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,†]

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

†Correspondence to: orkin@bloodgroup.tch.harvard.edu, maxim_pimkin@dfci.harvard.edu

Supplementary Note

Details of CORENODE development and validation

Definition of a core regulatory TF set for genome-wide network decomposition

The maximal practical set of TF regulators for genome-wide fitting is limited by two factors. First, the number of possible TF combinations rises rapidly with each additional TF (*i.e.* fitting term), complicating overfitting control and eventually exceeding the available computational resources. For example, 19 TFs give 969 3-mer combinations, 50 TFs give 19,600 3-mer combinations and 100 TFs give 161,700 3-mer combinations. Second, the ~600 TFs that meet minimal expression criteria in AML display significant internal collinearity (*i.e.* there are large groups of coexpressed TFs whose expression values display a high degree of correlation across AML samples; not shown), limiting the predictive power of a combinatorial regression approach. Therefore, we employed an integrative algorithm that prioritized TFs with known or predicted biological significance based on the published functional genomic datasets. Indeed, although an average mammalian cell expresses several hundred TFs, only a small number, variably referred to as reprogramming, master or core regulatory TFs, are principally important for lineage specification and, thus,

likely orchestrate the bulk of transcriptional regulation (49–53). These TFs have been shown to form highly interconnected transcriptional circuits (CRCs), are typically associated with superenhancers (SEs) and significantly overlap with context-specific gene dependencies (11,16,54,55).

To generate a reasonably sized TF set for genome-wide target fitting, the 225 selective AML dependencies (see Methods) were intersected with a published curated database of human TFs (56) and genes associated with SEs in at least 10 out of 49 primary AML samples from Ref.(9), resulting in the following list of candidate CR TFs: ARID2, CEBPA, E2F3, FLI1, FOSL2, GFI1, GFI1B, IRF8, LYL1, MEF2C, MEF2D, MEIS1, MYB, RUNX1, RUNX2, SPI1, SREBF1, STAT5B, ZEB2 (Supplementary Fig. 1). Genome-wide regressions and validations were performed on this 19-member set. For additional validation, MHC-II regressions were repeated on 31- and 37-member sets as described in the main text (Supplementary Fig. 2).

Inevitably, this approach leaves out many potentially important regulators which do not meet one or more of the above criteria, for example non-essential TFs or non-DNA-binding cofactors. Instead, for each target gene CORENODE generates a non-exclusive list of potentially regulating TFs.

***n*-mer construction**

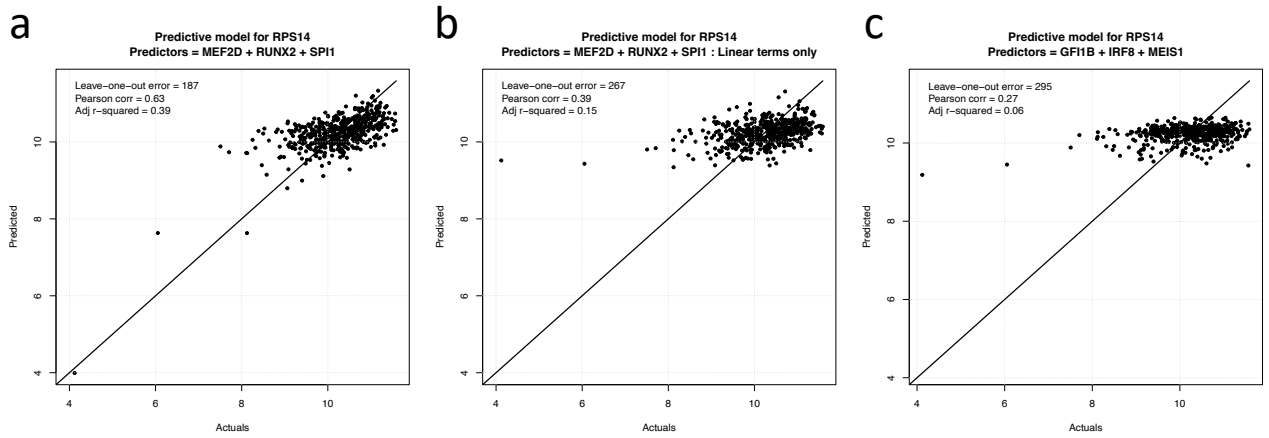
The 19 CR TF genes were combined into *n*-mers and used to regress expression of each component of the expressed genome across the BeatAML dataset (Supplementary Fig. 1). Each *n*-mer consisted of *n* TF genes (*i.e.*, a 3-mer contained 3 of the 19 TF genes). Therefore, there were 19 taken *n* at a time combinations of CR TFs: 171, 969, 3876, and 11628, respectively, for *n*=2–5. Each *n*-mer consisted of *n* terms linear in the component CR TF transformed expressions, *n* terms quadratic in the component CR TF transformed expressions, and *n* taken 2 at a time quadratic cross terms (yielding 1, 3, 6, and 10 cross terms, respectively, for *n*=2–5) in the component CR TF transformed expressions. For example, the regression equation for a 2-mer using MYB and CEBPA transformed mRNA to fit the transformed expression of gene A1BG would be as follows:

$$Y_{A1BG} = Int + A \times TF_{MYB} + B \times TF_{CEBPA} + C \times TF^2_{MYB} + D \times TF^2_{CEBPA} + E \times TF_{MYB} \times TF_{CEBPA}$$

where:
Y_{A1BG} is the transformed A1BG expression data
Int is the intercept term
TF_{MYB} and TF_{CEBPA} are the transformed expression data of the two CR TFs, and
A, *B*, *C*, *D*, and *E* are the regression coefficients

