

# Chromatin loop dynamics during cellular differentiation are associated with changes to both anchor and internal regulatory features

Marielle L. Bond<sup>1</sup>, Eric S. Davis<sup>2</sup>, Ivana Y. Quiroga<sup>3</sup>, Michael I. Love<sup>4,5,6</sup>, Hyejung Won<sup>5,7,+</sup>, Douglas H. Phanstiel<sup>1,2,3,6,8,+</sup>

<sup>1</sup> Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, NC, USA

<sup>2</sup> Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, NC, USA

<sup>3</sup> Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC, USA

<sup>4</sup> Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA

<sup>5</sup> Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

<sup>6</sup> Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

<sup>7</sup> Neuroscience Center, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

<sup>8</sup> Department of Cell Biology and Physiology, University of North Carolina, Chapel Hill, NC, USA

<sup>+</sup>Corresponding authors: Douglas H. Phanstiel (douglas\_phanstiel@med.unc.edu) and Hyejung Won (hyejung\_won@med.unc.edu)

## ABSTRACT

3D chromatin structure has been shown to play a role in regulating gene transcription during biological transitions. While our understanding of loop formation and maintenance is rapidly improving, much less is known about the mechanisms driving changes in looping and the impact of differential looping on gene transcription. One limitation has been a lack of well powered differential looping data sets. To address this, we conducted a deeply sequenced Hi-C time course of megakaryocyte development comprising 4 biological replicates and 6 billion reads per time point. Statistical analysis revealed 1,503 differential loops. Gained loops were enriched for AP-1 occupancy and correlated with increased expression of genes at their anchors. Lost loops were characterized by increases in expression of genes within the loop boundaries. Linear modeling revealed that changes in histone H3 K27 acetylation, chromatin accessibility, and JUN binding in between the loop anchors were as predictive of changes in loop strength as changes to CTCF and/or cohesin occupancy at loop anchors. Finally, we built linear models and found that incorporating the dynamics of enhancer acetylation and loop strength increased accuracy of gene expression predictions.

## INTRODUCTION

The three dimensional (3D) organization of chromatin is thought to play an important role in transcriptional regulation and has been implicated in many biological processes, including cellular differentiation and response to external stimuli<sup>1</sup>. While several types of 3D chromatin structures exist, chromatin loops are of particular interest as they are thought to regulate gene expression by bringing distal regulatory elements (e.g. enhancers) into close physical proximity with gene promoters via point to point interactions. Indeed, loop anchors are typically enriched for enhancers and promoters and correlate with differences in gene expression<sup>2-4</sup>. Aberrations to chromatin looping are associated with a variety of human diseases and developmental disorders such as Cornelia de Lange syndrome<sup>5</sup>, polydactyly<sup>6,7</sup>, and cancer<sup>8,9</sup>. While the basic mechanisms of loop formation have been established, major questions remain regarding the mechanisms driving differential looping during biological development and their functional impact.

The vast majority of chromatin loops are thought to form through a process called loop extrusion, in which the cohesin complex is loaded onto DNA and reels in

chromatin until it reaches convergently bound CTCF proteins<sup>10</sup>. In specific cell types and biological conditions chromatin loops can form through non-canonical mechanisms including phase separation<sup>8,11,12</sup> or binding of lineage-specific factors like LDB1<sup>8,11,12</sup>. While chromatin loops have been shown to change over cellular transitions, the mechanisms that govern these structural changes, and the impact of these changes on gene expression, remain poorly understood<sup>13,14</sup>.

The relationship between looping and gene expression is even less clear. Cell type specific loops correlate with differential expression patterns, supporting a role for loops in gene regulation<sup>15-17</sup>. Moreover, forced looping between enhancers and promoters at select loci has been shown to activate transcription<sup>18,19</sup>. However, several recent studies have called the role of loops in regulating gene expression into question. Live cell imaging of looping between SOX2 and an enhancer known to regulate its expression showed no correlation between enhancer-promoter proximity and gene transcription<sup>20</sup>. In another study, rapid and thorough degradation of cohesin led to a complete removal of cohesin-driven loops

in human cancer cells with only a minor impact on gene expression<sup>14</sup>. In summary, the degree to which chromatin looping regulates gene expression is still unresolved. One impediment to answering these questions is that identifying differential loops between cells and conditions remains challenging. Due to the depth of sequencing required for Hi-C data sets, the statistical requirements of differential analysis, and the cost of DNA sequencing, most existing differential looping studies lack the statistical power to adequately identify differential loops. And without comprehensive and rigorously defined sets of differential loops, it is challenging to determine what mechanisms drive differential looping and what transcriptional impact they have.

