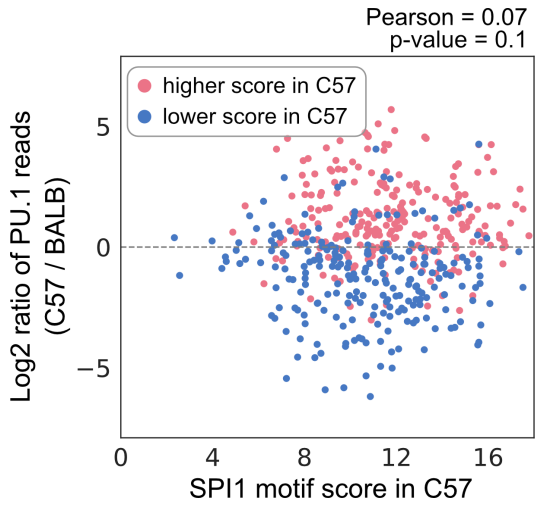
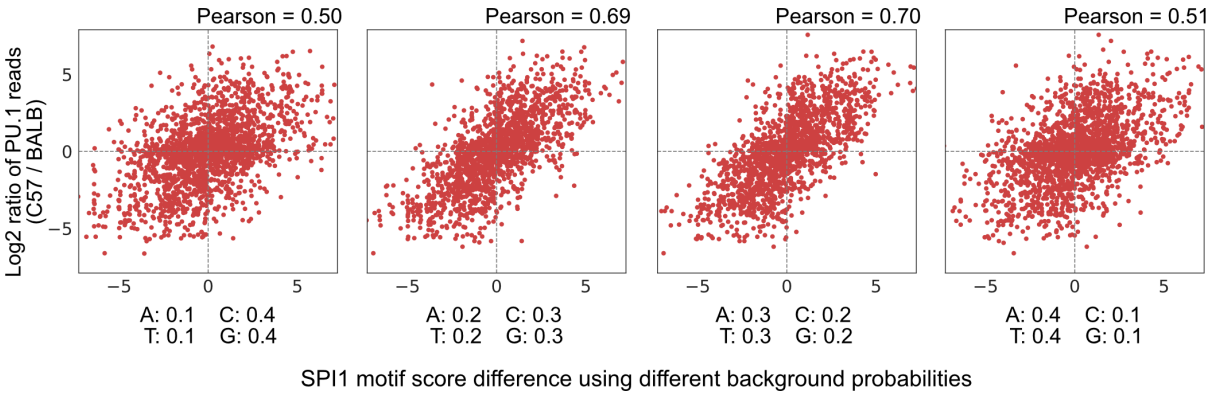
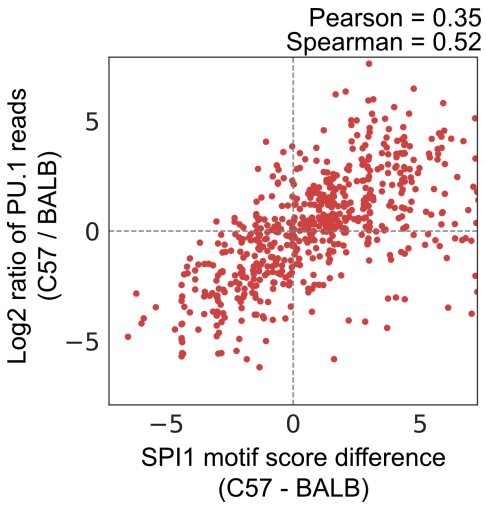


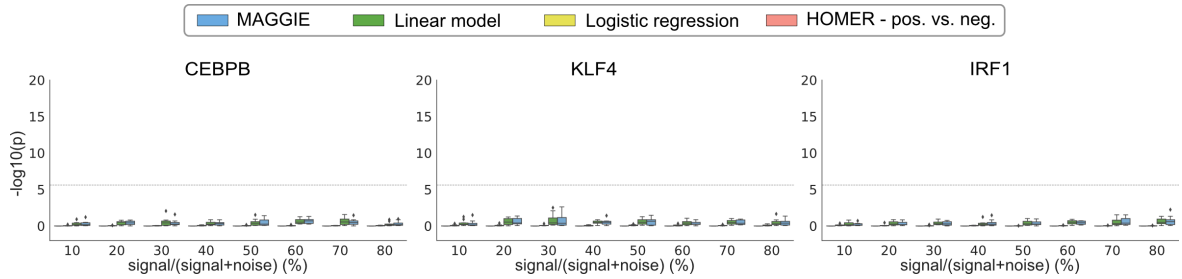
Supplementary Fig. S1. Relationship between change of PU.1 binding and SPI1 motif score at a similar level of SPI1 motif mutation. Each dot represents a PU.1 binding site that has mutation on SPI1 motif between BALB and C57 by a difference of motif score between 1 and 1.5. Red dots are binding sites with a higher motif score in C57, while blue dots are for sites with lower scores in C57. Change of PU.1 binding activity was calculated by the fold change of PU.1 ChIP-seq reads between BALB mice and C57 mice. Sites with a stronger PU.1 motif in BALB has an increase in PU.1 binding in general, but the level of binding increase is not affected by the actual motif score (Pearson coefficient = 0.07 with an insignificant p-value).



