Transcriptional plasticity drives leukemia immune escape

Kenneth Eagle[1,2*], Taku Harada[1,*], Jérémie Kalfon[3], Monika W. Perez[1], Yaser Heshmati[1], Jazmin Ewers[1], Jošt Vrabič Koren[4], Joshua M. Dempster[3], Guillaume Kugener[3], Vikram R. Paralkar[5], Charles Y. Lin[4], Neekesh V. Dharia[1,3], Kimberly Stegmaier[1,3], Stuart H. Orkin[1,6,†] and Maxim Pimkin[1,3,†]

Supplementary Figures 1-9

1. Cancer and Blood Disorders Center, Dana-Farber Cancer Institute and Boston Children's Hospital, Harvard Medical School, Boston, MA
2. Ken Eagle Consulting, Houston, TX
3. Broad Institute of MIT and Harvard, Cambridge, MA
4. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
5. Division of Hematology/Oncology, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA
6. Howard Hughes Medical Institute, Boston, MA

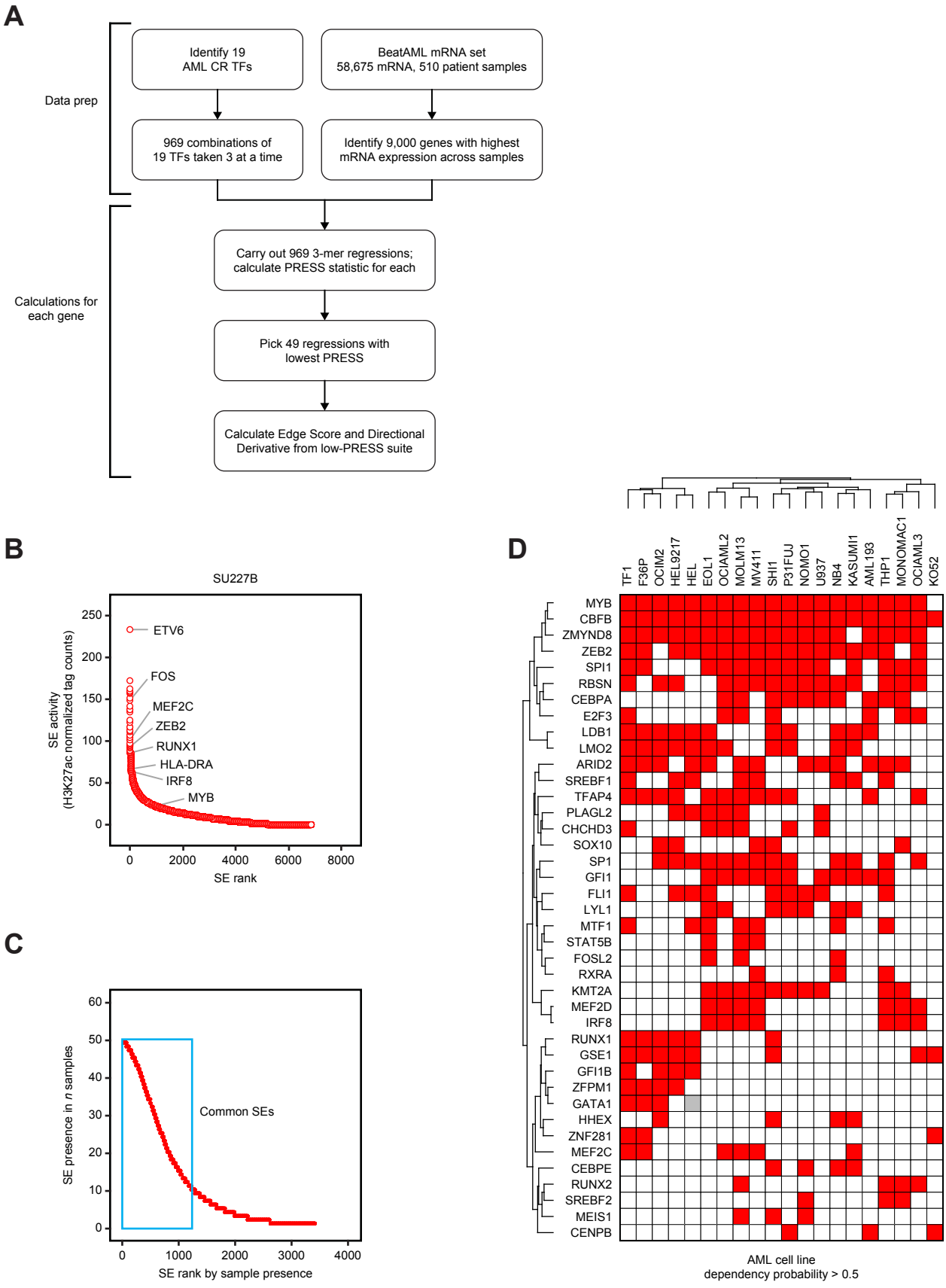
* These authors contributed equally

† These authors jointly supervised

Correspondence to:
Stuart H. Orkin
Division of Hematology/Oncology,
Boston Children's Hospital
1 Blackfan Circle,
Boston, MA 02115
Tel.: 617-919-2042
orkin@bloodgroup.tch.harvard.edu