To address this gap, we generated a deeply sequenced Hi-C data set characterizing the differentiation of K562 cells into a megakaryocyte-like state. By sequencing over 18 billion reads across three timepoints and four biological replicates, we achieved a statistical power of roughly 0.932 and identified 1,503 differential loops. Generation and intersection with accompanying maps of chromatin accessibility, histone acetylation, transcription factor (TF) binding, and gene expression revealed insights into both the mechanisms and the functional impacts of differential looping during cellular differentiation. Interestingly, we find that regulatory features both at and between loop anchors correlate with changes in loop strength. Finally, we show that incorporating H3 K27 acetylation and chromatin looping dynamics into linear models in addition to promoter acetylation improves predictions of changes in gene expression.

## RESULTS

### Differentiation of K562 cells induces large-scale changes to 3D chromatin structure across multiple scales

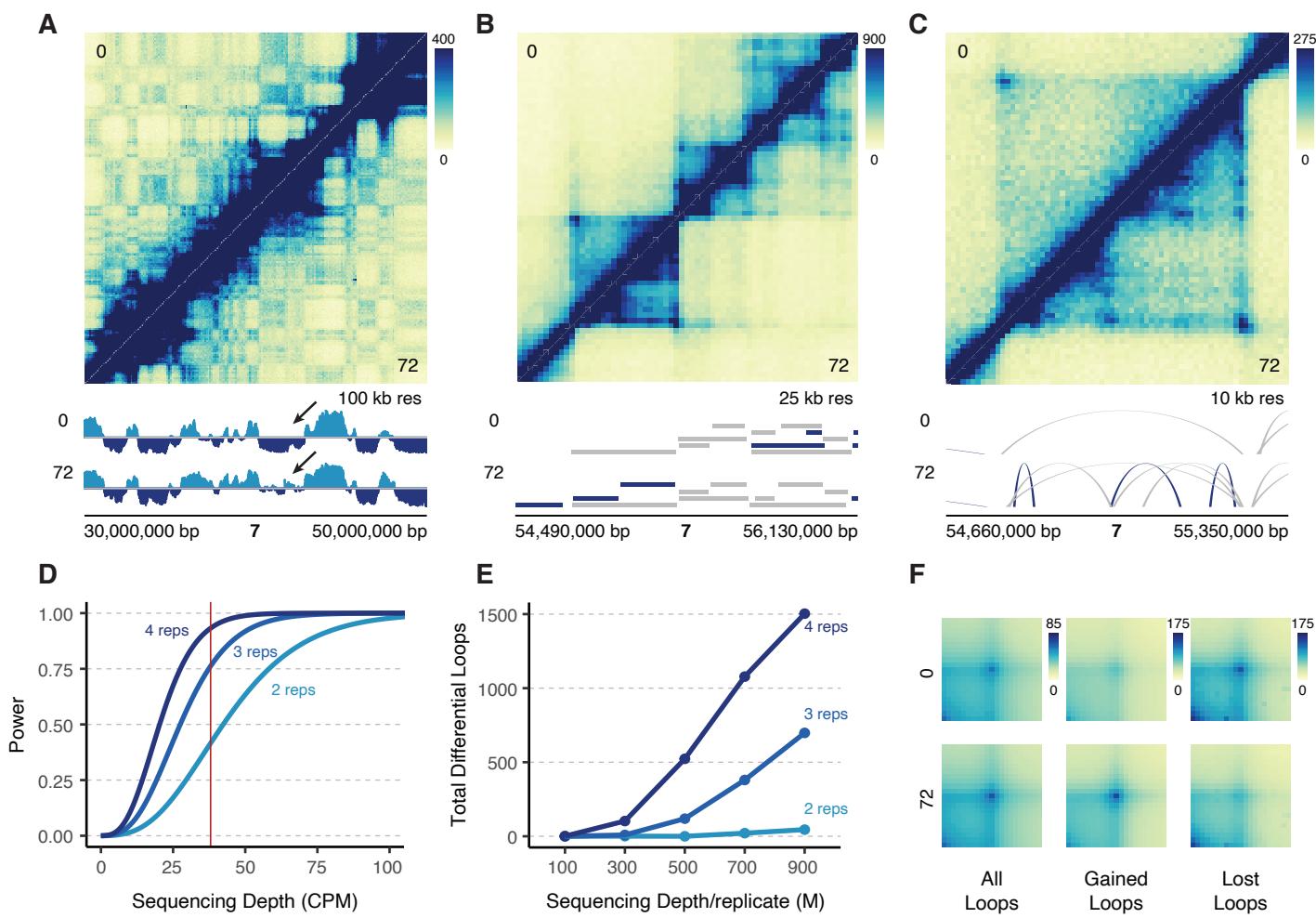
To understand how 3D chromatin structures change over cellular differentiation, we performed a deeply sequenced, 3 timepoint Hi-C time course tracking the differentiation of K562 cells into a megakaryocyte-like state<sup>21</sup>. We treated K562 cells with phorbol 12-myristate 13-acetate (PMA), which has been shown to induce a megakaryocyte-like phenotype<sup>21,22</sup>, for 0, 6, and 72h. We confirmed differentiation using qPCR for *ITGB3*, a megakaryocyte marker<sup>23</sup> (**Fig S1A**). As K562 cells differentiate into a megakaryocyte-like state, they lose their potential to differentiate into erythroid cells. We confirmed this with qPCR for *KLF1*, an erythroid marker, which decreases in expression over differentiation<sup>24</sup> (**Fig S1B**). We then performed *in situ* Hi-C on four biological replicates and sequenced them to a depth of roughly 6 billion reads per