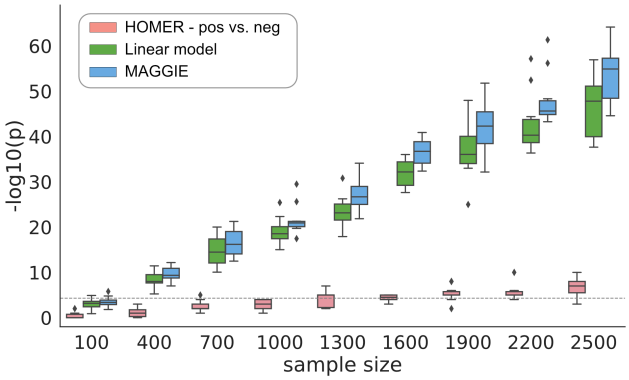
Supplementary Fig. S2. Diminished correlations between motif score differences of SPI1 motif and fold changes of PU.1 binding activity using non-uniform background probabilities. Background probabilities are displayed under each plot with GC contents ranging from 20% (rightmost) to 80% (leftmost). The PU.1 binding sites are the same as those shown in Fig. 1C.



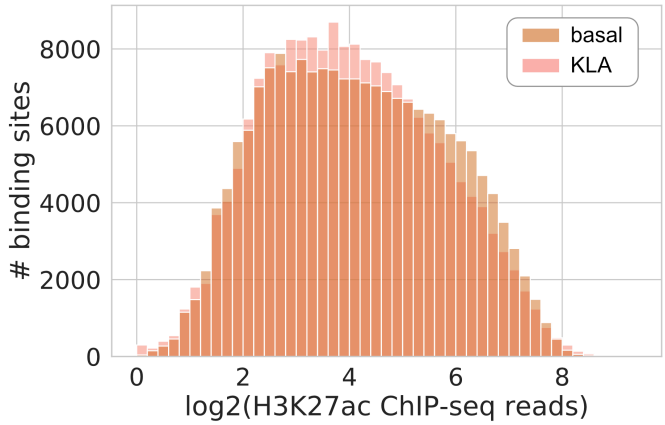
Supplementary Fig. S3. Diminished correlation between motif score differences of SPI1 motif and fold changes of PU.1 binding activity using motif at the same locations. By restricting motifs at the same locations, we ended up with fewer PU.1 binding sites having mutations on SPI1 motif.



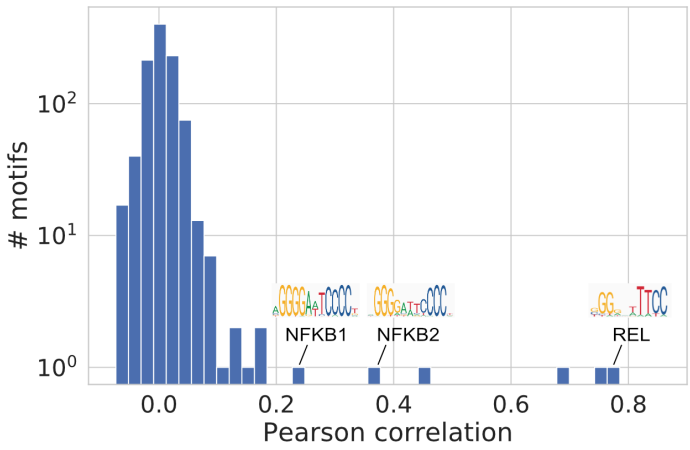
Supplementary Fig. S4. Significance values of simulated experiments from the comparative approaches for the unmutated motifs. All of the methods recognize unmutated motifs as insignificant for all levels of signal-to-noise ratio. The simulated datasets used here were the same as those in Fig. 2.