Regression and leave-one-out error

Standard multiple linear regression was carried out using the R routine "lm" with default options. The regressed "*y*" values were the transformed mRNA expression (described above) across the relevant



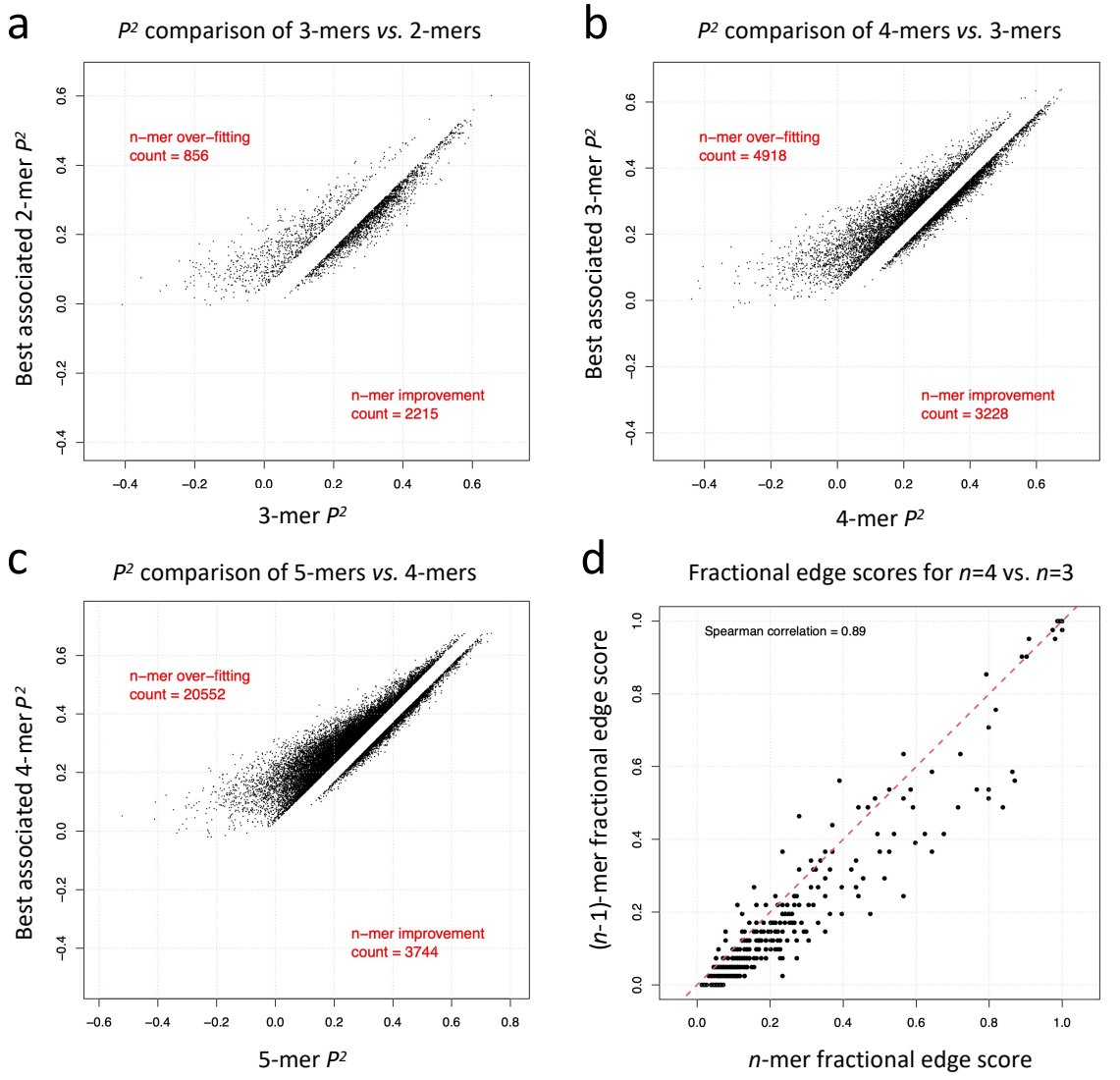
Box 1. Examples of good and bad predictive models using RPS14 as a target. a, Expression levels of MEF2D, RUNX2 and SPI1 are used as predictors; quadratic and cross terms are included in the model. **b**, Expression levels of MEF2D, RUNX2 and SPI1 are used as predictors; only linear terms are included in the model. **c**, Expression levels of GFI1B, IRF8 and MEIS1 are used as predictors; despite inclusion of quadratic and cross terms this model has no predictive power. Each data point is a sample from the BeatAML dataset (*n*=510). Predicted and actual expression values are given as $\log_2$(TMP+1).

samples; each *n*-mer combination was regressed separately against these values. Leave-one-out error (LOO), a type of cross-validation, was calculated analytically for each *n*-mer using the "PRESS" function from the R "MPV" package and the regression results from "lm".

