Cell Reports, Volume *29*

Supplemental Information

Global Analysis of Enhancer Targets Reveals Convergent Enhancer-Driven Regulatory Modules

Shiqi Xie, Daniel Armendariz, Pei Zhou, Jialei Duan, and Gary C. Hon

Figure S1. Mosaic-seq reveals the primary hits of enhancers.

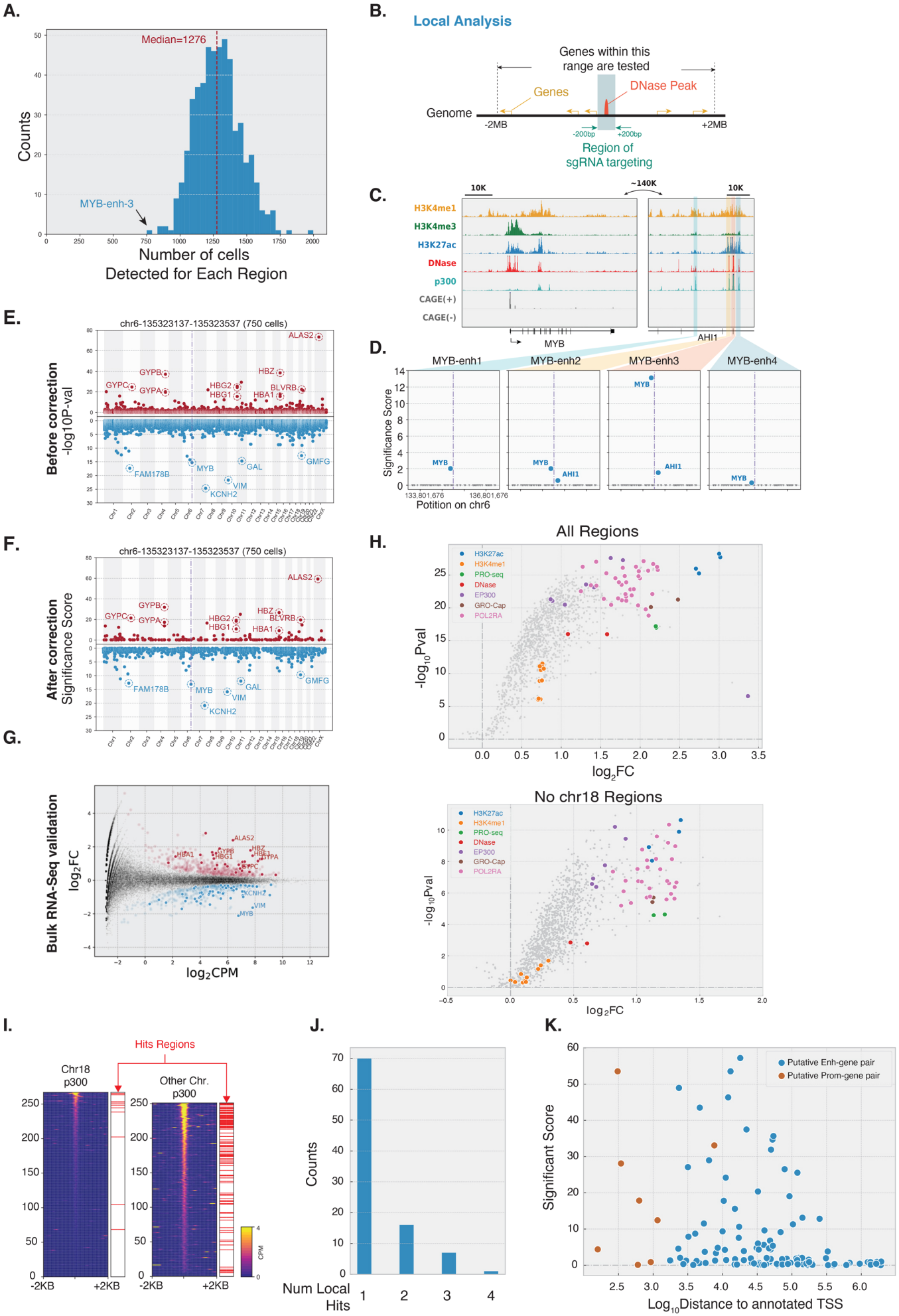

Figure S1. Mosaic-seq reveals the primary hits of enhancers. Related to Figure 1.

