zhao, Regina Liu, Sarah C. Pyfrom, Jacqueline E. Payton, Eugene M. Oltz

Correspondence

jpayton@wustl.edu (J.E.P.), eoltz@wustl.edu (E.M.O.)

In Brief

Huang et al. functionally dissect cis-regulatory circuits associated with NEK6, a mitotic kinase overexpressed in B cell lymphoma. Only a subset of predicted enhancers and CTCF sites cooperatively constructs the regulatory hub of NEK6. A super-enhancer is completely dispensable for maintaining NEK6 expression and architecture in transformed B cells.

Accession Numbers

GSE87323

Huang et al., 2017, Cell Reports 18, 2918–2931
March 21, 2017 © 2017 The Author(s).
http://dx.doi.org/10.1016/j.celrep.2017.02.067
cis-Regulatory Circuits Regulating NEK6 Kinase Overexpression in Transformed B Cells Are Super-Enhancer Independent

Yue Huang,1 Olivia I. Koues,1,2 Jiang-yang Zhao,1 Regina Liu,1 Sarah C. Pyfrom,1 Jacqueline E. Payton,1,* and Eugene M. Oltz1,3,*
1Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA
2Present address: Vanderbilt Technologies for Advanced Genomics, Vanderbilt University Medical Center, Nashville, TN 37232, USA
3Lead Contact
*Correspondence: jpayton@wustl.edu (J.E.P.), eoltz@wustl.edu (E.M.O.)
http://dx.doi.org/10.1016/j.celrep.2017.02.067

SUMMARY

Alterations in distal regulatory elements that control gene expression underlie many diseases, including cancer. Epigenomic analyses of normal and diseased cells have produced correlative predictions for connections between dysregulated enhancers and target genes involved in pathogenesis. However, with few exceptions, these predicted cis-regulatory circuits remain untested. Here, we dissect cis-regulatory circuits that lead to overexpression of NEK6, a mitosis-associated kinase, in human B cell lymphoma. We find that only a minor subset of predicted enhancers is required for NEK6 expression. Indeed, an annotated super-enhancer is dispensable for NEK6 overexpression and for maintaining the architecture of a B cell-specific regulatory hub. A CTCF cluster serves as a chromatin and architectural boundary to block communication of the NEK6 regulatory hub with neighboring genes. Our findings emphasize that validation of predicted cis-regulatory circuits and super-enhancers is needed to prioritize transcriptional control elements as therapeutic targets.

INTRODUCTION

Cell identity and function rely on stringently controlled programs of gene expression, perturbations of which underlie diseases, including autoimmunity and cancer. Genome-wide association studies have revealed that most pathogenic changes in gene expression are linked to variants in regulatory elements rather than coding sequences (Maurano et al., 2012). A dissection of cis-regulatory circuits controlling transcriptomes in normal and diseased cells remains an important objective. Most cis-regulatory circuits are composed of gene-proximal promoters and distal enhancers, which serve as conduits for transcription factors (TFs) and communicate with each other via physical contact, forming a series of loops in nuclear chromatin (Bulger and Groudine, 2011).

Conventional enhancers (CEs), both active and poised, can be identified in the genome as nucleosome-free regions. The activity level of each conventional enhancer is revealed by the density of certain histone modifications, prototypically histone H3 acetylated at lysine 27 (H3K27ac) (Bulger and Groudine, 2011). Recent epigenome analyses have revealed a new class of regulatory regions, coined super-enhancers (SEs) (Whyte et al., 2013), which are characterized by broad stretches of H3K27ac. Most SEs are dense clusters of highly active CEs, which bind lineage-restricted TFs. Indeed, SEs normally co-localize with a limited set of genes that are most essential for cell identity and function. The acquisition or amplification of SEs near oncogenes contributes to several classes of cancer (Hnisz et al., 2013; Mansour et al., 2014). SEs are also enriched for disease-associated sequence variants, some of which presumably disrupt TF binding sites to alter super-enhancer function and expression of its associated gene(s) (Hnisz et al., 2013; Koues et al., 2016). However, contributions of SEs to gene expression programs have been mostly assumed from correlative chromatin profiling rather than by direct testing (Proudhon et al., 2016). Furthermore, it remains controversial whether SEs represent a new paradigm in transcriptional regulation or merely clusters of CEs that additively promote transcription (Dukler et al., 2016; Hay et al., 2016).

In addition to cis-regulatory elements, gene expression programs are significantly influenced by chromosome architecture, which facilitates or impairs promoter-enhancer contacts. The architecture of mammalian genomes is compartmentalized into topologically associated domains (TADs), which are highly conserved among cell types and species (Dixon et al., 2012). Loci within each TAD interact with one another but are largely cordoned off from neighboring TADs. Each of these architectural building blocks is subdivided into structures called sub-TADs or contact domains, which are composed of loops between CTCF binding elements (structural loops) or between promoters and enhancers (regulatory loops). At a biochemical level, structural loops form via dimeric interactions between CTCF proteins bound in a convergent orientation at two distinct sites and are stabilized by association with the ring-like cohesin complex...
for the mitosis-associated kinase NEK6 overexpressed in BCL (Mareschal et al., 2015). We find that completely dispensable for expression. Strikingly, a B cell-specific super-enhancer is comprised of multiple CEs, predicted by correlative algorithms to but also emphasizes the need to rigorously test predictions, expression is regulated in normal and pathogenic B cells NEK6 only a subset of CEs, predicted by correlative algorithms to communicate between enhancers and alternative genes (Lupiáñez et al., 2015; Hnisz et al., 2016b).

Similarly, a deeper understanding of the regulatory determinants that underlie oncogenic gene expression programs remains a basic mission of cancer research (Sur and Taipale, 2016). Pathogenic expression programs have been characterized for many cancers, including various types of B cell lymphoma (BCL) (Jiang et al., 2016; Morin et al., 2010). A common class of BCL, termed follicular lymphoma (FL), is incurable. Most FLs exhibit an indolent clinical course but often transform to a more aggressive cancer, termed diffuse large BCL (DLBCL) (Lenz and Staudt, 2010). Recently, we showed that pathogenic gene expression programs in FL are coordinated by a common set of TFs that, in turn, augment or attenuate activities of their target enhancers when compared with normal B cell counterparts, termed centrocytes (CCs) (Koues et al., 2015). Integrative transcriptome and epigenome analyses revealed a blueprint of pathogenic cis-regulatory circuits associated with FL, which predicted connections between distal enhancers and promoters of dysregulated genes. Similar correlation-based circuitries governing gene expression have been constructed for many normal and transformed cell types (Thurman et al., 2012), revealing a new collection of potential targets for epigenetic therapeutics. However, the validity of predicted circuits remains largely untested at the functional level. This gap is particularly important given that a majority of predicted cis-regulatory circuits consist of multiple enhancers connected to a single gene or, conversely, multiple genes connected to a single enhancer (Thurman et al., 2012).

Here, we functionally dissect a predicted cis-regulatory circuit for the mitosis-associated kinase NEK6, which is commonly overexpressed in BCL (Mareschal et al., 2015). We find that only a subset of CEs, predicted by correlative algorithms to regulate NEK6 in BCL, is required to maintain its elevated expression. Strikingly, a B cell-specific super-enhancer is completely dispensable for NEK6 expression and maintenance of a regulatory hub that co-localizes its promoter with many distal CEs. A cluster of CTCF sites at one border of the NEK6 contact domain serves as a chromatin and architectural boundary to minimize the functional impact of its regulatory hub with neighboring genes. Our study not only provides insights into how NEK6 expression is regulated in normal and pathogenic B cells but also emphasizes the need to rigorously test predictions, based solely on chromatin landscapes, regarding cis-regulatory circuits and super-enhancer function.