time point (**Table S1**). We generated Hi-C contact maps using the Juicer pipeline<sup>15</sup>, identified compartments using the EigenVector package<sup>25</sup>, topologically associating domains (TADs) using arrowhead<sup>15</sup>, and chromatin loops using SIP<sup>26</sup> (**Fig S1C**). Replicates exhibited high similarity as measured by principal component analysis (PCA) (**Fig S1D**).

Visual inspection of the data revealed clear changes at multiple scales including nuclear compartments, TADs, and chromatin loops (**Fig 1A-C**). To assess the sequencing depth and replicates required to achieve sufficient statistical power, we analyzed our data set using the RNASeqPower package<sup>27</sup>. Using our dispersion of 0.0019 and median sequencing depth of 38 counts per million (CPM) per loop, the statistical power to detect 2-fold changes was 0.932, which is generally considered to be well-powered<sup>28</sup> (**Fig 1D**). We used the dispersion from our Hi-C data to model predicted statistical power across multiple different sequencing depths and numbers of replicates (**Fig 1D**). Holding sequencing depth per replicate constant, we found that decreasing to 3 or 2 replicates reduced the power estimates to 0.762 and 0.416 respectively. To determine how this increase in power translated into the number of differential loops, we subsampled our data to multiple depths and analyzed it with 2, 3, or all 4 replicates (**Fig 1E**). Using our full data set of four biological replicates (~900 million reads per replicate) we identified 1,503 differential loops using DESeq2<sup>29</sup> with a log-2 fold change (LFC) greater than 1.5 (adjusted p-value < 0.05) compared to only 698 and 45 identified with 3 or 2 replicates respectively (**Fig 1E**). This underscores how critical sequencing depth is in the sensitivity to detect differential loops. Finally, we confirmed the quality of these differential loops through aggregate peak analysis<sup>15</sup> (APA), showing that gained loops have higher contact at 72 vs 0 h, and lost loops have higher contact at 0 vs 72 h (**Fig 1F**).

### Genes at the anchors of gained, but not lost, loops exhibit concordant changes in expression

To assess the potential transcriptional impacts of differential loops, we performed RNA-seq across 8 time points (0, 0.5, 1.5, 3, 6, 24, 48, and 72 h) in K562 cells treated with PMA. DESeq2 analysis identified 3,190 differential genes (adjusted p-value < 0.05, LFC > 2) which were grouped into 6 clusters using k-means clustering (**Fig 2A**). The biggest and most unique changes were observed at 6 and 72 hours (1,619 differential, 527 unique genes after 6 h; 1,925 differential, 236 unique genes after 72 h), which coincide exactly with our selected Hi-C timepoints. Consistent with the differentiation of these cells into a megakaryocyte-like state, the upregulated genes were enriched for Gene Ontology (GO) terms relating to cell