A. Distribution of cells detected per targeted region. Note that MYB-enh-3 has the lowest cell number (750 cells), which indicates that knocking down of MYB affects the cell proliferation.
B. Shown is a schematic representations of how local hits are identified.
C. A snapshot of genome browser tracks indicating the epigenetic features of MYB neighboring regions in K562 cells. ChIP-seq and DNase-seq data are from ENCODE, and CAGE data are from FANTOM.
D. Manhattan plot represents the genes that are significantly changed upon suppression of indicated enhancers. Y-axis represents the significance score (see methods). Genes are ordered based on the positions on different chromosomes (X-axis). The vertical dashed line indicates the position of target regions. The genes with positive SS are shown in blue dots and the other genes are show in inverted triangles. For visualization purposes, all genes with zero or negative SS are aligned at y=0.
E. Manhattan plot represents uncorrected p-values of genes upon suppression of MYB-enh3. Y-axis represents the raw p-values (see methods). Genes are ordered based on the positions on different chromosomes (X-axis). The vertical dashed line indicates the position of target regions. For visualization purposes, all genes with zero log10(p-values) not shown.
F. As above, after correction with the significance score (SS). For visualization purposes, only genes with SS greater than zero are shown.
G. MA-plot from bulk RNA-Seq where MYB-enh3 is repressed. The MA-plot compares 10 sgRNAs (one perturbing each enhancer, with two biological replicates) with 2 negative control sgRNAs (with at least two biological replicates). Grey transparent dots represents non-significant genes. Red and blue transparent dots represent significantly up- or down-regulated genes, respectively. Solid red / blue dots represent the hits that are overlapped with the Mosaic-seq.
H. Enrichment analysis for different epigenetic features. (top) Using ChIP-Seq and other datasets generated by the ENCODE Consortium in K562 cells, we compared genomic signal at all 94 enhancers with local hits to all the remaining targeted enhancers. (bottom) We also performed the same analysis after excluding enhancers on chromosome 18. Each dot represents a single experiment. Multiple dots of the same color represent different biological replicates or growth conditions, produced either in the same or different laboratory.
I. Enhancer hits are depleted from chromosome 18. Heatmap shows the p300 ChIP-seq signals for all the regions targeted, the red bar to the right of each panel indicates the enhancers identified as having local hits.
J. Distribution of the number of primary gene targets of enhancers.
K. Distribution of enhancer-target gene distance. All the TSS are based on the annotation. The targeted region that regulates RP11-461C13.1 (chr5:54325645-54326045) is likely a promoter.