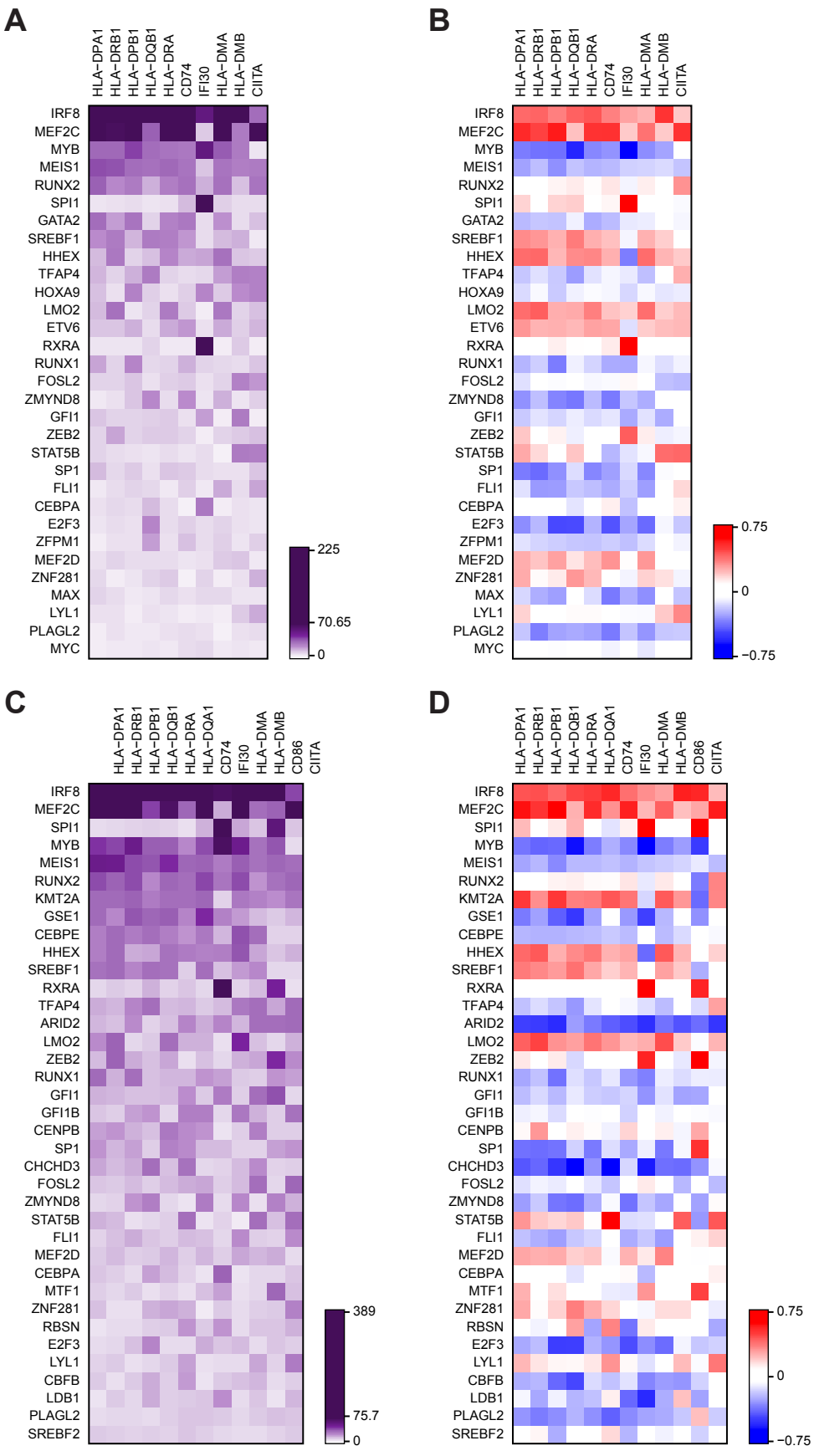
Maxim Pimkin
Department of Pediatric Oncology,
Dana-Farber Cancer Institute,
450 Brookline Ave.,
Boston, MA 02215
Tel.: 617-632-3400
maxim_pimkin@dfci.harvard.edu