**Figure 1. Deeply sequenced Hi-C experiments provide sensitive detection of differential chromatin loops.**

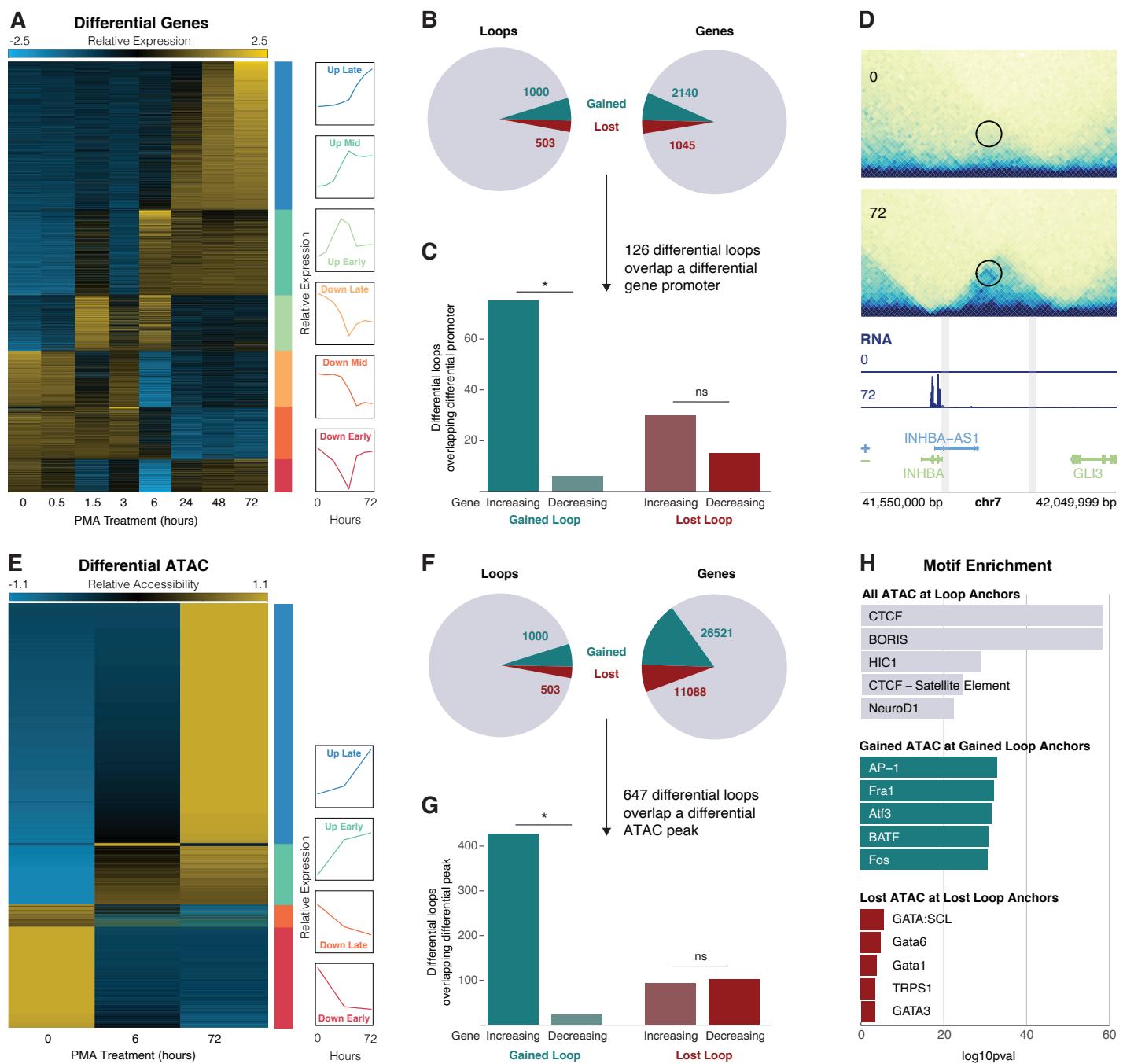
**(A)** 20 Mb region on chromosome 7 at 100kb resolution comparing K562s at 0h (top) to differentiated megakaryocytes at 72h (bottom). Signal tracks show compartmental eigenvector calls (light blue = compartment A, dark blue = compartment B). The arrow points to qualitative changes in compartmentalization. **(B)** Zoom in of a 1.6 Mb region of chromosome 7 at 25 kb resolution. TAD calls are indicated by ranges below the Hi-C map for each cell type (dark blue = cell type specific, gray = shared across cell types). **(C)** Zoom in of a 690 kb region in figure B at 10 kb resolution showing a region with differential loops. Arches indicate loop calls (dark blue = cell type specific, gray = shared across cell types). **(D)** Statistical power modeled across various theoretical sequencing depths at 2, 3, and 4 biological replicates. Red line indicates the median sequencing depth per loop (CPM: counts per million). **(E)** Actual number of differential loops called at multiple different subsampled sequencing depths for different numbers of replicates (M: millions). **(F)** Aggregate peak analysis for all loops, gained loops, and lost loops.

differentiation, cell adhesion, and morphogenesis, and pathways including focal adhesion, hematopoietic cell lineage, and regulation of actin cytoskeleton (**Fig S2A-B**). Upregulated genes include megakaryocyte markers *VWF*, *FLI1*, and *ITGB3*, further supporting acquisition of a megakaryocyte-like phenotype.