RESULTS

The NEK6 cis-Regulatory Circuit Distinguishes FL Subsets

Very few correlation-based predictions for cis-regulatory circuits in normal or transformed cells have been validated functionally by targeted engineering of control elements within their native chromosomal context (Sur and Taipale, 2016). To rigorously test a manageable set of predictions, we prioritized pathogenic cis-regulatory circuits associated with CC transformation into FL (Koues et al., 2015). Prioritization of differentially expressed genes and their corresponding regulomes was tiered for recurrence of pathogenic enhancers in FL samples, altered levels of gene expression, relevant TF binding, and gene function (Figure S1A; Supplemental Experimental Procedures; Table S1). The scheme yielded seven regulatory clusters and accompanying genes, which we considered to be of high priority for functional dissection (Table S2). Each of the seven regions consists of multiple enhancers and potential target genes, which renders comprehensive analysis of all prioritized circuits unwieldy. From the seven, we selected a region spanning NEK6 and several neighboring genes for in-depth functional studies, based on multiple criteria. We first tested enhancer activities using luciferase reporters for a series of regulatory elements from the seven surviving regions, each of which displays augmented H3K27ac in FL compared with CCs. A regulatory element in the NEK6 region (CE1) displays the most robust enhancer activity in both an EBV-transformed B cell line (GM12878) and a human BCL line (Farage; Figure S1B). Moreover, NEK6, a central gene in the identified circuit, encodes a serine/threonine kinase that mediates mitotic progression, is overexpressed in many cancers, and is essential for sustained growth of tumors derived from numerous tissues (Fry et al., 2012).

With regard to B cell oncogenesis, NEK6 expression distinguishes the two known subtypes of DLBCL, exhibiting elevated expression in the germinal center (GC) compared with the activated B cell (ABC) subtype (Mareschal et al., 2015). Epigenome analyses revealed that FL also segregates into two analogous classes (Koues et al., 2015), subtype 1 (GC-like) and subtype 2 (ABC-like). Strikingly, NEK6 expression is significantly elevated in subtype 1 FL, further highlighting its similarity to GC-DLBCL (Figure 1A). One final criterion in selecting the NEK6 region for further study is its rich regulatory landscape, which seemingly consists of multiple enhancers augmented in BCL and a series of potential architectural elements (see below). Thus, we suspected that analysis of NEK6 cis-regulatory circuits would provide insights into enhancer and architectural elements important for cell-type-, lymphoma-, or FL-subtype-specific expression of this mitosis-associated kinase.

The NEK6 Regulatory Landscape

To identify the collection of distal architectural and regulatory elements that contribute to elevated NEK6 expression in BCL, we leveraged data from public databases (ENCODE Project Consortium, 2012; Koues et al., 2015). Nucleosome-depleted
regions demarcate more than a dozen active or poised elements spread over a 500-kb region encompassing NEK6 and its neighboring genes (Figure 1B; formaldehyde-assisted isolation of regulatory elements sequencing [FAIRE-seq]/DNase I hypersensitive site sequencing [DNase-seq]). Several of these regions are bound by architectural factors, CTCF and RAD21, in GM12878, suggesting they may serve as structural or boundary elements (CTCF sites, CS1–7). NEK6 has two annotated transcription start sites (TSSs), which are both active in human B cells and GM12878 (Figure S1C). H3K27ac peaks coincide with 14 nucleosome-depleted regions in FL samples, indicating positions of active conventional enhancers (CE1–14). Importantly, many of these enhancers exhibit a higher density of H3K27ac in FL compared with normal CC counterparts, suggesting they are hyperactive in transformed B cells. A subset of active enhancers (CE3–9) is clustered in a region ~63 to ~40 kb upstream of NEK6, which is designated as a super-enhancer (SE1) in both FL and CC samples using the rank ordering of super-enhancers (ROSE) algorithm (Lovén et al., 2013; Whyte et al., 2013) to analyze H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) data (Figure 1C). When compared with other cell types, the activities of CE1, CE10, and SE1 are primarily restricted to B cells (Figure S1D). Another conventional enhancer region, CE13–14, is also active in a subset of other cell types that express NEK6. These genomipen analyses suggest that CE1, CE10, SE1 and, perhaps, CE13–14 are critical enhancers for driving high levels of NEK6 expression in activated or transformed B cells.

Extensive genetic manipulations are required to dissect the NEK6 regulome; however, this approach is currently infeasible using primary human B cells. As such, we identified a tractable cell model that mirrors the NEK6 chromatin landscape in primary FL. As shown in Figure 1B, the transformed human B cell line GM12878 meets this criterion, while the human T lymphocyte cell line Jurkat exhibits a chromatin landscape largely devoid of active regulatory elements near NEK6, thus providing a negative control. In addition to recapitulating patterns of active enhancers in primary B and FL cells, the CE3–9 region is classified as a super-enhancer in GM12878 (Figure 1C). NEK6 expression in GM12878 is comparable to levels observed in tonsillar B cells, the majority of which are activated, whereas NEK6 transcripts are nearly undetectable in Jurkat (Figure 1D).

In addition to NEK6, two neighboring genes, LHX2 and PSMB7, are predicted to connect with many of the B cell-restricted enhancers in FL using a gene circuitry algorithm (Koues et al., 2015). LHX2 is a TF involved in the differentiation of developing lymphoid and neural cell precursors and is a putative oncogene for pancreatic tumors (Zhou et al., 2014). PSMB7 is a proteasome subunit that was identified as a biomarker for breast and colon cancers (Munkácsy et al., 2010). As shown in Figure 1E, expression of these two genes, but not the more distal DENND1A, are modestly elevated in FL and/or tonsillar B cells compared with human CCs. All of these genes are expressed at varying levels in GM12878 (Figure S1E). As such, functional dissection of the NEK6 cis-regulatory circuit can be achieved using GM12878, which recapitulates prominent features of the FL regulome.

Spatial Convergence of NEK6 Distal Regulatory Elements

Proper control of gene expression requires direct contact of distal regulatory elements with their target promoters. Many cell-type-specific contacts between enhancers and promoters are confined within TADs and further restricted by boundary elements to minimize inappropriate enhancer-promoter communication. To elucidate the NEK6 interactome within its chromosomal neighborhood, we analyzed publicly available genome-wide chromosome conformation capture (Hi-C) data for a ~2-Mb region in GM12878 (Figure 2A) (Rao et al., 2014). Based on interactomes conserved among cell types, the TAD containing NEK6 spans ~1 Mb encompassing DENND1A, LHX2, NEK6, and PSMB7. In GM12878, this region also contains several sub-TADs, one of which includes NEK6, spanning from the DENND1A promoter to PSMB7 (~500 kb). Within the sub-TAD, there is a robust contact domain spanning from the cluster of upstream CTCF sites (CS2–4) to the downstream NEK6 promoters (TSS1–2). More focal contacts are observed between both NEK6 promoters and pockets of upstream regulatory elements, especially with CE1 and SE1. Hi-C data revealed associations of the NEK6 locus with PSMB7 and, to a lesser extent, with LHX2, suggesting a potential mechanism for their elevated expression in FL. Finally, NEK6 is flanked by two sets of CTCF sites pointing in convergent orientations, a trio located ~130 kb upstream of TSS1 (CS2–4) and a pair located in a NEK6 intron (CS5) and near the PSMB7 promoter (CS6). The
Figure 2. The NEK6 Regulatory Hub

(A) Hi-C data for the NEK6 region in GM12878, as visualized in Juicebox (Rao et al., 2014). The intensity of each pixel represents relative normalized numbers of contact between corresponding regions, for which red and blue represent enriched or depleted interaction frequencies, respectively. Knight and Ruiz normalization (balanced) is applied to remove locus-specific biases. The observed over expected (O/E) signal is displayed to account for a higher number of interactions with closer regions due to one-dimensional proximity (Rao et al., 2014). Several chromatin structures and contact points are highlighted with black boxes. In the left panel, genes within the NEK6-TAD are colored red and remaining genes are colored blue.