By including quadratic and cross terms, we hypothesized that the *n*-mers would be able to capture non-linear behaviors and TF interactions. For example, fitting expression of RPS14 with the expression of 3 CR TFs MEF2D, RUNX2 and SPI1 yields an adjusted *r*-squared of 0.39 when including the non-linear terms, but the adjusted *r*-squared is only 0.15 using just the linear terms; plots of actual *vs.* predicted expression for these two fits are shown in **Box 1**. While both regressions pass through the cluster of points around mRNA=10, the fit with the non-linear terms makes much better predictions at both lower and higher expression values. Including the non-linear terms also reduces the aggregate leave-one-out error (see discussion of this statistic below) by 30%. To demonstrate that not all *n*-mers have predictive power, even with the non-linear terms, we show the best possible fit for the same RPS14 expression data using expression of CR TFs GFI1B, IRF8 and MEIS1; clearly this *3*-mer has no predictive power (**Box 1c**).

n-mer benchmarking

We undertook benchmarking to understand the appropriate level of *n* that would provide sufficient prediction accuracy, minimize the risk overfitting and have reasonable computation times. Indeed, increasing the number of fitting terms poses two major obstacles. First, given our intention to evaluate all *n*-mer combinations across each of 9000 genes, there are significant computational difficulties posed by the



Box 2. CORENODE n-mer benchmarking. Panels **a**, **b** and **c** illustrate P2 difference between 2- and 3-mers, 3- and 4-mers, and 4- and 5-mers, respectively, using a diagnostic set of combinatorial regressions with the 19 CR TFs as targets. **d**, Similar fractional edge counts are obtained using 3-mers versus 4-mers for the diagnostic set of 19 TFs used as targets.

increasing number of combinations of the CR TFs with higher *n* (969 combinations at n=3, 3876 at n=4, 11628 at n=5, and 27132 at n=6). More importantly, while increasing *n* will always increase the r^2 of the regressions, there is a significant risk of overfitting at higher values of *n* because of the increased number of regression terms (10 terms including an intercept at *n*=3, 15 terms at *n*=4, 21 terms at *n*=5, *etc*.). We chose to measure the overfitting risk using leave-one-out (LOO) cross validation, which measures the error in predicting each sample value using all the other sample values as inputs (see p.24 in Ref.57). To that end, for every regression we calculated aggregate LOO across the set of samples used for regression, which is also referred to as the PRESS statistic.

Thus, as a diagnostic test, we compared LOO results across *n*-mers fit to expression of the 19 CR TF genes as targets. Each *n*-mer can be thought of as a combination of several (*n*-1)-mers; in particular, three 2-mers can be associated with each 3-mer, four 3-mers can be associated with each 4-mer, and ten 4-mers can be associated with each 5-mer. We compared the best associated 2-mers (in terms of LOO) to each 3-mer to calculate the degree of LOO improvement or degradation from adding the third component. Similar analyses compared 3-mers to 4-mers and 4-mers to 5-mers. This benchmarking did not include any *n*-mers where the same CR TF was both a predictor and the predicted target; such situations would necessarily allow perfect fit and distort any statistical results. To ascertain risk of overfitting versus goodness of regression, we used two quantitative tests:

- First, for the CR TFs as regression targets, we calculated the Spearman correlation of the fractional edge scores (defined below) from each (*n*-1)-mer to the fractional edge scores from the associated *n*-mer.
- Second, we use a P^2 statistic derived from the PRESS statistic (58), comparing the P^2 from each *n*-mer and the best associated (*n*-1)-mer. After eliminating all cases where the change in P^2 was less than one standard deviation of all such changes in P^2, we counted the number of instances where P^2 was improved by adding the extra term *vs*. the number of instances where P^2 was degraded. Note that where the standard R^2 measure of goodness of regression fit will always improve with added terms, P^2 can improve or degrade depending on the balance between improved capture of the underlying behavior *vs.* overfitting.

As demonstrated in **Box 2**, approximately 72% of the non-equivalent associations show a higher P^2 value using the 3-mers, demonstrating that on average 3-mers contain useful information compared to the associated 2-mers and are not overfitting. By comparison, only 39% of 4-mers have useful information compared to the associated 3-mers, while for 5-mers only 15% have useful information compared to associated 4-mers. Therefore, we concluded that using 3-mers for our model would provide the best balance between goodness-of-fit and overfitting genome-wide.

Additionally, we examined the specific CR TFs that are predicted to regulate the other CR TFs, using a normalized version of the edge scores as described below. The results using 3-mers correlate strongly with those using 4-mers (Spearman correlation = 0.89; **Box 2d**), again confirming that 3 is an appropriate value of *n*.

***3*-mer edge scores and derivatives**

We regressed transformed gene expression data against the transformed expression of three CR TFs, including the linear, quadratic, and cross terms as described above, yielding equations for each gene of the form:

$$Y = Int + A \times TF_1 + B \times TF_2 + C \times TF_3 + D \times TF_1^2 + E \times TF_2^2 + F \times TF_3^2 + G \times TF_1 \times TF_2 + H \times TF_1 \times TF_3 + I \times TF_2 \times TF_3$$

where:

Y is the transformed target gene expression
Int is the intercept term
TF_1, TF_2, and TF_3 are the transformed CR TF expression data, and
A, *B*, *C*, …, *I* are the regression coefficients

One such equation was generated for each of the 969 3-mer combinations of the 19 CR TFs, for each of the 8,981 expressed target genes. Each of the 3-mers was evaluated using the PRESS statistic described above; we chose the 49 best 3-mers (the 5% of the 969 3-mer combinations with the smallest PRESS statistics) to represent a suite of good fits to the data. Edge scores for each gene were calculated by counting the number of the 49 best 3-mers that included terms for each of the 19 CR TFs. The edge scores therefore ranged from 49, where a particular CR TF was found in every one of the best 49 3-mers, to 0, where none of the best 3-mers included that CR TF. By definition, the sum of the 3-mer edge scores for each gene was 147.