Supplementary Figure 1



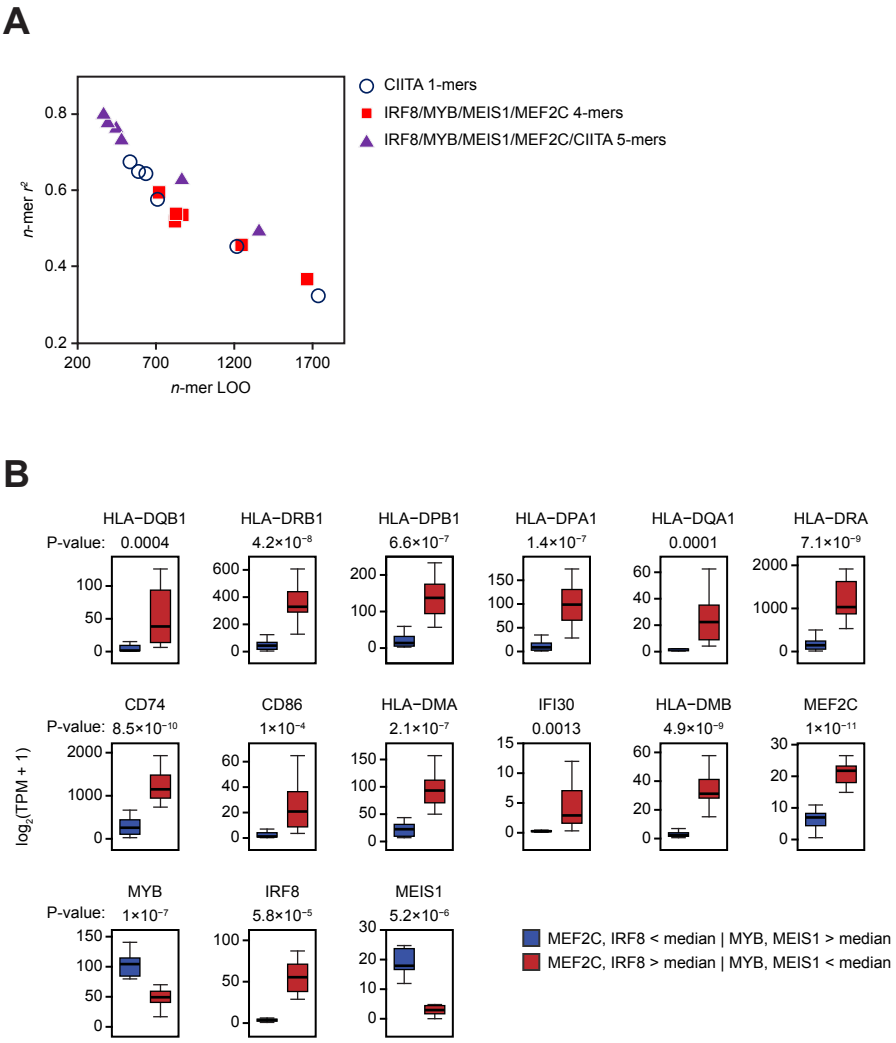
Supplementary Fig. 1. CORENODE development. A, Data and computational flowchart. **B**, SE signal and ranks in a representative primary patient sample with high MHC type II expression. **C**, SE presence in 49 primary AML samples. Data in **B** and **C** are from Ref.(9). **D**, Transcription factors identified as selective AML dependencies by the Broad Cancer Dependency Map project. The heatmap reflects probability of dependency > 0.5. Forty TF genes with dependency probability >0.5 in 3 or more AML cell lines are shown.

Supplementary Figure 2



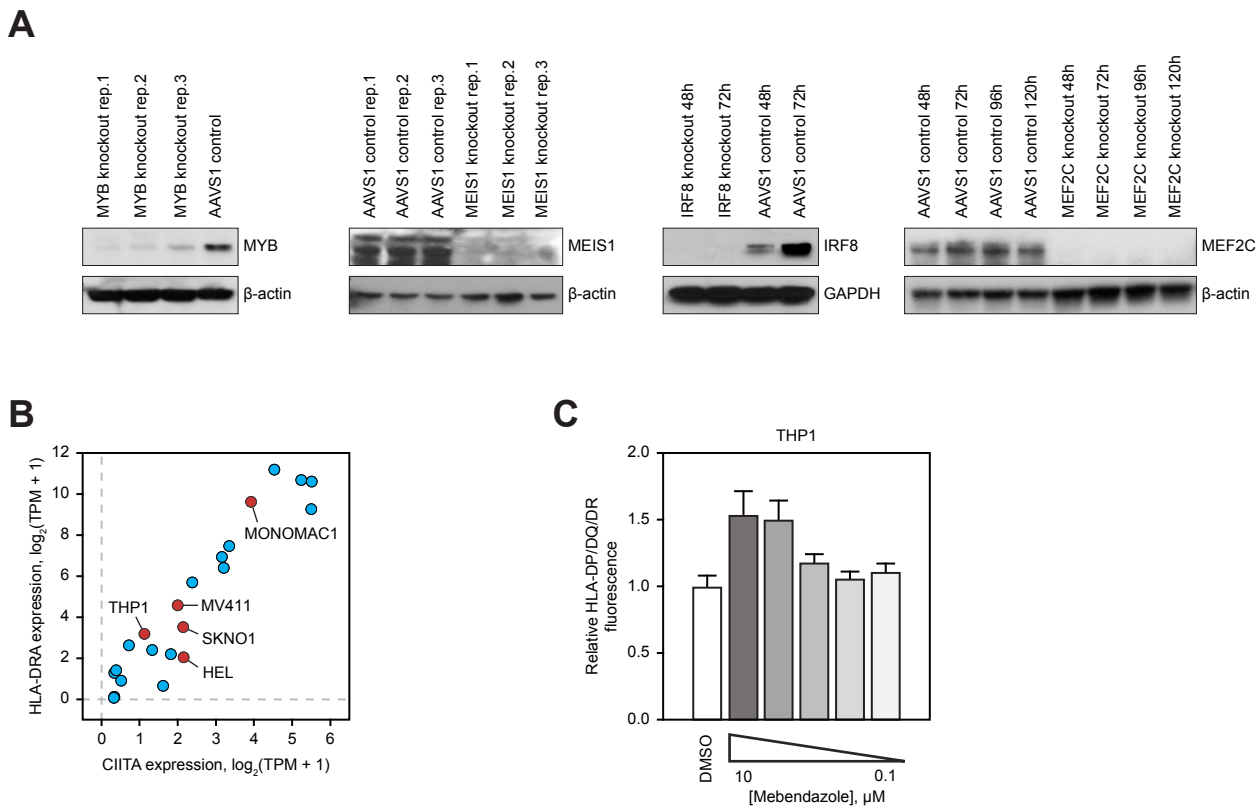
Supplementary Fig. 2. CORENODE identifies the same 4 TFs as top MHC-II regulators regardless of the CR TF set. Heatmaps of edge scores (ES) and directional derivatives (DD) representing computationally established edges between target genes (columns) and TFs (rows, sorted by average ES). Higher ES corresponds to higher confidence edges, and DD predicts amplitude and directionality (positive *vs.* negative) of TF-target regulation. Panels **A** and **B** represent CORENODE output derived from an extended set of 31 TFs defined in our parallel work on AML core regulatory circuitry (16); **A** depicts ES and **B** depicts DD. Panels **C** and **D** represent CORENODE output derived from a 37-TF set representing all selectively essential AML TFs above the minimal average expression cutoff (i.e. within the 9000 top expressed genes) regardless of SE association; **C** depicts ES and **D** depicts DD.