(legend continued on next page)
convergent orientation favors loop formation between CTCF regions (Ghirlando and Felsenfeld, 2016; Rao et al., 2014), perhaps spatially sequestering the NEK6 regulome.

To determine whether this regulatory architecture is cell-type specific, we performed chromosome conformation capture (3C) assays in GM12878 (NEK6+) and Jurkat (NEK6−), which directly probes interactions between a given viewpoint and selected regions of the NEK6 chromosomal neighborhood. As shown in Figure 2B, a viewpoint spanning TSS1 interacts with upstream regulatory regions and with TSS2 at significantly higher frequencies in GM12878 compared with Jurkat. Peak TSS1 associations are with the CTCF cluster (CS2–4), CE1, CE2, and sites within SE1. To further validate the NEK6 element, we assayed a number of complementary viewpoints. Interactions with the distal CE1 element are significantly higher throughout the NEK6 sub-TAD in GM12878 compared with Jurkat. The enhanced CE1–PSMB7 contacts were confirmed using a PSMB7 promoter viewpoint (Figure S2A). Coupled with 3C assays using viewpoints in SE1 (Figures S2B and S2C), TSS2 (Figure S2D), and the CTCF cluster (Figure S2E), we conclude that the upstream region of NEK6 folds into a cell-type-specific regulatory conformation, forming a hub for enhancers, promoters, and CTCF sites, which likely drives higher levels of NEK6 expression in activated B cells.

Conventional Enhancers Augment NEK6 Expression in Transformed B Cells

Our ultimate goal is to test predictions for key components of the cis-regulatory circuit associated with elevated NEK6 expression in transformed B cells. Chromatin profiling and interactome analyses revealed over a dozen enhancer elements that could potentially augment NEK6 expression in FL. To prioritize functional analyses, we first measured enhancer activities for each candidate regulatory element in GM12878 and Jurkat (Figures 3A and S3A). In addition to the robust, GM12878-specific enhancer activity of CE1, four other elements augment luciferase expression from SV40-promoter-driven reporters. These include two regions in SE1 (CE5 and CE9), the CE10 region upstream of TSS1 and the CE13 region upstream of TSS2. Despite its significant levels of interaction with NEK6 promoters (Figure 2B), CE2 lacks enhancer activity in GM12878, which is consistent with minimal deposition of H3K27ac over this region (Figure 1B). The activity status of CEs was bolstered by ChiP-seq data from GM12878 (ENCOD Edwards Project Consortium, 2012), which reveals significant peaks for EP300 and TFs important in B cell biology, including EBF1, OCT2, PU.1, PAX5, RELA, and TCF3 (Figure S3B). In contrast, CE2 lacks significant binding by any of these factors. These functional data led us to first focus on the role of three CEs located outside of SE1, which had the most robust activities in GM12878 (CE1, CE10, and CE13).

To test the contributions of selected CEs to NEK6 expression, we individually deleted each enhancer from its endogenous site in GM12878 using CRISPR/Cas9 technology (Table S3). Deletion of CE13, which is proximal to TSS2, produces a modest but significant decrease in NEK6 expression when compared with subclones retaining the enhancer on both alleles (Figure 3B). Ablation of CE10 has no significant impact on NEK6 expression, despite its enhancer activity in luciferase assays. Importantly, NEK6 expression is attenuated substantially in subclones lacking the most distal enhancer, CE1, located 120 kb from TSS1. Consistently, NEK6 protein levels are dramatically reduced in CE1−/− subclones as measured by western blotting (Figure S3C). The effects of each enhancer deletion are indistinguishable for transcripts derived from either TSS1 or TSS2 (Figure S3D). Moreover, neither the CE1 nor the CE13 enhancer deletion impacts expression of neighboring LHX2 and PSMB7 genes (Figure S3E). These data suggest that CE1 and CE13 both contribute to augmented NEK6 expression in transformed B cells. Indeed, compound deletion of both elements further diminishes NEK6 mRNA and protein expression (Figures 3C and S3C). We conclude that two conventional enhancers, positioned outside of the large super-enhancer, additively potentiate NEK6 expression in GM12878.

To probe the effects of enhancer deletions on NEK6 chromatin and interaction landscapes, we analyzed subclones using ChIP and 3C, respectively. Deletion of CE13 reduces H3K27ac to near-background levels at an adjacent region, verifying removal of the core enhancer (Figure 3D). H3K27ac levels in CE1−/− mutants are unaffected at all other NEK6 enhancers tested. In sharp contrast, deletion of CE1 leads to significant reductions in H3K27ac not only at an adjacent region but also at many locations within SE1 and other enhancers that associate with CE1. These data suggest that CE1 is a dominant element in sculpting the active epigenetic landscape near NEK6, perhaps through spatial interactions that form its regulatory hub. In this regard, the TSS1 interactome is unaffected by deletion of either CE13 or CE1 (Figure 3E). Likewise, CE1 deletion does not alter long-range interactions between this region and downstream regulatory elements, including the TSS6 (Figure S3F). However, deletion of CE13 slightly boosts associations of CE1 with downstream enhancers, as well as NEK6 TSSs (Figure S3G). This finding suggests that CE13 may partially compete with CE1 for association with TSSs and other elements of the regulatory hub. When CE13 is deleted, there may be a compensatory increase in CE1 interactions.

To further test whether the dominant CE1 element is dispensable for maintaining the NEK6 interactome, we performed circular chromosome conformation capture sequencing (4C-seq) on GM12878, as well as three independent CE1−/− and two wild-type subclones. Genome-wide interactome data probed from CE1 and TSS1 viewpoints show that CE1 deletion subclones have no significant differences for interactions with regions between CS2 and downstream of TSS2 (Figures 3F and S3H), validating our 3C findings. These data indicate that maintenance of the NEK6 regulatory hub, which includes the distal CTCF cluster, CE1, SE1, CE13, and TSSs, is independent of the dominant NEK6 element.
conventional enhancer, CE1. However, this element contributes significantly to the maintenance of active chromatin marks at other CEs in the regulatory hub, boosting NEK6 expression in GM12878.

The NEK6 Super-Enhancer Is a Bystander

Super-enhancers are thought to be dominant regulatory elements for genes controlling cell identity, major cellular functions, and, in some cases, oncogenesis (Hnisz et al., 2013). Our chromatin analysis identified SE1, a 23-kb region located between CE1 and the TSSs, as a B cell-specific NEK6 super-enhancer. Although two conventional enhancers (CE1 and CE13) contribute to NEK6 expression, a substantial level of transcripts remains following their deletion, further implicating SE1 as an important regulatory element. To test this directly, we deleted the entire SE1 region from both alleles of GM12878 using CRISPR/Cas9. Surprisingly, multiple independent clones lacking SE1 consistently express NEK6 mRNA at modestly higher levels when compared with subclones retaining an SE1+/+ configuration (Figure 4A). Removal of SE1 also enhances or has minimal impact on NEK6 protein expression (Figure S3C). ChIP analysis revealed a depletion of H3K27ac neighboring the deleted SE1, confirming removal of the super-enhancer (Figure 4B). However, SE1 deletion does not impact H3K27ac levels at other tested CEs. Moreover, compund deletion of SE1 on one allele of CE1+/− clones has no significant impact on NEK6 expression (Figure 4A).