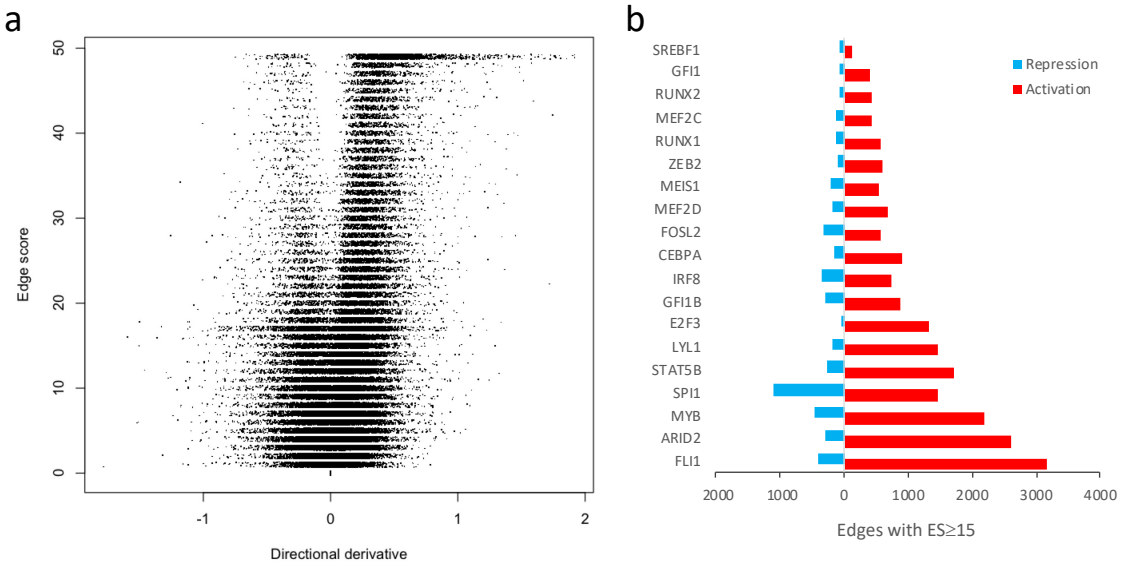
In order to predict the amplitude and direction (positive versus negative) of TF-gene regulatory connections, we calculated a directional derivative (*DD*). For each gene, for each of the 49 best 3-mers, analytic derivatives of the regressed expression function above were taken with respect to the TF_i, yielding:

$$\frac{dY}{dTF_1} = A + 2D \times TF_1 + G \times TF_2 + H \times TF_3$$

$$\frac{dY}{dTF_2} = B + 2E \times TF_2 + G \times TF_1 + I \times TF_3$$

$$\frac{dY}{dTF_3} = C + 2F \times TF_3 + H \times TF_1 + I \times TF_2$$

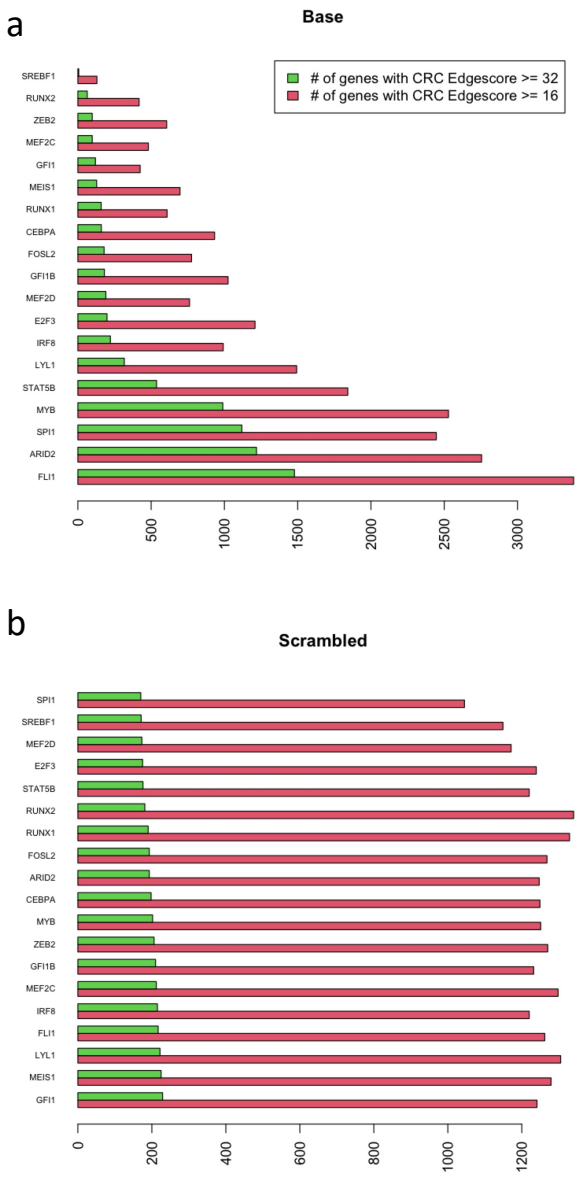
CR TF expression data for each sample was plugged into these equations, yielding a large number of estimated slopes for each gene (49 3-mers × 3 CR TFs per 3-mer × the number of samples; see **Fig. 1D** for a graphic illustration). These estimates were aggregated for each target gene by CR TF across all the samples and 3-mers, yielding a distribution of values for the slope of that gene's response to a change in the expression of the CR TF. We calculated the mean of this distribution, and also transformed the mean into a z-score by dividing by the standard deviation of the distribution.