Supplementary Figure 3



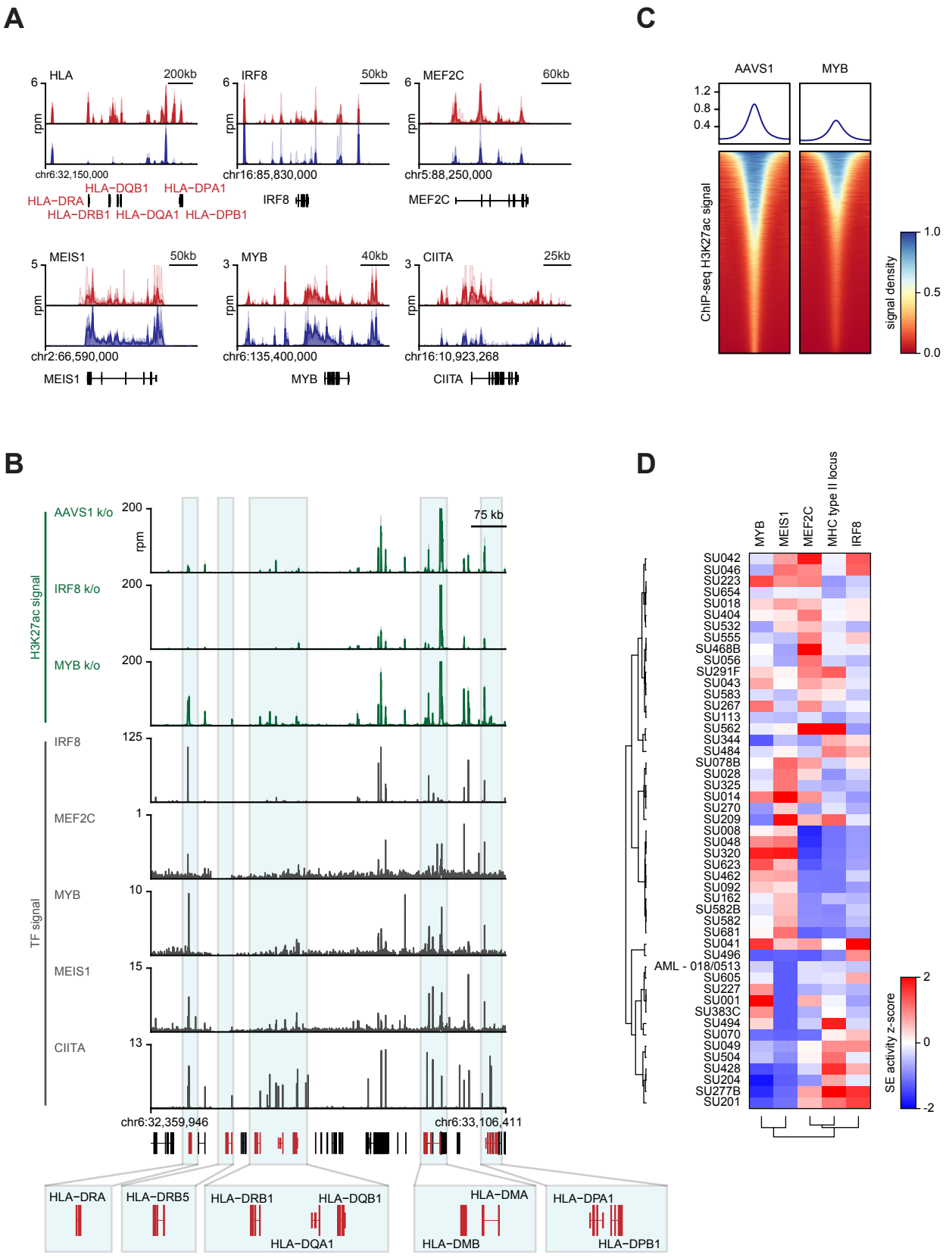
Supplementary Fig. 3. Prediction of MHC-II gene expression by CORENODE. A, Combinatorial regression of MHC-II genes. Comparison of leave-one-out error (LOO) and r^2 for 1-, 4- and 5-mers describing expression of 6 MHC type II genes (HLA-DPA1, HLA-DBP1, HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1). 5-mers where CIITA is the 5th term display the smallest LOO and the largest r^2 values indicating improved fit without overfitting. **B**, Gene expression in two populations of the TCGA dataset. The blue population represents patient samples with below-median expression of IRF8 and MEF2C and above-median expression of MYB and MEIS1, and the red population represents patient samples with the opposite pattern of TF expression.

