Supplemental Information

cis-Regulatory Circuits Regulating \textit{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

ND

C

RNA

GM

actB

TSS1

TSS2

D

DENND1A

LHX2

CS1

CS2-4

CE1

SE1(3-9)

NEK6

PSMB7

FL

CC

HepG2

HCT-116

NHLF

HUVEC

E

DENND1A

LHX2

PSMB7

Relative to ACTB

GM

Jurkat

Relative to ACTB

GM

Jurkat

Relative to ACTB

GM

Jurkat
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 (ENCODEx) 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

B

C

D

E
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−/−). 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, and ***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^{-16}).
Fig S4

A

<table>
<thead>
<tr>
<th>shGFP</th>
<th>shNEK6-1</th>
<th>shNEK6-2</th>
<th>shNEK6-3</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEK6</td>
<td>GAPDH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>1</td>
</tr>
</tbody>
</table>

B

Relative to ACTB

C

CFSE MFI

d0 d1 d2

D

% Dead cells (Annexin V)

E

Significance
- Not Significant
- p < 0.05
**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 sh*GFP* 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 Sall 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 glycerine, 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 (M-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: AGCACAAGCTGGAGTACAACA) or NEK6 (target sequences 1, 2, and 3: CGGGCCAGAGTGTCCAGGGCAA, AGGAGAGGACAGTATGGAAGTA, AGGAGAGGACAGTATGGAAGTA, and 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