One potential explanation for enhanced NEK6 expression following SE1 removal is that CE1 resides 23 kb closer to its promoters. However, this would imply that SE1 itself does not contribute fundamentally to NEK6 expression. To explore the impact of SE1 on the NEK6 regulatory hub, we performed 3C. As shown in Figure 4C, SE1 deletion potentiates interactions between TSS1 and more distal elements (CE1 and CE2). The SE1+/− clones also show enhanced associations between TSS1 and more proximal regulatory regions (CE10 and TSS2), whose linear distances are unaffected by SE1 deletion. These data suggest that SE1 has a modest inhibitory impact on the frequency of enhancer associations in the NEK6 regulatory hub, as well as overall expression of this gene in GM12878.

An alternative explanation for the lack of SE1 regulatory function is that removal of critical enhancer elements drop NEK6 levels below a threshold required for GM12878 proliferation or survival. To test this possibility, we depleted NEK6 using several independent small hairpin RNAs (shRNAs). Reduced levels of NEK6 protein (20%–30% normal; Figures S4A and S4B) have no detectable impact on either proliferation or survival of GM12878 (Figures S4C and S4D). The lack of a biological phenotype may also stem from expression of NEK7 in these cells, a closely related kinase with significant functional overlap (Fry et al., 2012). These data indicate that selective pressure from reduced NEK6 levels cannot reasonably explain the lack of a significant expression phenotype in SE1-deficient cells.

Although SE1 is dispensable for NEK6 expression in GM12878, it remains possible that this broad regulatory region may target another gene in its chromosomal neighborhood. Focused qRT-PCR analysis of SE1+/− clones revealed no significant change in PSMB7 expression (Figure 4D). Similar to its effect on NEK6, SE1 deletion modestly enhances levels of LHX2 transcripts. To explore potential SE1 roles on a more global level, we analyzed three independent GM12878 subclones with SE1+/+ or SE1+/− genotypes using RNA sequencing (RNA-seq). SE1 deletion does not significantly change steady-state expression of any gene located within 5 Mb (Figure 4E). On the transcriptome level, six genes are significantly increased or decreased in SE1+/− clones compared with their wild-type counterparts (Figure S4E). The six genes are located on five different chromosomes; however, published promoter-capture Hi-C data reveal no significant inter-chromosomal interactions between any of the gene promoters and SE1 in GM12878 (Mifsud et al., 2015). We conclude that SE1, although clearly assigned as a super-enhancer using current algorithms, has no identifiable regulatory impact for maintaining expression of its nearest neighbors or any gene in a large chromosomal swath centered on NEK6.

A CTCF Cluster Establishes the NEK6 Contact Domain, but Not the Regulatory Hub

Our functional data clearly demonstrate that two conventional enhancers, CE1 and CE13, additively increase NEK6 expression in transformed B cells. The more distal of these two elements,
CE1, requires long-range looping (>120 kb) to communicate with NEK6 promoters. Architectural elements, largely consisting of CTCF sites, are common mediators of long-range looping that facilitate enhancer contact with gene promoters. Moreover, some CTCF sites serve as boundary elements to compartmentalize chromatin domains and inhibit inappropriate communication between enhancers and other neighboring genes (Ghirlando and Felsenfeld, 2016). CE1 is flanked by a cluster of CTCF sites positioned at one border of a robust contact domain containing NEK6. All three sites in this cluster are oriented convergently with a pair of downstream CTCF sites, located in a NEK6 intron (CS5) and near the PSMB7 promoter (CS6). The convergent orientation favors intermolecular CTCF interactions, which could form loops to cordon off NEK6-associated enhancers from other genes in...

Figure 4. SE1 Is a Dispensable Element in the NEK6 Regulome
(A) NEK6 transcripts measured by qRT-PCR of SE1 deletion subclones. Each dot represents a unique subclone, which is reported as the average of two independent experiments. See Figures 3B and 3C for details. For (A)–(D), statistical significance was determined using an unpaired t test with Welch’s correction (*p < 0.05).
(B) H3K27ac ChIP assays in SE1 deletion subclones. See Figure 3D for details. For (B) and (C), each bar represents the mean ± SEM of two subclones, each of which includes two independent experiments.
(C) Interaction frequencies, as measured by 3C-qPCR, in SE1 deletion subclones for NEK6 TSS1 (left) and CE1 (right) viewpoints.
(D) LHX2 and PSMB7 transcripts measured by qRT-PCR in SE1 deletion subclones. Each dot represents a unique subclone, which is reported as the average of two independent experiments.
(E) Expression profile for all genes located within 5 Mb of SE1, as measured by RNA-seq, in SE1 wild-type and deletion subclones of GM12878. Average logCPM indicates the average expression level of each gene among three wild-type and three deletion subclones, reported as log2 read counts per million mapped reads. Log(SE1 Del/WT) represents the log2 fold-change of each gene between the average CPM of deletion versus wild-type subclones. Blue lines denote 2-fold differences.
See also Figure S4.
the TAD. To explore architectural logic in the NEK6 cis-regulatory circuit, we deleted a region spanning all three sites in the upstream CTCF cluster (CS2–4). Minimal CTCF binding is detected at sites flanking CS2–4 following its deletion when compared with wild-type loci (Figure 5A), whereas CTCF ChIP signals are unaffacted at CS5 and CS6. NEK6 expression is reduced ~20% in subclones harboring the CS2–4 deletion on both alleles (Figure 5B). In contrast, LHX2 expression is enhanced ~60% in knockout subclones, while expression of the two other genes in this TAD, DENND1A and PSMB7, remains unchanged.

These data suggest that CS2–4 serves as a boundary element to prevent the spread of active chromatin from NEK6 to LHX2, minimize long-range interactions between NEK6 enhancers and LHX2, or both (Ghirlando and Felsenfeld, 2016; Ong and Corces, 2014). To test the first possibility, we measured H3K27ac densities at sites in the NEK6 contact domain and adjacent LHX2 regions (Figure 5C). Consistent with a role for CS2–4 as a chromatin boundary, its deletion permits H3K27ac spreading upstream of CE1 into the LHX2 locus. CS2–4 deletion had an opposite effect on H3K27ac densities within the NEK6 contact domain, which are significantly reduced, accompanied by an increase in the H3K27me3 modification (Figure S5A). Thus, perturbed patterns of chromatin modifications correlate well with altered gene expression upon deletion of the 5’ CTCF cluster, supporting its functional assignment as a boundary element.

To determine whether CS2–4 also serves as a spatial boundary, precluding communication between NEK6 enhancers and other promoters, we performed 3C on subclones with wild-type and CS2–4–/– genotypes. As expected, mutant subclones generate no 3C signal for interactions between TSS1 and the deleted CS3 region (Figure 5D). All other interactions between TSS1 and NEK6 regulatory elements are unaffected by CS2–4 deletion. In contrast, TSS1 interactions with the LHX2 and DENND1A promoters, located further upstream in the sub-TAD, are significantly increased in mutant subclones. A similar enhancement of upstream interactions is observed for the CE1 element with LHX2, but not DENND1A, which correlates with the differential impacts of CS2–4 deletion on expression levels. Conversely, CE1 associations are decreased with downstream regions, including CS5 and the PSMB7 promoter. The enhanced interactions with LHX2 were confirmed using a complementary viewpoint corresponding to its promoter (Figure 5E).

To support these findings, we performed 4C-seq on GM12878, as well as independent CS2–4–/– and wild-type subclones (Figures 5F, S5B, and SSC). Genome-wide interactome data probed from TSS1 and CE1 viewpoints reveal that, in general, CS2–4–/– subclones have more robust associations with upstream regions in the sub-TAD, reaching to the DENND1A promoter, as reflected in percent total normalized reads (Figure 5F) (Guo et al., 2015). In contrast, interactions within the NEK6 contact domain itself are slightly attenuated following CS2–4 deletion (diminished percent normalized reads in Figures 5F and S5B). In addition, 4C-seq data identify several interactions that differ significantly between CS2–4–/– and control clones. Deletion of the CTCF cluster significantly augments interactions between CE1 and several regions upstream (Figure 5F, green asterisks), as well as with the LHX2 promoter, although the latter does not attain statistical significance in 4C data. Conversely, multiple interactions of CE1 with downstream regions in the NEK6 gene body and PSMB7 promoter region are significantly diminished following CS2–4 removal (Figure 5F, red asterisks), consistent with our 3C data (Figure 5D). Similarly, upon CS2–4 deletion, TSS1 has significantly elevated associations with the DENND1A and LHX2 promoters (Figure 5F).