Figure S2. Optimization of Mosaic-seq parameters

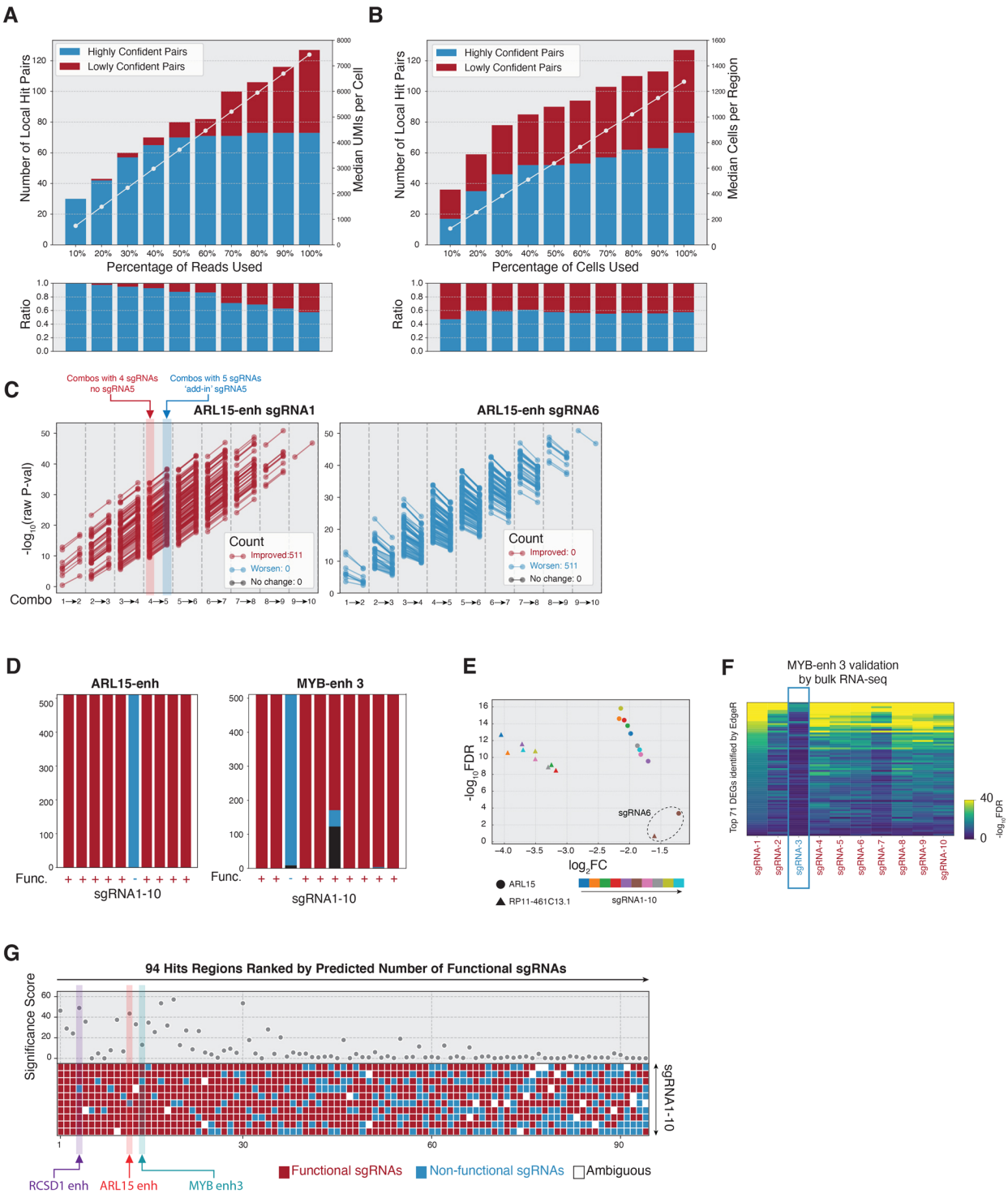

Figure S2. Optimization of Mosaic-seq parameters. Related to Figure 2.

A. Number of hits identified when downsampling the reads sequenced per cell. Bars indicate the number of local enhancer-target gene pairs identified with the indicated percentage of reads used. Counts of pairs with high confidence (SS >= 1) are shown in red and the ones with low confidence (SS < 1) are shown in blue. Dotted line indicates the median UMIs per cell (y-axis, right). The bottom panel indicates the ratio between high confidence and low confidence pairs for each downsampling. See STAR Methods for details.
B. Similar to A, but downsampling the number of cells sequenced.
C. Examples of two sgRNAs targeting ARL15-enh. Lines connect two points: the left point represents a combination of sgRNAs; the right point represents the same combination plus the indicated sgRNA. Red indicates increased statistical significance; blue indicates decreased significance.
D. Bar chart summarizing the combinatorial analysis in A for the 10 sgRNAs targeting ARL15-enh and MYB-enh3 (counting the frequency of 'improved', 'worsen' and 'no change').
E. RNA-seq validation of individual sgRNAs from ARL15-enh. Different colors represent individual sgRNAs. Circles and triangles represent two major target genes. FDR and fold change values are calculated by EdgeR. Note that sgRNA-6 has much weaker effect on both targets comparing with the other sgRNAs.
F. RNA-seq validation of individual sgRNAs from MYB-enh3. Heatmap showing summary of FDR for the top 71 DEGs for sgRNAs targeting MYB-enh-3. Note that sgRNA-3 has the worst performance among all 10 sgRNAs. Using data from bulk RNA-seq.
G. Functional evaluation for individual sgRNAs targeting all regions with local hits. Top panel, the most significance score observed for each region. Bottom panel, heatmaps shows the functional evaluation for individual sgRNAs in a region (See STAR Methods). Regions are ranked based on the number of sgRNAs that are predicted to be functional.

Figure S3. More examples for global analysis of enhancer primary/secondary targets.