Box 3. Distribution of edge scores and directional derivatives. a, Correlation between edge scores and directional derivatives. ES and DD values are shown for all edges in the expressed genome (all CR TFs vs. all target genes), demonstrating that edges with higher edge scores tend to have higher directional derivatives. It also demonstrates the preponderance of positive over negative regulation in the genome. **b**, Numbers of positive (DD>0) vs. negative (DD<0) high confidence edges (ES≥15) are shown for each CR TF, highlighting preponderance of predicted gene activation over repression and significant variability of high confidence edges between TFs.

As expected, edge scores correlated with DD scores with higher edge scores predicting higher amplitude of TF-target regulation (**Box 3a**). Importantly, CORENODE identified a significantly higher number of edges with positive DD values indicating positive regulation (i.e. gene activation), rather than negative DD values indicating negative TF-target regulation (i.e. repression) (**Box 3b**).

Ultimately, our approach assigns two values to each TF-target connection, or edge, in the expressed genome. The edge scores (ES) reflect the predicted confidence of the TF-target connection, while directional derivatives (DD) predict the direction and amplitude of regulation. Using aggregated information

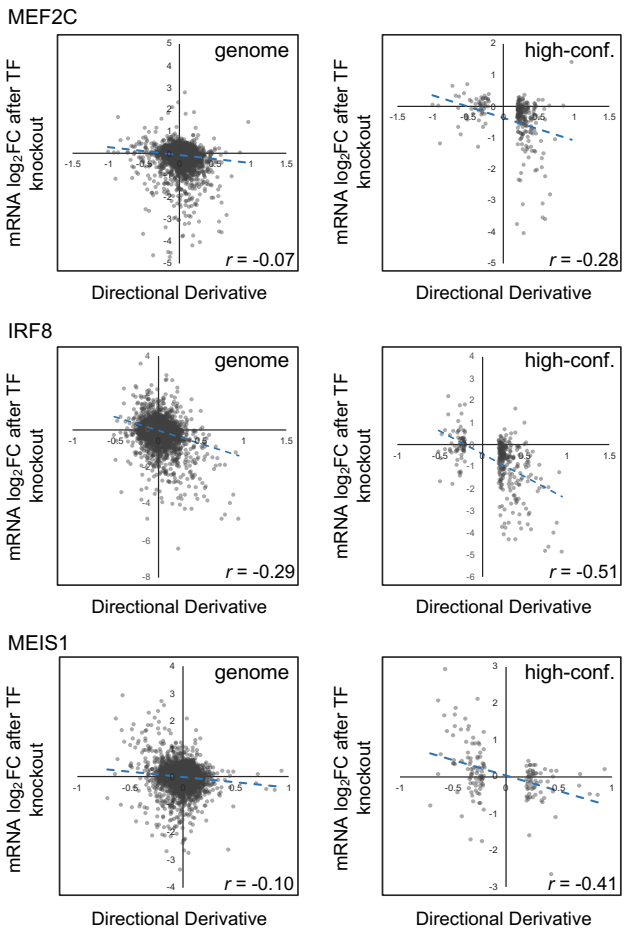


Box 4. Distribution of high-confidence edges by CR TF in the genome-wide model compared with scrambled control. Number of edges with high edge scores varies by CR TF (**a**). Calculating edge scores on scrambled data results in loss of this variance and lower edge scores overall (**b**).

from the top 5% of regressions for calculating these edge characteristics (rather than simply taking the best fit) takes into account that the predictions from the top *n*-mers are relatively similar and picking the single best *n*-mer would unrealistically restrict the available information from the fitting process. Indeed, aggregating these results across the genome-wide set of all 9000 genes results in a marked contrast between TFs that are predicted to regulate many genes across the genome (FLI1, ARID2, SPI1) *vs*. those that are predicted to regulate relatively few genes (SREBF1 and RUNX2) (**Box 4a**). As a computational control, we scrambled each genes' mRNA data among the 510 samples and repeated the genome-wide *ES* calculations (**Box 4b**). The lack of structure, with all CR TFs showing similar numbers of influenced genes, suggests that the patient results are demonstrating significant biological structure.