Supplementary Figure 4



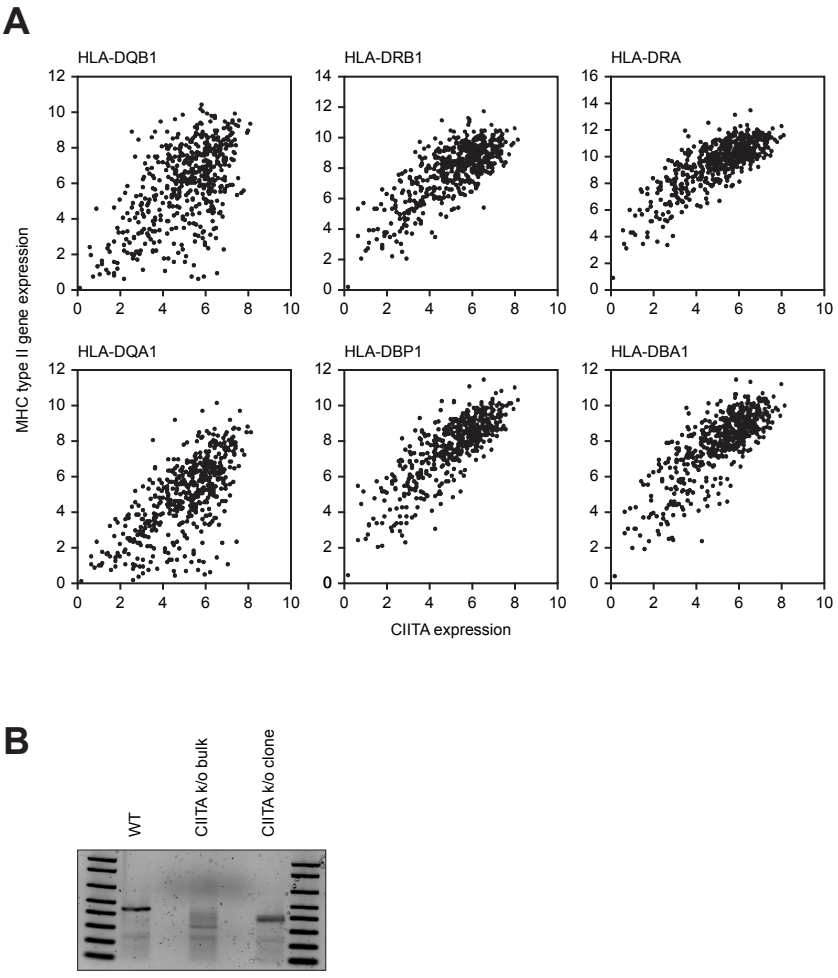
Supplementary Fig. 4. TF knockouts and MYB inhibition. A, Validation of gene knockouts. MV411 cells were electroporated with RNPs targeting the indicated genes and protein depletion was verified by Western blot. Unless indicated otherwise, imaging was performed 72 hours post electroporation. **B**, Relationship between CIITA expression and HLA-DRA expression in AML cell lines. Expression data were downloaded from the CCLE database (https://sites.broadinstitute.org/ccle/). Cell lines used for TF knockout experiments are marked in red. **C**, Treatment of THP1 cells with mebendazole for 48 hours induced surface expression of MHC-II, measured by FACS with a pan-specific anti-HLA-DP/DQ/DR antibody.