A potential explanation for the latter finding is that new contact loops may be formed between NEK6-proximal CTCF sites (e.g., CS5) and the properly oriented CTCF site upstream of the deleted CS2–4 region. A CTCF site located between the DENND1A promoter and LHX2, designated as CS0, has the same orientation as those deleted from the CS2–4 cluster (Figure 5F). Indeed, 3C analyses indicate that the CS2–4 deletion enhances CS0-CS5 interactions, whereas CS0-CS6 crosslinking remains unaffected (Figure 5G). The architectural remodeling of CTCF interactions, which may place the NEK6 gene in closer proximity to LHX2 and DENND1A, was confirmed using the complementary CS5 viewpoint (Figure 5G). Together, these data indicate that CS2–4 contributes modestly to establishing the regulatory hub between NEK6 promoters and enhancers. Instead, this CTCF cluster predominantly functions as a chromatin and
architectural boundary, minimizing the impact of the NEK6 regulatory hub on neighboring genes in its TAD.

**DISCUSSION**

Developmental and cell-type-specific regulation of genes is orchestrated by changes in TF expression, enhancer activation, and alterations in chromatin landscapes, including architecture. Deciphering the contributions of each process to gene regulation is especially important given that a vast majority of disease-associated changes in the genome affect expression levels rather than coding potentials (Maurano et al., 2012). A prerequisite for understanding cis-regulatory circuits that govern normal or pathogenic gene expression is the profiling of enhancers and their contacts in distinct cell types. This milestone has largely been achieved in several hematologic malignancies and normal cellular counterparts (Chapuy et al., 2013; Koues et al., 2015). Based on chromatin and architectural profiles, pattern-based algorithms have been used to predict key regulatory connections between enhancers and their target genes. However, there is a critical need to test predicted circuits using reductionist, genetic approaches.

In this study, we dissected cis-regulatory circuits within a chromosomal neighborhood spanning at least three genes over-expressed in human BCL. Importantly, many predictions from pattern-based algorithms for NEK6 were not substantiated when tested directly. The predicted circuitry for pathogenic NEK6 expression involved at least a dozen enhancers with augmented H3K27ac loads in FL versus normal B cells. All of the CEs, including those comprising a super-enhancer, directly contact the NEK6 promoter in transformed B cells, further strengthening their predicted contributions to its elevated expression in BCL. Instead, we find that the NEK6 regulome is dominated by two conventional enhancers: one located near the TSSs (CE13) and a second, more powerful enhancer (CE1) located ~100 kb upstream. Although some of the predicted enhancers for NEK6 bind an overlapping set of factors, CE1 exhibits higher loads of TF binding than other enhancers (Figure S3B), potentially explaining its dominant regulatory function. CE13 has lower levels of bound TFs and enhancer activity in luciferase assays, yet its proximity to TSSs may elevate its role in NEK6 regulation. The remaining CEs and, surprisingly, the super-enhancer, are all dispensable for NEK6 expression in transformed B cells, despite correlative changes in epigenetic and architectural landscapes. Thus, our study underscores the pressing need to hone predicted circuitry through rigorous testing. Although tedious, the emergence of high-throughput methods for genetic dissection of TFs, enhancers, and chromosome architecture will speed achievement of this goal.

We suspect several potential reasons for disconnects between predictive algorithms and direct validation of cis-regulatory circuits. First, as shown here for NEK6, a dominant enhancer can affect the chromatin profile of other regulatory elements in its interactome. Deletion of CE1 attenuated H3K27ac loads on other CEs spread throughout the NEK6 region. Thus, increased CE1 activity in BCL likely augments H3K27ac on other elements in the regulatory hub, even if they do not contribute substantially to enhanced gene expression. Second, we cannot rule out that some CEs function as “backup” elements to partially sustain NEK6 expression if CE1 activity is destroyed. This may be true for CE13, which contributes modestly to NEK6 expression in the absence of CE1. However, SE1 does not appear to have such a backup role, since deletion of the entire region, or its composite CEs (data not shown), has no significant effect on NEK6 expression, whether CE1 is present or not.

The most surprising and significant finding from our study is that a clearly established super-enhancer has no discernable impact on the expression of NEK6 or any other gene on its chromosome. This finding is especially notable given the building dogma that SEs are a collection of key elements controlling high-level expression of genes critical for cell identity and function, as well as oncogenesis (Lovén et al., 2013). This finding not only underscores the need for functional evaluation of SEs in many cell types but also brings to light a third potential explanation for discrepancies between predicted and validated cis-regulatory circuits. Although the super-enhancer and a subset of other CEs are dispensable for NEK6 expression, these elements may be required earlier in B cell development or transformation to initially activate or augment transcription of this kinase gene. After these key activation events, SE1 or other CEs may become dispensable, with CE1 primarily maintaining elevated levels of NEK6 expression. These issues are currently intractable in primary human B cells but may be approached in future studies by deletion of analogous regulatory regions for mouse NEK6. Notwithstanding, our findings indicate that at least a subset of SEs associated with oncogenesis would not be priority targets for current epigenetic-based therapeutic strategies to squelch expression of associated genes (Lovén et al., 2013).

A second surprise to emerge from our studies concerned determinants for regulatory architecture of the NEK6 chromosomal neighborhood. We found that most enhancers in this region converge spatially to form a regulatory hub with NEK6 promoters and flanking CTCF clusters. Although CE1 is the dominant NEK6 enhancer, its deletion does not significantly affect maintenance of the regulatory hub. Likewise, deletion of CS2–4 has only a modest impact on spatial interactions within this hub. These findings suggest several intriguing possibilities for architectural determinants of regulatory hubs, which await future dissection, including (1) direct CE1-promoter interactions are redundant, structurally, with CS2–4 looping to downstream CTCF sites; (2) another element, excluding SE1 and CE1, is the key determinant for initiating regulatory hub formation; or (3) once the NEK6 sub-TAD is decorated with active histone modifications, homotypic chromatin interactions drive close association of the promoter with regional enhancers (Lieberman-Aiden et al., 2009). Nevertheless, our study identifies important dual roles for CS2–4 as a chromatin and architectural boundary, impairing the spread of active chromatin and enhancer interactions upstream of NEK6 into LHX2. Thus, many CTCF sites or clusters predicted to be important for formation of architectural loops may be more critical in establishing or maintaining borders of regulatory domains.

Our findings will also inform future studies to determine how NEK6 contributes to B lymphomagenesis. Despite consistent overexpression of the mitosis-associated kinase in BCL, NEK6 depletion had no detectable impact on viability or proliferation...
of transformed human B cells, including complete NEK6 knockout in two BCL lines (data not shown). In contrast, NEK6 knockout in other cancer models significantly attenuated cell growth (Fry et al., 2012). We suspect that, in BCL, partial functional overlap with the closely related kinase NEK7 may explain the lack of cellular phenotype. Indeed, NEK7 is overexpressed in primary cells derived from BCL biopsy specimens compared with their normal counterparts (Koues et al., 2015). Human NEK6 and NEK7 loci appear to be partial duplicates of one another, since both are flanked upstream by additional LHX and DENND genes. However, unlike NEK6, the NEK7 locus is devoid of chromatin hallmarks for active distal enhancers in B lymphocytes, FL, or other cell types (ENCODE Project Consortium, 2012; Koues et al., 2015). These correlative data suggest that NEK family kinases are essential components of the program for lymphomagenesis, requiring transformed B cells to augment NEK6 as a complement, or a backup, to NEK7 overexpression, or vice versa. Thus, our dissection of the NEK6 regulome will be an important starting point to test such requirements in the GC program and oncogenic conversion to BCL.