CORENODE validation

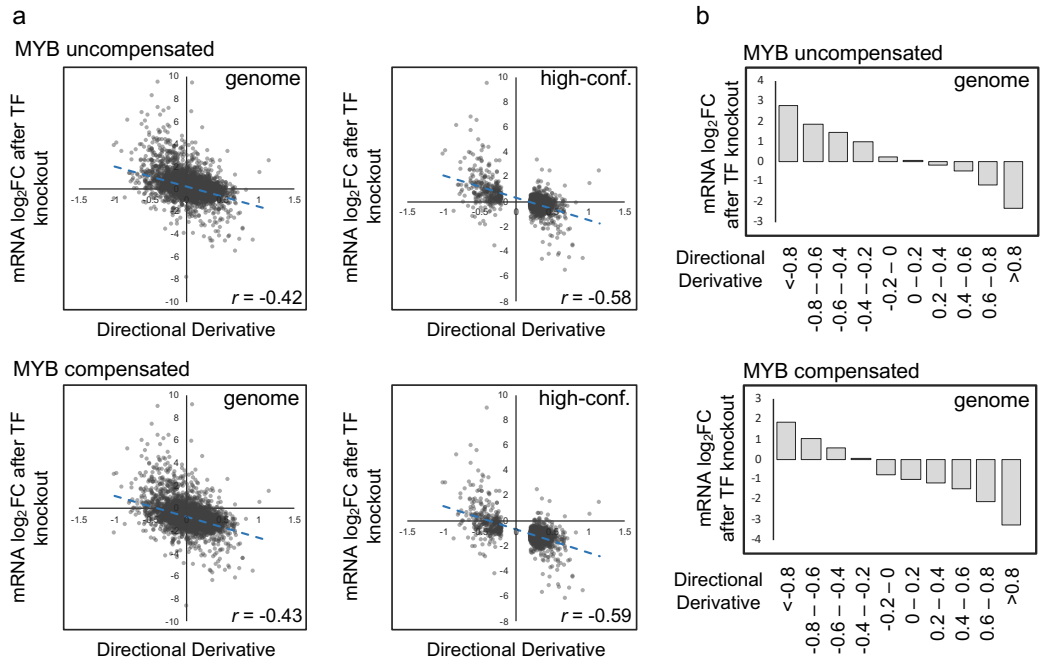
We set out to validate CORENODE predictions by performing CRISPR knockouts of the CR TFs predicted to regulate the MHC type II genes followed by an assessment of global transcriptional response using RNA-seq. We delivered *in vitro* assembled RNPs with 2-3 gRNAs by nucleofection for an efficient and rapid depletion of the CR TFs (59). Knockout efficiency was confirmed by Western blot (Supplementary Fig. 4A). Total RNA was extracted and mRNA-seq was performed 72 hours after electroporation.



Box 5. Genome-wide validation of CORENODE edges: IRF8, MEF2C and MEIS1. Correlation between directionality derivative (DD) and actual $\log_2$ fold change 72 hours after TF knockout measured by RNA-seq, is shown genome-wide (*left*) and for high-confidence edges (ES ≥ 15 and [DD] ≥ 0.2, *right*), with Pearson correlation (*r*) for each plot.

Next, for each CR TF we generated a set of high confidence CORENODE edges by taking all genes with ES ≥ 15 and [DD] ≥ 0.2 and plotted their DD scores against the measured response in the RNA-seq experiment after the CR TF knockout. This analysis demonstrated a strong correlation between predicted and actual target gene response (**Boxes 5,6**). Although not every gene showed the predicted response, this is expected for several reasons. First, there may be substantial sample-to-sample differences in TF regulators of some genes (60), and such divergence is especially likely in a cell line model. Second, a complete TF knockout may produce effects that are different from physiologic TF dose variation on which our model is built. Third, there are mathematical and practical limitations that cannot be avoided, such as the impact of post-transcriptional regulation. For example, protein levels of regulating TFs may differ from their mRNA expression levels, which will necessarily limit CORENODE's predictive power. However, despite these limitations, most discrepancies are observed at the lower DD and ES values and the higher confidence edges tend to be highly accurate.

Given that CR TFs are essential for cell survival, we calibrated our RNA-seq experiments to synthetic spike-in constructs in order to compensate for any global changes in mRNA expression (61,62). As expected, we observed signs of a global transcriptional collapse following MYB knockout; the increase in the MHC type II expression was preserved after adjusting for this global change in the mRNA pools. However, the genome-wide mRNA response to MYB knockout displayed a better alignment with CORENODE predictions when