Supplementary Figure 5



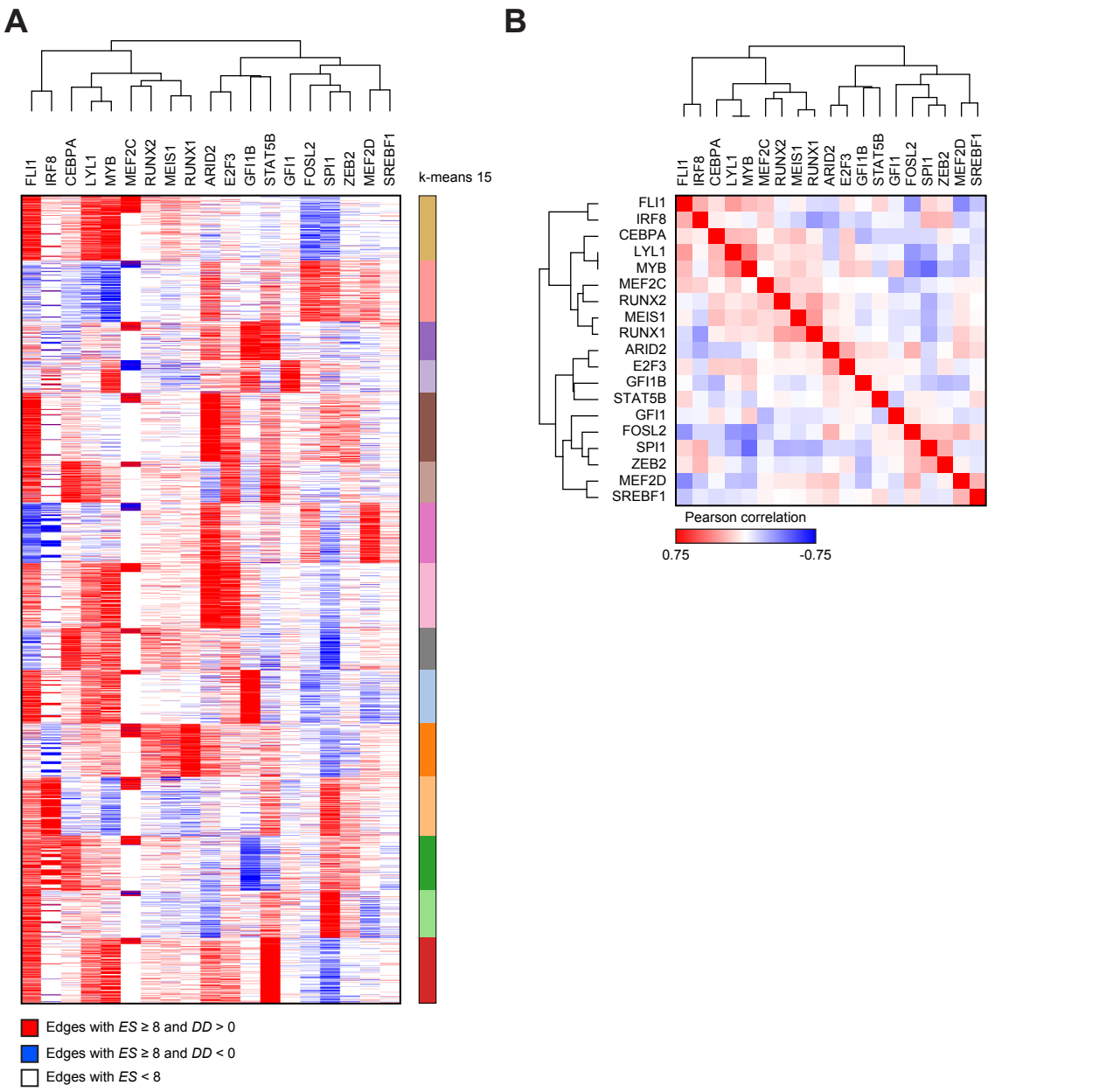
Supplementary Fig. 5. Histone and TF architecture in the MHC-II locus. A, H3K27ac metatracks. *Red*, tracks from 10 AML patients with the highest MHC type II expression; *blue*, tracks from the 10 AML patients with the lowest MHC type II expression. Each metatrack is a collection of semi-transparent area plots representing individual samples and the average profile is represented by a thick line. **B**, H3K27 acetylation and TF binding in the MHC type II locus. Green tracks are H3K27ac metatracks composed by overlaying semi-transparent area plots representing 2 biological replicates of the indicated TF knockouts compared to a control locus-targeting sgRNA (AAVS1). The average profile is represented by a thick line. Grey tracks represent binding of the indicated TFs. All of the shown ChIP-seq experiments were performed in MV411 except for the CIITA track which was downloaded from Ref.(38) and represents a lymphoid cell line. **C**, Density plots of spike-in controlled H3K27ac ChIP-seq experiment showing genome-wide histone acetylation changes after MYB knockout, compared to a control locus-targeting sgRNA (AAVS1). Each row represents a single peak. **D**, Heatmap of aggregated SE activities associated with the MHC-II locus and the tetrad TFs. H3K27ac ChIP-seq data are from Ref.(9). Multiple SEs were called by ROSE2 for each locus and their signal was aggregated and z-scored. Rows and columns were hierarchically clustered by Pearson correlation with complete linkage.