EXPERIMENTAL PROCEDURES

Details for prioritization, 3C, 4C-seq, RT-PCR, RNA-seq, western blotting, and NEK6 knockdown experiments are in Supplemental Experimental Procedures.

Patient Samples

All human samples were obtained under IRB-approved protocols as previously described (Koues et al., 2015).

Luciferase Assay

Candidate enhancers (~800 bp) were PCR amplified (Table S4) and cloned into SV40 promoter-driven pGL3 plasmid (Promega). Reporters were transfected into GM12878 and Farage (Roche 06366236001) or electroporated into Jurkat.

Super Enhancer Calling

H3K27ac ChIP-seq data for primary B cells (Koues et al., 2015) and GM12878 (ENCODE Project Consortium, 2012) were aligned to the reference human genome (hg19) with Bowtie2 (Langmead and Salzberg, 2012). Peaks were called using model-based analysis of ChIP-seq (MACS), and SEs were called using ROSE under default settings (Loven et al., 2013; Whyte et al., 2013).

3C and 4C-Seq

3C and 4C-seq assays were performed as described previously (Hagège et al., 2007; Majumder et al., 2015; Splinter et al., 2012) using strategies detailed in Supplemental Experimental Procedures. Primers and probes are shown in Table S4. 4C-seq statistics are shown in Table S5.

Chromatin Immunoprecipitation

ChIP assays were performed as described previously (Koues et al., 2015) using the following antibodies: 1 μg anti-H3K27ac (ab4729), 1 μg anti-H3K27me3 (ab6002), 8 μL anti-CTCF (Cell Signaling Technology 2899), and anti-rabbit immunoglobulin G (IgG) (sc2027). ChIP DNA was analyzed with SYBR qPCR primer pairs outside and inside of the gRNA target sites. gRNA sequences are shown in Table S3. Most gRNAs were cloned into the Addgene vector pKLV-U6gRNA(BbsI)-PGKpuro2ABFP (Addgene 50946), PCR primers for screening deletions are provided in Table S3. PCR products spanning deletion sites were purified and Sanger sequenced (Table S3). All molecular analyses were performed on sibling subclones corresponding to parental and mutant genotypes in the same experiment to avoid complications that might arise from drifts in bulk GM12878 cultures and experimental variations.

ACCESSION NUMBERS

The accession number for the raw reads and processed files for the RNA-seq and 4C-seq datasets reported in this paper is GEO: GSE87323.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.02.067.

AUTHOR CONTRIBUTIONS

E.M.O., J.E.P., Y.H. and O.I.K. conceptualized the study and designed experiments. E.M.O. supervised all aspects of the project. Experiments and data analysis were performed by all authors. The manuscript was written by E.M.O. and Y.H. with input from J.E.P.

ACKNOWLEDGMENTS

This work was supported by NIH grants CA156690, CA188286 (J.E.P. and E.M.O.), AI122726, and AI115734 (E.M.O.), as well as WU-ICTS (TR000448) and Siteman Cancer Center (CA18142) grants. We thank Patrick Collins for helpful comments and the Genome Technology Access Center for assistance with -omics analyses.

Received: September 11, 2016
Revised: January 13, 2017
Accepted: February 21, 2017
Published: March 21, 2017

REFERENCES


Supplemental Information

cis-Regulatory Circuits Regulating NEK6 Kinase Overexpression in Transformed B Cells Are Super-Enhancer Independent

Yue Huang, Olivia I. Koues, Jiang-yang Zhao, Regina Liu, Sarah C. Pyfrom, Jacqueline E. Payton, and Eugene M. Oltz
**Fig S1**

**A**

All predicted gene-enhancer pairs

≤ 1,607,099

Recurrence in FL

≤ 112,878

Robust levels of gene expression and enhancer marks

4,784

Gene function

≤ 2,021

TF binding

≤ 1,997

Ranking and manual selection

Seven prioritized circuits

**B**

Luminescence fold change

Promoter only

IGH enhancer

NEK6

RUVBL1

RGS13

EEPD1

CASC4

EIF3I

KIAA1199

GM12878

Farage

**C**

RNA

GM

actB

TSS1

TSS2

**D**

DENND1A

LHX2

CS1

CS2-4

CE1

SE1(3-9)

NEK6

PSMB7

**E**

Relative to ACTB

GM

Jurkat

DENND1A

LHX2

PSMB7
Figure S1 (related to Figure 1). Prioritization scheme, luciferase assays of putative enhancers, and regulatory landscape of NEK6 in distinct cell types

(A) A flowchart showing how seven regulatory circuits were selected from genome-wide predicted gene-enhancer pairs. See Supplemental Experimental Procedures for details.

(B) Luciferase reporter assays for DREs connected to seven high priority genes with elevated expression in FL. Enhancer activities were measured transiently in GM12878 and Farage B cell lines and are reported relative to a control reporter construct containing only the SV40 promoter. The potent B cell enhancer associated with the human IGH locus is included as a positive control. Results represent the mean ± SEM of at least two independent experiments. ND: Not done.

(C) UCSC Genome Browser views of annotated NEK6 transcript isoforms and RNA-seq data from GM12878 (ENCODEN) or in vitro activated B cells (Koues et al. 2015). The two active TSSs for NEK6 in B cells are indicated. RNA data are presented as the number of aligned, in silico extended reads per 10 bp.

(D) UCSC Genome Browser views of H3K27ac ChIP-seq data from FL, CC and other distinct ENCODE cell types.

(E) Transcript abundance of NEK6 neighboring genes measured by RT-qPCR in GM12878 and Jurkat cells. Results represent the mean ± SEM of three independent experiments. Statistical significance (unpaired t-test with Welch's correction): *p<0.05, and ****p<0.001.
Fig S2

A

Relative crosslinking

LHX2

CS3

CE1

SE1

TSS1

PSMB7

B

Relative crosslinking

LHX2

CS3

CE1

SE1(CE4-6)

C

Relative crosslinking

LHX2

CS3

CE1

SE1(CE9)

D

Relative crosslinking

LHX2

CS3

CE1

TSS2

E

Relative crosslinking

LHX2

CS3
Figure S2 (related to Figure 2). Interaction frequencies of five additional viewpoints within the NEK6 sub-TAD

(A-E) Interaction frequencies measured by 3C-qPCR in GM12878 (NEK6 expressing) and Jurkat (NEK6 silent) for the indicated viewpoints: PSMB7 promoter (A), SE1 (CE4-6) (B), SE1 (CE9) (C), NEK6 TSS2 (D), and CS3 (E). Results represent the mean ± SEM of two independent experiments. Statistical significance (unpaired t-test with Welch's correction): *p<0.05.
Figure S3 (related to Figure 3). Luciferase assays, TF binding, expression and interaction analyses of CEs

(A) Luciferase reporter assays for control constructs. Enhancer activities were measured transiently in GM12878 or Jurkat cells and calculated relative to an SV40 promoter-only reporter construct. The human IGH enhancer and mouse Tcrb enhancer were included as positive controls. Results show the mean ± SEM of two independent experiments.

(B) Approximate relative TF binding intensities for six NEK6 CEs, derived from ChIP-seq data for TFs important in B cell biology in GM12878 (ENCODE).

(C) Immunoblots probed with antibodies specific for NEK6 or GAPDH, in different GM12878-derived CRISPR deletion subclones with the indicated genotypes, including parental wild-type cells (WT) and NEK6 knockout subclones (NEK6<sup>−/−</sup>). Normalized NEK6 protein levels relative to WT, as measured by ImageJ, are indicated at the bottom.