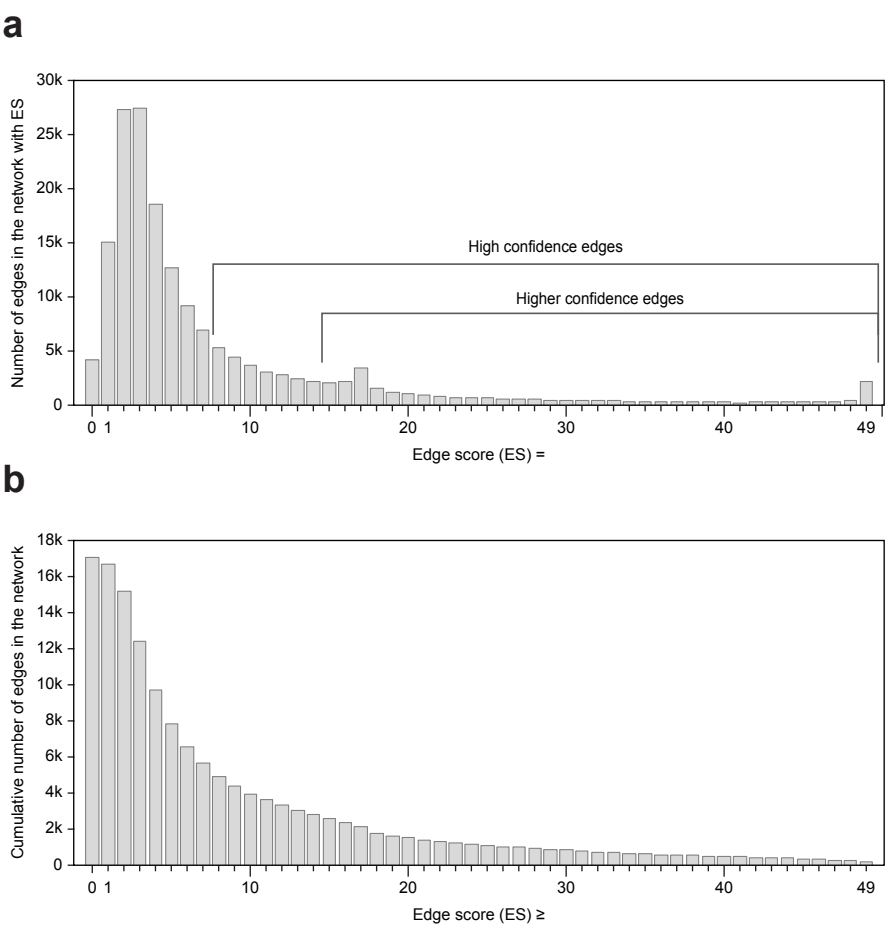


Box 6. Genome-wide validation of CORENODE edges: MYB. a, Correlation between directionality derivative (DD) and actual $\log_2$ fold change 72 hours after TF knockout measure by RNA-seq, genome-wide (*left*) and for high-confidence edges (ES ≥ 15 and [DD] ≥ 0.2, *right*), with Pearson correlation (*r*) for each plot. RNA-seq data were processed in DEseq2 and normalized using the standard median of ratios method (*uncompensated*) or to synthetic ERCC spike-in control to account for the global transcriptional collapse (*compensated*). **b**, Average $\log_2$ fold change in mRNA expression: transcripts are binned according to the directionality derivative (DD). The analysis includes all expressed genes regardless of the edge score. Use of uncompensated data (without spike-in normalization) results in a more precise prediction of the inflection point between positive and negative regulation.

spike-in calibration was omitted (**Box 6**). Indeed, CORENODE is based on steady-state variations of TF levels that, unlike a full TF knockout, do not result in secondary transcriptional collapse. Therefore, knockout experiments likely overestimate physiologically relevant global responses to naturally occurring changes in the levels of critical lineage TFs. More specifically, depending on the time of measurement, many mRNAs with early positive response to TF deprivation will appear downregulated after a global transcriptional

collapse ensues. Strikingly, CORENODE accurately predicts the inflection point between negative and positive regulation in uncompensated data (**Box 6b**), providing additional proof that, while spike-in normalization is of critical importance for accurate assessment of global transcriptional response, it requires careful interpretation.

Spike-in was also used in IRF8, MEF2C and MEIS1 knockout but no global transcriptional collapse was observed upon analysis of spike-in read ratios and the RNA-seq data was normalized by DEseq2 in a regular fashion for these TFs knockouts.



Box 7. Distribution of edge scores in the CORENODE network. **a**, Number of edges with specific ES value. **b**, Cumulative number of edges in the network above progressively increasing ES cutoff.

Genome-wide transcription network decomposition and graphic representation

To build a genome-wide map of TF-gene edges, we started with the ES matrix and converted it to a ternary network representation, where all edges above a certain ES cutoff receive a value of 1 or -1, when DD is >0 or <0, respectively, and edges below the cutoff receive a value of 0. We used an ES cutoff of 8 (rounded network-wide mean) to assign high confidence edges (**Box 7**). For some analyses we used a more conservative ES cutoff of 15 for highest confidence. The resulting matrix was analyzed in Morpheus (https://software.broadinstitute.org/morpheus/) as shown in Supplementary Fig. 7.

To build an optimal graphic representation of the network in Fig. 4B, we used an ES cutoff of 8 to create a binary matrix similar to what is described above, except all edges above the cutoff receiving a value of 1 regardless of the DD. The resulting matrix was clustered using the R routine “kmeans”, with eight centers and nstart=5. The eight clusters were placed in random order around an outer circle of radius one, spaced proportionally to cluster size. TF-to-cluster edges were then considered only for edges of sufficient strength, as indicated by the cluster center value for that cluster/TF pair from the kmeans clustering being at or above the median of all cluster center values. Each TF was then placed at the geometric median point of the cluster centers to which it had strong edges, using R routine “geo_median” from the package “pracma” to calculate the geometric median; this point minimizes the sum of the distances from each TF to its connected clusters (solving the unweighted Weber problem). The process is then repeated for all permutations of ordering the eight outer clusters. Finally, we picked the permutation that results in the shortest aggregate distance of the TF-to-cluster edges. By minimizing the aggregate distance between the TFs and gene cluster centers, we hoped to accomplish several goals: (1) the representation is provably optimal in terms of minimizing edge distances, (2) the graphic is easily understandable and unbiased, and (3) the concept is extendable to a weighted representation, taking into account the relative strength of the TF-to-cluster connections, by moving to a solution of the weighted Weber problem. That the aggregate distance is minimized can be seen from a combination of the following considerations: (1) given the cluster placements, each TF is independently minimizing its aggregate edge distance using the geometric median point, which means that the sum across all the TFs is also minimized, and (2) we test all possible cluster placements, so no better option remains.

Evaluation of gene-regulatory equation terms

To evaluate the impact and significance of each term in the 3-, 4- and 5-mer gene-regulatory equations, we considered the following parameters: (1) the calculated regression coefficient, (2) a standard-error uncertainty estimate of each regression coefficient demonstrating whether the regression coefficient is significantly different from zero, (3) *t*-value (coefficient/standard error) and (4) *p*-value (probability that the *t*-

value is different from zero). Terms with the lowest *p*-value were considered to have the largest predicted impact on the overall fit. These statistics are summarized in Supplementary Data 6.