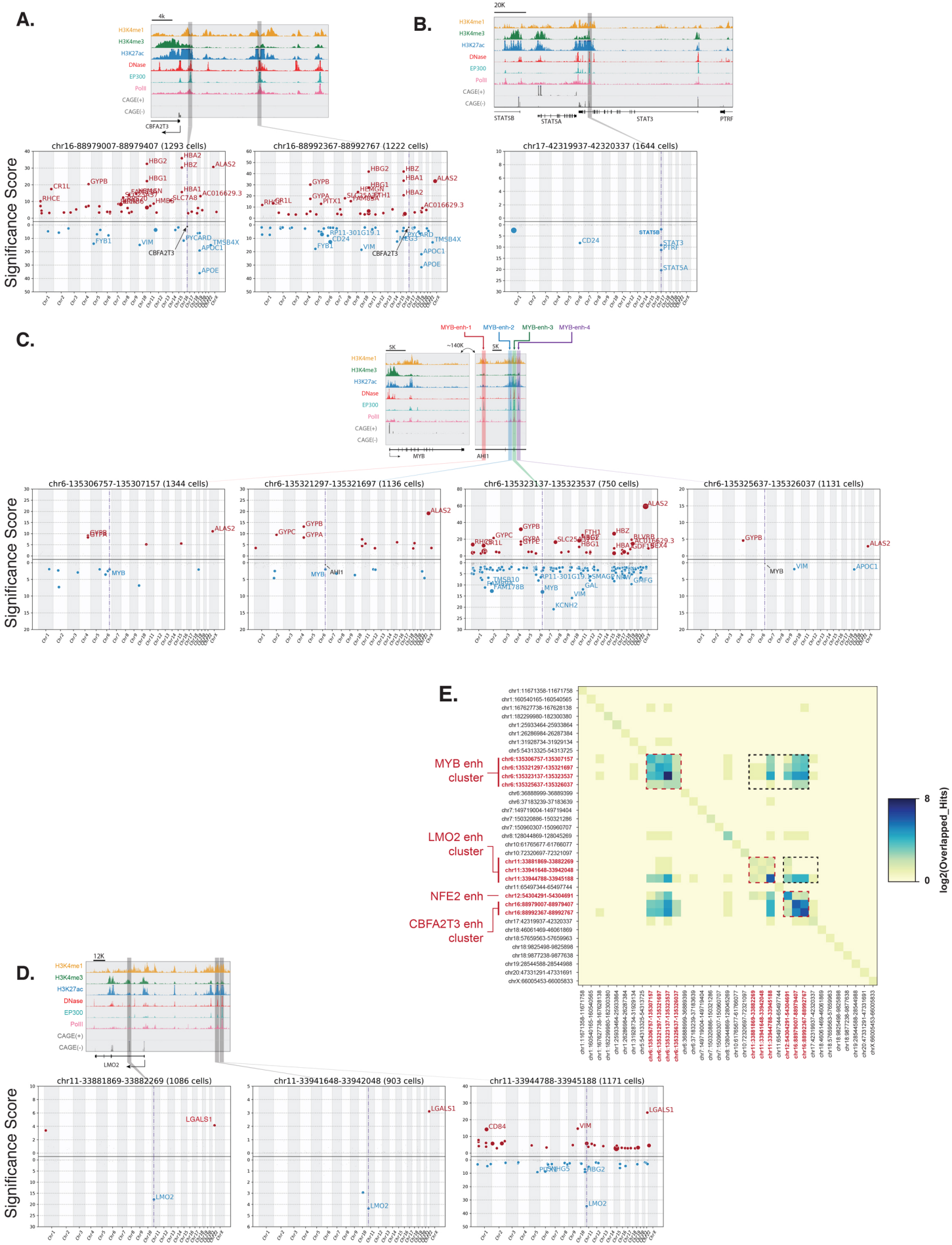

Figure S3. More examples for global analysis of enhancer primary/secondary targets. Related to Figure 3.

A-D: Similar to Figure 3A and B. The black arrows indicate local hits that do not pass global hits threshold.

E: This heatmap shows the number of overlapped gene targets for all 35 enhancers with global targets. The overlapped count was log2 transformed after a pseudo-count of 1 was added to each pair. Enhancers in the same loci (red squares) and across different loci (black squares) can share target genes.