(D) NEK6 transcripts derived from the two TSSs, as measured by RT-qPCR, in deletion subclones of CE13 (top), CE10 (middle) and CE1 (bottom). For panels D and E, each dot represents an independent subclone, which is reported as the average of two independent experiments. See Fig. 3B and C for details. Statistical significance (unpaired t-test with Welch's correction): **p<0.01, ***p<0.005, ****p<0.001.

(E) LHX2 and PSMB7 transcripts measured by RT-qPCR in CE13 (top) and CE1 (bottom) deletion subclones.

(F and G) Interaction frequencies, as measured by 3C-qPCR, in deletion subclones of CE1 (F) and CE13 (G) for CE1 viewpoint. Each bar represents the mean ± SEM of two independent subclones, each of which includes two independent experiments. Statistical significance (unpaired t-test with Welch's correction): *p<0.05.

(H) Interaction profiles, as measured by 4C-seq, for CE1 wild-type and deletion samples using CE1 and NEK6-TSS1 as anchors. For each viewpoint, reads per HindIII fragment normalized by DESeq2 are shown for three wild-type (blue), and three CE1 deletion lines (orange). The deleted CE1 region is shown as a yellow rectangle. Spearman’s rank correlation coefficients, as shown on the right of sample names, are calculated for each pair of samples of the same genotype, and are all significant (asymptotic t approximation, p<2.2 x 10<sup>-16</sup>).
Figure S4 (related to Figure 4). NEK6 knockdowns in GM12878 and global transcription profiles in SE1 deletion subclones

(A) Immunoblots probed with antibodies specific for NEK6 or GAPDH, in GM12878 cells transduced with shRNAs targeting GFP (control) or different regions of NEK6 transcripts and purified at 72h. Normalized NEK6 protein levels relative to shGFP are indicated at the bottom.

(B) NEK6 transcripts, as measured by RT-qPCR, in GM12878 cells transduced with either GFP- or NEK6-specific shRNAs and purified at 72h. Statistical significance (unpaired t-test with Welch's correction): **p<0.01.

(C) Proliferation rates, as measured by CSFE dilution (flow cytometry), in GM12878 cells transduced with either GFP- or NEK6-specific shRNAs and analyzed from 72h, which is labeled as d0. Median fluorescence intensities of CFSE are shown on the Y-axis.

(D) Cell death, as measured by Annexin V staining (flow cytometry), in GM12878 cells transduced with either GFP- or NEK6-specific shRNAs and analyzed at 72h. Statistical significance (unpaired t-test with Welch's correction): p<0.05.

(E) Global transcription profiles, as measured by RNA-seq, in SE1 wild-type and deletion subclones. Average logCPM indicates the average expression level of each gene among three wild-type and three deletion subclones, reported as log2 read counts per million mapped reads. Log(SE1 Del/WT) represents the log2 fold change of each gene between the average CPM of deletion subclones versus wild-type subclones. Statistical significance is generated using generalized linear model with p-values adjusted by Benjamini-Hochberg procedure. Six genes with p<0.05 are labeled with red color. Blue lines denote two-fold differences.
Figure S5 (related to Figure 5). H3K27me3 ChIP assays and interaction profiles in C2-4 deletion subclones

(A) ChIP-DNAs were analyzed by qPCR using primers near the indicated CEs. Each bar represents the mean ± SEM of two independent subclones, each of which includes two independent experiments. Statistical significance (unpaired t-test with Welch's correction): *p<0.05. ChIP assays with a non-specific IgG antibody are shown as controls.

(B) Zoomed-in UCSC Genome Browser views of interaction profiles, as measured by 4C-seq, for CS2-4 wild-type and deletion subclones using CE1 and NEK6-TSS1 as anchors. For each viewpoint, the average reads per HindIII fragment normalized by DESeq2 are shown for three wild-type (red), and three CS2-4 deletion lines (green). Reads located within the deleted CS2-4 region are removed from all samples. Also shown is a plot for differential signal between deletion and wild-type samples in natural log scale, ln (Del-WT). Statistical significance (generalized linear model adjusted by Benjamini-Hochberg procedure): p<0.05, are denoted by green or red asterisks for interactions that are increased or decreased in CS2-4 mutants, respectively.

(C) Interaction profiles, as measured by 4C-seq, for CS2-4 wild-type and deletion samples using CE1 and NEK6-TSS1 as anchors. For each viewpoint, reads per HindIII fragment normalized by DESeq2 are shown for three wild-type (blue), and three CE1 deletion lines (orange). The deleted CS2-4 region is shown as a yellow rectangle. Spearman’s rank correlation coefficients, as shown on the right of sample names, are calculated for each pair of samples of the same genotype, and are all significant (asymptotic t approximation, p<2.2 x 10^{-16}).
Table S1 (related to Figure 1, provided in .xlsx format). A manually curated list of genes used for prioritization

Table S2 (related to Figure 1, provided in .xlsx format). Predicted circuits of seven prioritized regions

Table S3 (related to Experimental Procedures, provided in .xlsx format). Sanger Sequencing results of CRISPR deletion subclones, and primers and sequences related to CRISPR experiments

Top: Sequences flanking the deletion junctions are shown for a reference wild-type allele and sequenced deleted alleles. gRNA target sequences are colored green, genomic DNA sequences between two gRNA target sites are shown as “--” for clarity, “(--)” denote that this region is deleted and “**” represent extra deleted nucleotides beyond this region in sequenced homozygous subclones.

Middle: gRNA target sites, sequences and cloning oligos are shown.

Bottom: For each targeted region for deletion, combinations of two gRNA constructs and multiple genotyping primer pairs are listed.

Table S4 (related to Experimental Procedures, provided in .xlsx format). Primers used for molecular assays

Primers used to clone CEs into luciferase reporter constructs: Primer sequences include added nucleotides of BamHI or SalI restriction sites for cloning (lowercase), and genomic DNA sequences (uppercase).