Comparison with ARACNe

We sought to compare the predictive power of CORENODE to ARACNe, a popular algorithm of network decomposition based on mutual information (63–65). Using BeatAML mRNA expression data as input, ARACNe prioritized the same 4 TFs (IRF8, MEF2C, MYB and MEIS1) as top predicted regulators of MHC-II expression in AML (**Box 8**). However, it should be noted that, since ARACNe is not a regression-based approach, it does not allow modeling of combinatorial and dose-response effects.

a

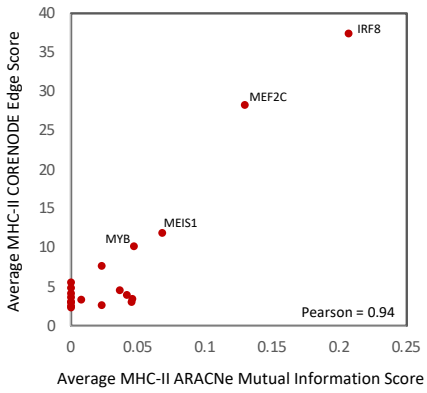
Corenode Edge Score

	HLA-DPA1	HLA-DRB1	HLA-DPB1	HLA-DQB1	HLA-DRA	HLA-DQA1	HLA-DMA	HLA-DMB	Average
IRF8	32	41	21	49	42	49	17	49	37.5
MEF2C	34	25	45	17	24	17	48	17	28.4
MEIS1	17	11	17	14	10	7	13	7	12
MYB	9	16	8	17	12	3	10	7	10.3
RUNX2	7	6	7	4	10	12	12	4	7.75
SREBF1	7	8	3	6	7	2	6	6	5.63
STAT5B	4	2	3	3	2	17	1	7	4.88
GFI1	3	4	3	3	3	4	2	15	4.63
ARID2	4	5	3	3	4	4	4	7	4.25
ZEB2	4	6	4	4	6	2	4	2	4
FOSL2	4	4	2	2	5	3	3	7	3.75
MEF2D	2	4	4	3	3	3	5	4	3.5
GFI1B	3	2	4	4	2	4	6	2	3.38
CEBPA	5	2	2	4	4	5	1	2	3.13
FLI1	2	2	3	3	2	4	6	3	3.13
RUNX1	3	3	10	2	2	3	1	1	3.13
SPI1	2	2	3	3	4	3	3	2	2.75
LYL1	3	2	2	3	2	3	2	4	2.63
E2F3	2	2	3	3	3	2	3	1	2.38

ARACNe Mutual Information Score

	HLA-DPA1	HLA-DRB1	HLA-DPB1	HLA-DQB1	HLA-DRA	HLA-DQA1	HLA-DMA	HLA-DMB	Average
IRF8	0.19	0.2	0.18	0.15	0.24	0.21	0.13	0.35	0.21
MEF2C	0.14	0.13	0.16	0.06	0.16	0.11	0.15	0.14	0.13
MEIS1	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.09	0.07
MYB	0.06	0.06	0	0.08	0.06	0	0.06	0.06	0.05
RUNX2	0	0	0.06	0	0.06	0	0.07	0	0.02
SREBF1	0	0	0	0	0	0	0	0	0
STAT5B	0	0	0	0	0	0	0	0	0
GFI1	0	0	0.06	0	0.05	0.05	0.06	0.07	0.04
ARID2	0	0	0	0	0	0	0	0	0
ZEB2	0.06	0.08	0	0	0.09	0	0	0.09	0.04
FOSL2	0	0	0	0	0	0	0	0	0
MEF2D	0.07	0.07	0.06	0	0.08	0	0.08	0	0.05
GFI1B	0	0	0	0	0.06	0	0	0	0.01
CEBPA	0.07	0.06	0.06	0.05	0.06	0	0	0.05	0.04
FLI1	0	0	0	0	0	0	0	0	0
RUNX1	0	0	0	0	0	0	0	0	0
SPI1	0	0	0	0.07	0.05	0	0	0.06	0.02
LYL1	0	0	0	0	0	0	0	0	0
E2F3	0	0	0	0	0	0	0	0	0

b



Box 8. Comparison between ARACNe and CORENODE. **a**, CORENODE Edge Scores and ARACNe Mutual Information Scores between the 19 core TFs and MHC-II genes. **b**, A cross-plot of average CORENODE Edge Scores and ARACNe Mutual Information Scores between the 19 core TFs and MHC-II genes.

SUPPLEMENTARY NOTE REFERENCES

49. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell. 2006;126:663–76.

50. Graf T, Enver T. Forcing cells to change lineages. Nature. 2009;462:587–94.

51. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell. 1987;51:987–1000.

52. Orkin SH, Hochedlinger K. Chromatin Connections to Pluripotency and Cellular Reprogramming. Cell. 2011;145:835–50.

53. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005;122:947–56.

54. Hnisz D, Schuijers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, et al. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. Mol Cell. 2015;58:362–70.

55. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155:934–47.

56. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;172:650–65.

57. Murphy KP. Machine Learning. MIT Press; 2012.

58. Allen DM. The Relationship Between Variable Selection and Data Agumentation and a Method for Prediction. Technometrics. 1974;16:125–7.

59. Seki A, Rutz S. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. J Exp Med. 2018;215:985–97.

60. Sorrells TR, Johnson AD. Making Sense of Transcription Networks. Cell. 2015;161:714–23.

61. Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. Molecular and cellular biology. 2015;36:662–7.

62. Lovén J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, et al. Revisiting global gene expression analysis. Cell. 2012;151:476–82.

63. Margolin AA, Nemenman I, Basso K, Klein U, Wiggins C, Stolovitzky G, et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. Bmc Bioinformatics. 2006;7:S7.

64. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. Nature genetics. 2005;37:382–90.

65. Lachmann A, Giorgi FM, Lopez G, Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics. 2016;32:2233–5.