Supplementary Figure 6



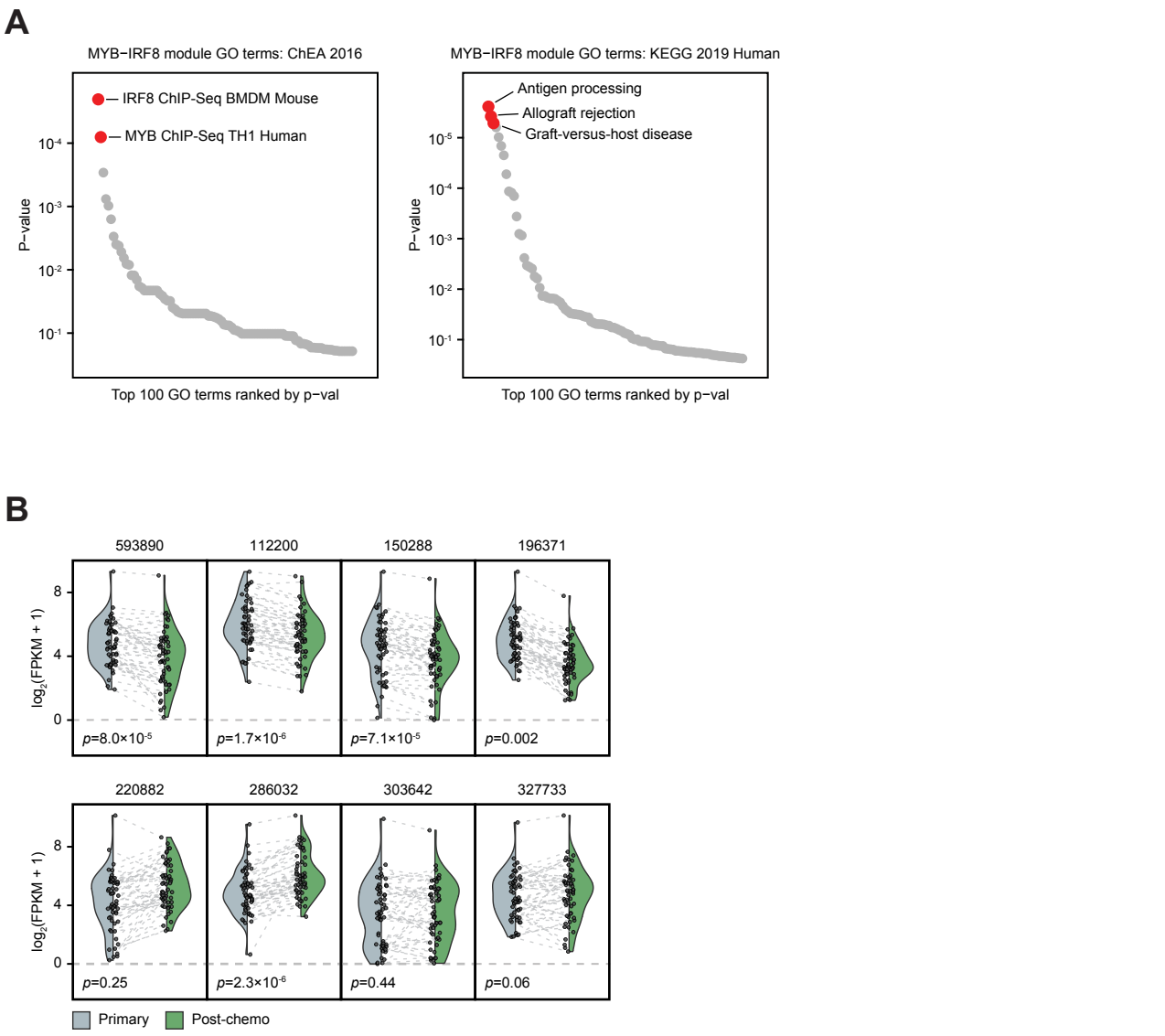
Supplementary Fig. 6. The role of CIITA in MHC-II regulation. A, Correlation between CIITA and MHC-II gene expression in the Beat AML dataset. Note that these plots represent Pearson correlation and are different from CIITA 1-mers (such as the one in Fig. 2A) which have a quadratic term in addition to a linear term. **B**, Validation of a CIITA knockout in MV411 cells by PCR amplification followed by agarose gel electrophoresis with ethidium bromide staining. Three closely positioned gRNAs were used for the CIITA knockout, producing a size shift on gel electrophoresis indicating complete editing efficiency in the bulk population and a homogenous excision in the clonally selected line.