Primers used for 3C and 4C assays

Primers used for RT-qPCR assays

Primers used for ChIP-qPCR assays
<table>
<thead>
<tr>
<th>Sample name</th>
<th># of total mapped reads (reads located within two restriction fragments of the viewpoint are removed)</th>
<th>Fraction of total mapped reads located in the cis chromosome (chr9)</th>
<th>Fraction of total mapped reads located in the NEK6 sub-TAD (chr9:126,130,000-127,200,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE1viewpoint_GM12878</td>
<td>3,299,896</td>
<td>61%</td>
<td>41%</td>
</tr>
<tr>
<td>CE1viewpoint_CE1_WT_clone1</td>
<td>4,712,155</td>
<td>58%</td>
<td>38%</td>
</tr>
<tr>
<td>CE1viewpoint_CE1_WT_clone2</td>
<td>4,708,887</td>
<td>57%</td>
<td>39%</td>
</tr>
<tr>
<td>CE1viewpoint_CE1_Del_clone1</td>
<td>3,923,007</td>
<td>52%</td>
<td>31%</td>
</tr>
<tr>
<td>CE1viewpoint_CE1_Del_clone2</td>
<td>4,000,761</td>
<td>57%</td>
<td>37%</td>
</tr>
<tr>
<td>CE1viewpoint_CE1_Del_clone3</td>
<td>4,121,564</td>
<td>64%</td>
<td>44%</td>
</tr>
<tr>
<td>CE1viewpoint_CS2-4_WT_clone1</td>
<td>3,235,282</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>CE1viewpoint_CS2-4_WT_clone2</td>
<td>5,606,882</td>
<td>59%</td>
<td>35%</td>
</tr>
<tr>
<td>CE1viewpoint_CS2-4_Del_clone1</td>
<td>4,088,825</td>
<td>64%</td>
<td>47%</td>
</tr>
<tr>
<td>CE1viewpoint_CS2-4_Del_clone2</td>
<td>5,208,257</td>
<td>63%</td>
<td>45%</td>
</tr>
<tr>
<td>CE1viewpoint_CS2-4_Del_clone3</td>
<td>5,819,853</td>
<td>67%</td>
<td>48%</td>
</tr>
<tr>
<td>TSS1viewpoint_GM12878</td>
<td>3,650,318</td>
<td>61%</td>
<td>40%</td>
</tr>
<tr>
<td>TSS1viewpoint_CE1_WT_clone1</td>
<td>5,801,414</td>
<td>60%</td>
<td>39%</td>
</tr>
<tr>
<td>TSS1viewpoint_CE1_WT_clone2</td>
<td>6,220,468</td>
<td>61%</td>
<td>45%</td>
</tr>
<tr>
<td>TSS1viewpoint_CE1_Del_clone1</td>
<td>5,300,021</td>
<td>55%</td>
<td>37%</td>
</tr>
<tr>
<td>TSS1viewpoint_CE1_Del_clone2</td>
<td>5,587,267</td>
<td>63%</td>
<td>43%</td>
</tr>
<tr>
<td>TSS1viewpoint_CE1_Del_clone3</td>
<td>5,307,205</td>
<td>71%</td>
<td>55%</td>
</tr>
<tr>
<td>TSS1viewpoint_CS2-4_WT_clone1</td>
<td>3,824,313</td>
<td>63%</td>
<td>43%</td>
</tr>
<tr>
<td>TSS1viewpoint_CS2-4_WT_clone2</td>
<td>2,733,325</td>
<td>64%</td>
<td>39%</td>
</tr>
<tr>
<td>TSS1viewpoint_CS2-4_Del_clone1</td>
<td>4,785,052</td>
<td>60%</td>
<td>38%</td>
</tr>
<tr>
<td>TSS1viewpoint_CS2-4_Del_clone2</td>
<td>3,281,914</td>
<td>68%</td>
<td>52%</td>
</tr>
<tr>
<td>TSS1viewpoint_CS2-4_Del_clone3</td>
<td>2,793,105</td>
<td>66%</td>
<td>49%</td>
</tr>
<tr>
<td>TSS1viewpoint_CS2-4_Del_clone3</td>
<td>4,328,501</td>
<td>74%</td>
<td>57%</td>
</tr>
</tbody>
</table>

Table S5 (related to Experimental Procedures). 4C-seq statistics
SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Prioritization Scheme for cis-Regulatory Circuits in FL. We prioritized FL circuits as follows. First, we selected circuits with at least one gene-enhancer pair that was recurrently augmented in more than 6/15 FL samples (Koues et al., 2015): gene expression FL/average CC>1 (quantified by microarray analysis), enhancer histone marks FL/average CC>1.5 (quantified by H3ac, H3K27ac or FAIRE-seq). Second, predicted target gene(s) and relevant enhancers were required to exhibit robust levels of RNA expression (normalized microarray signal > 120) and histone marks (H3K27ac, H3ac and H3K4me1 ChIP-seq RPM > 100) in FL. Third, the surviving list of enhancer-gene combinations was intersected with a manually curated list of ~7000 genes that have been implicated in general oncogenesis, immune modulation, or chromatin modification (Table S1). Finally, the remaining genetic loci were examined for binding of TFs known to be important for B cell function (EBF1, PU.1, IRF4, IKZF1, POU2F2, PAX5, MEF2A, MEF2C, RUNX3, RELA, TCF3, TCF12, YY1, MAX, STAT1, STAT3, STAT5A, SP1), or the enhancer-associated acetyltransferase, EP300, using public ChIP-seq data for the transformed B cell line, GM12878, in UCSC Genome Browser (ENCODE Project Consortium, 2012). The remaining list of ~2000 gene-enhancer pairs were ranked based on levels of RNA expression and histone marks, recurrence in FL samples, as well as concordance between expression and histone modifications at putative enhancers. Manual inspection of the top ~200 highest ranked enhancer-promoter pairs yielded seven genetic loci that we considered to be of highest priority. See also Figure S1A.

3C. In brief, 10^7 cells were crosslinked with 1% formaldehyde, quenched with glycine, lysed, digested with HindIII, religated, and purified with phenol-chloroform followed by Qiagen PCR purification columns. Interactions were measured using a Taqman qPCR assay for ligation products between each anchor HindIII fragment and each target HindIII fragment. Interaction frequencies were normalized for signals obtained from nearest neighbor fragments in the EEF1G gene. Standard curves were generated using HindIII digested and religated bacterial artificial chromosomes (RP11-1123P20, RP11-15B9, RP11-902D21 and RP11-259I15 for NEK6, RP11-993C15 for EEF1G). Amplicons with extreme Ct values in standard curves were either discarded or analyzed using delta Ct values. Statistical analysis was performed using Prism.

4C-Seq. In brief, 3C DNAs were digested with a second restriction enzyme, DpnII, religated, and purified using Qiagen PCR purification columns. The circularized DNA was amplified using inverse PCR and nested inverse PCR reactions with primers in the anchor HindIII-DpnII fragment. PCR products were used to prepare indexed sequencing libraries. All twelve libraries were pooled in one lane for 50 bp single-end deep sequencing (Illumina HiSeq2500). Reads were aligned to the reference human genome (build hg19) with Bowtie2 2.2.9 (Langmead and Salzberg, 2012). Reads for each HindIII fragment were calculated using r3Cseq 1.18.0 (Thongjuea et al., 2013) and normalized using DESeq2 1.14.1 (Love et al., 2014). Statistical analysis for differential interactions between genotypes were performed using DESeq2. Spearman correlation of each genotype was performed using R.

RT-PCR and RNA-Seq. For RT-PCR, total RNA was extracted using TRIzol (Invitrogen), reverse-transcribed (MuLV reverse transcriptase, New England Biolabs). SYBR qPCR was carried out using primers in Table S4. Statistical analysis was performed using Prism. For RNA-seq, total RNA was extracted (RNeasy, Qiagen). Poly (A) mRNA was purified (Dynabeads mRNA Direct, Thermo Fisher Scientific), reverse-transcribed, and used for preparation of indexed libraries. All six libraries were pooled in one lane for 50 bp single-end deep sequencing (Illumina HiSeq2500). RNA-seq reads were aligned to the reference human genome (Ensembl 76) with STAR 2.0.4b (Dobin et al., 2013). Gene counts were derived by Subread:featureCounts 1.4.5 (Liao et al., 2014). Statistical analysis was performed using edgeR 3.14.0 (Robinson et al., 2010).

Western Blotting. Western blotting was performed using standard protocols with the following antibodies: NEK6 (ab133494), GAPDH (sc365062).

NEK6 Knockdown. GM12878 cells were transduced with retroviral vectors containing shRNAs specific for either GFP (target sequence: AGCACAAGCTGGAGTACAACTA) or NEK6 (target sequences 1, 2, and 3: CGGCCAAGTGTCAACAGGCAA, AGGAGAGGACAGTAGGGAAGTA, AGCAGATGACTCAAGTTTAA) and an hCD2 marker as previously described (Bednarski et al., 2012). Transduced cells were subjected to the following assays. Cell death was quantified by Annexin V (BD Biosciences 556422) and hCD2 (BD Biosciences 560642) double staining 72 h post-transduction. Cell proliferation was...
measured by CFSE dilution (Life technologies C34554), staining cells with CFSE 48 h post-transduction, then with anti-hCD2 at 72, 96 and 120 h post-transduction. Knockdown efficiencies were assessed for hCD2⁺ cells purified 72 h post transfection using magnetic beads (Miltenyi Biotec 130-091-114) by western blotting.
SUPPLEMENTAL REFERENCES