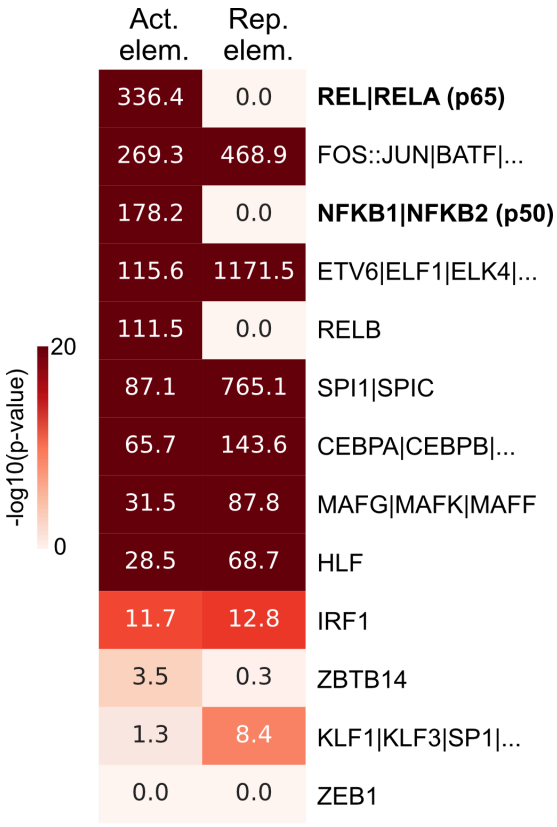
Supplementary Fig. S5. Effect of sample size on the outputs from the comparative approaches. Different number of input sequences were simulated and embedded with SPI1 and CEBPB motifs, among which 50% of total sequences experienced mutation on the SPI1 motifs. Ten simulations were repeated for each sample size. Overall, significance values increased with a larger sample size for all the methods.



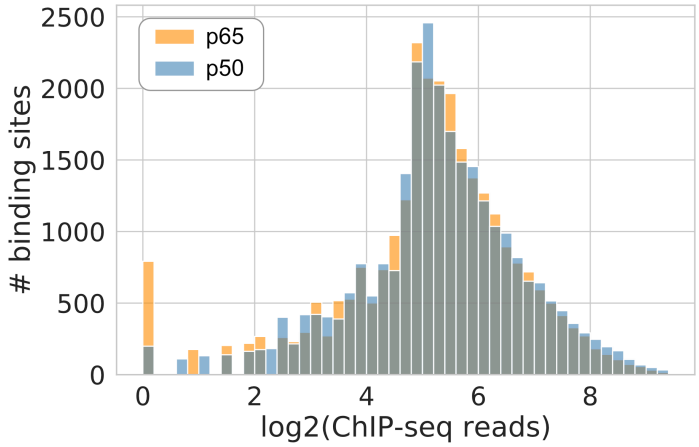
Supplementary Fig. S6. Distribution of H3K27ac ChIP-seq reads for extended open chromatin of macrophages at basal and KLA-treated conditions. ChIP-seq reads were counted within 1000-bp extended regions around open chromatin regions identified from ATAC-seq. Read counts were pooled for the four testing strains of mice. Regions with larger than 16 reads are within the more active half of the total regions.



Supplementary Fig. S7. Correlations of motif score differences between RELA motif and other testing motifs. NFKB1 and NFKB2 motif have relatively better correlations with RELA motif due to their motif similarity. However, it is not comparable to the extreme similarity between REL and RELA motif, which has a Pearson correlation coefficient larger than 0.7. Motif score differences are based on strain-specific KLA-activated regulatory elements.



Supplementary Fig. S8. Motif enrichment results from HOMER for KLA-activated and KLA-repressed regulatory elements of C57 mice. Motif enrichment was calculated by comparing to random backgrounds with default parameters. The motifs here are the same as those in Fig. 4B and are ranked by the enrichment p-values for activated elements.



Supplementary Fig. S9. Distribution of ChIP-seq reads for p65 and p50 at their respective binding sites. ChIP-seq reads were counted within 200-bp binding sites called for p65 and p50 using HOMER "findpeaks -size 200". The median counts for both factors are roughly at 5 = log2(32 reads).