Figure S4. Linking distal MYB enhancers to blood-associated traits

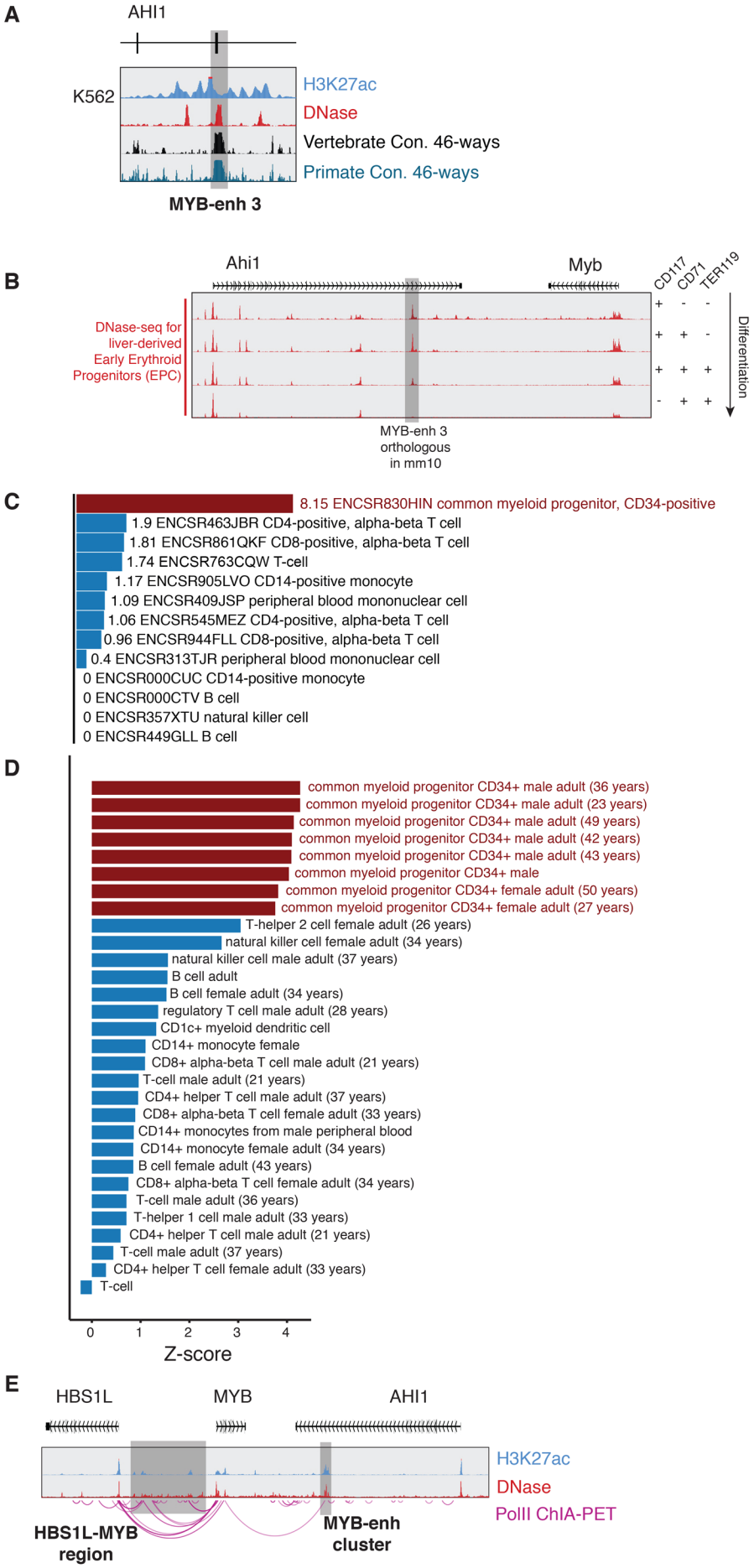

Figure S4. More information about MYB-enh-3. Related to Figure 4.

A. MYB-enh-3 is highly conserved in all vertebrate. Note that the enhancer spans the one exon and part of one intron of AHI1.
B. DNase-seq pattern of mouse early erythroid progenitors (EPCs). Note that CD117-CD71+TER119+ labels the differentiating erythroblast, which is depleted of this enhancer.
C. Expression level of MYB in different primary blood cells. The numbers indicate the log2 TPM of the gene. The figure is generated by the SCREEN web interface from ENCODE project (http://screen.encodeproject.org).
D. DNase I hypersensitivity at MYB-enh-3 in different primary blood cells. The x-axis indicates the Z-scores of the cell for the region. The figure is generated by the SCREEN web interface from ENCODE project (http://screen.encodeproject.org).
E. ChIA-PET from K562 cells demonstrate the direct interactions between MYB-enh studied in this work and the promoter of MYB.

Supplementary Tables

Table S1. All the oligos used in this study. Related to STAR Methods.

Table S2. Sequencing statistics. Related to STAR Methods.

Table S3. Primary and secondary enhancer targets identified in this study. Related to Figure 1.

Table S4. Differentially expressed genes identified in bulk RNA-seq experiments. Related to Figure 3.