Supplementary Figure 7



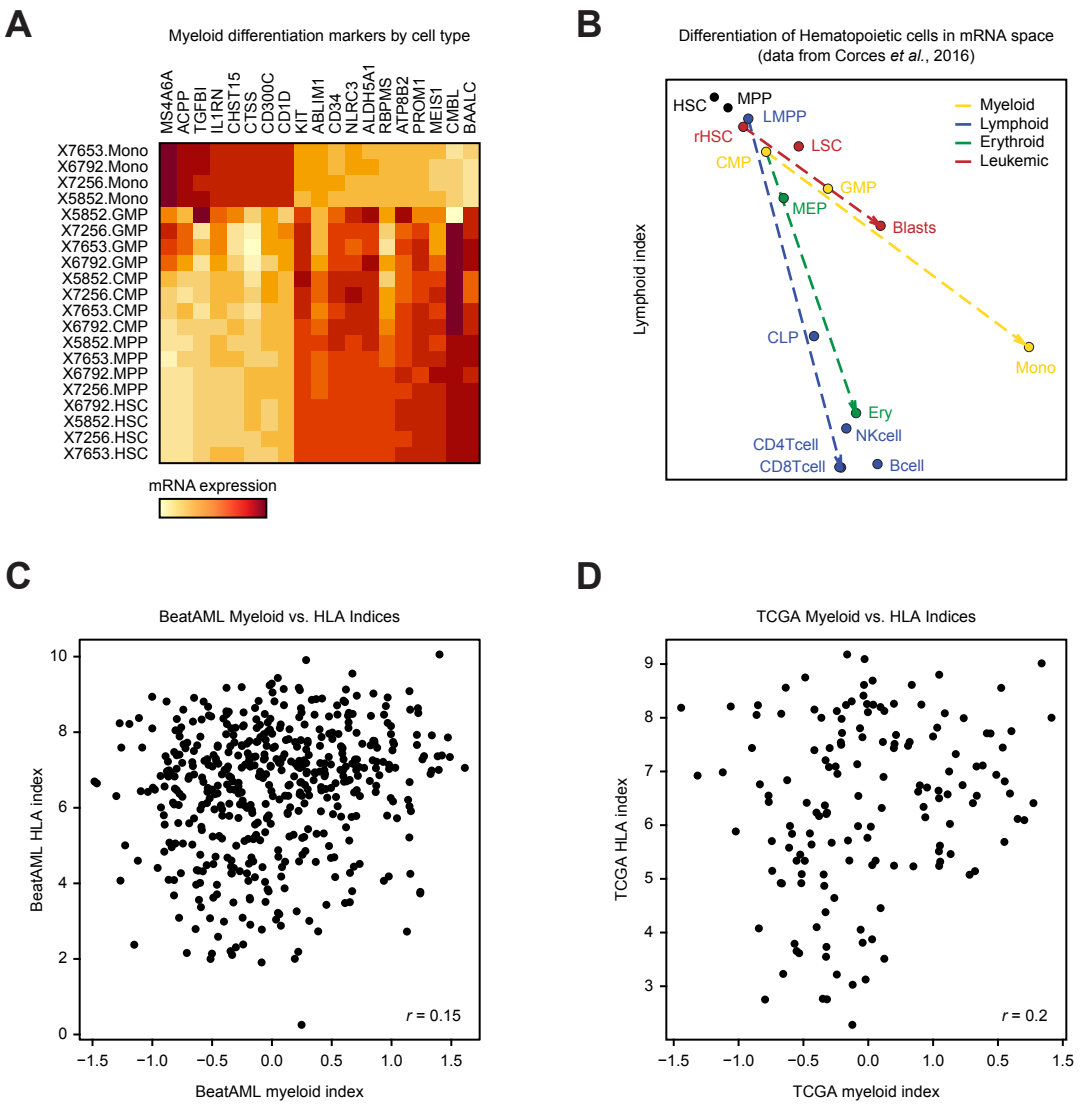
Supplementary Fig. 7. Global network decomposition by CORENODE identifies TF modules. A, A genome-wide heatmap and clustering of edges. Edges with scores ≥8 were converted to 1 for predicted positive regulation (DD > 0) or -1 for negative regulation (DD < 0). The resulting matrix was clustered by Pearson correlation (CR TFs (columns): hierarchical, complete linkage; target genes (rows): k-means with k=15) using Morpheus (https://software.broadinstitute.org/morpheus). **B**, A TF similarity matrix produced from the regulation matrix depicted in panel **A**.

Supplementary Figure 8



Supplementary Fig. 8. The GvL module. A, Gene-set enrichment analysis of the GvL module. The ChEA dataset shows enrichment of TF binding in the gene promoters, whereas the KEGG dataset shows gene pathway enrichment. Enrichments were computed using Enrichr (43). **B**, Expression of the GvL module in paired primary and post-chemotherapy relapse patient samples from Ref.(3).

Supplementary Figure 9



Supplementary Fig. 9. MHC-II expression and myeloid differentiation status. A, Expression of the 19 myeloid markers in sorted normal progenitors. **B**, Normal progenitors and leukemic cells are plotted according to their indices of myeloid and lymphoid development (see Supplementary Note for details). Panels **A** and **B** were computed with data from Ref.(42). **C** and **D**, Poor correlation MHC-II expression and myeloid differentiation state. A composite score reflecting MHC-II expression (HLA index) is plotted against the index of myeloid development using Beat AML (15) and TCGA (17) datasets, respectively.