To determine the relationship between differential loops and gene transcription, we intersected the promoters of differential genes with differential loop calls (**Fig 2B**). We found that 9% (126 of 1,503) of differential loop anchors overlapped a differential gene promoter. Interestingly, 93% (75 out of 81) of gained loops that overlapped a differential gene promoter showed the same direction of change as the gene ( $p = 2.91 \times 10^{-16}$ , binomial test) (**Fig 2C**). This is consistent with our previous work in macrophages<sup>30,31</sup> and suggests that gained loops are relevant in increasing transcription of genes at their anchors. Upregulated genes found at the anchors of gained loops include *TGFB1* and *THBS1*, both of which are megakaryocyte-related<sup>32,33</sup>. An example of a gained loop with a concordant increase in gene expression is present at the *INHBA* locus (**Fig 2D**).

In contrast, lost loops that overlapped a differential gene promoter showed no such concordant behavior, with only 33% (15 out of 45) exhibiting the same directional change as the gene ( $p = 0.04$ , binomial test) (**Fig 2C**). These findings again agree with our previous work in macrophages<sup>30,31</sup>, where we did not see a significant decrease in expression of genes at lost loop anchors, suggesting that loss of looping is not sufficient to decrease transcrip-

tion. In fact, we found that genes at lost loop anchors were significantly more likely to be downregulated in differentiated megakaryocytes compared to K562s (Fig 2E). This is consistent with our previous work in macrophages<sup>30,31</sup> and suggests that loss of looping is not sufficient to decrease transcription. In fact, we found that genes at lost loop anchors were significantly more likely to be downregulated in differentiated megakaryocytes compared to K562s (Fig 2E). This is consistent with our previous work in macrophages<sup>30,31</sup> and suggests that loss of looping is not sufficient to decrease transcription.



**Figure 2. Gene expression and chromatin accessibility changes at differential loops.**

**(A)** RNA-seq normalized counts for all differential genes. Clusters are indicated by bars on the right side of the heatmap. Line plots show the mean expression per cluster. **(B)** Pie charts showing the proportion of differential loops (left) and genes (right). **(C)** Concordance analysis for the 126 differential loops that had a differential gene promoter at an anchor. Asterisks represent  $p < 0.05$  (binomial test). **(D)** Example region of a gained loop with an increased gene at the INHBA locus. **(E)** ATAC-seq normalized counts for all differential peaks. Clusters indicated by bars on the right side of the heatmap. Line plots show the mean expression per cluster. **(F)** Pie charts showing the proportion of differential loops (left) and ATAC peaks (right). **(G)** Concordance analysis for the 647 differential loops that had a differential ATAC peak at a promoter. Asterisks represent  $p < 0.05$  (binomial test). **(H)** TF motif enrichment analysis on all ATAC peaks at all loops (top), concordant gained ATAC peaks at gained loop anchors (middle), and concordant lost ATAC peaks at lost loop anchors (bottom).

tional output<sup>30,31</sup>. This is also consistent with work by Rao et al that found almost no change in gene transcription following virtually complete abrogation of loop-extrusion<sup>14</sup>. Taken together, this suggests that loss of looping is generally not sufficient to induce a decrease in transcriptional

output of genes at loop anchors. The anti-concordance we observed at loop anchors was also observed when looking at genes found between loop anchors. 45% of lost loops (228 out of 503) had a differential gene between their anchors, 64% of which (146 out of 228) overlapped

a gene that was increasing ( $p = 2.69 \times 10^{-5}$ , binomial test) (**Fig S2C**). Genes that overlapped the interior of lost loops were expressed at significantly higher levels compared to genes within gained loops ( $p = 1.85 \times 10^{-23}$ , Wilcoxon rank sum test) (**Fig S2D**). This also agrees with our work in macrophages<sup>30,31</sup> showing that extremely high expression of genes within loop boundaries was associated with a weakening of the loop, suggesting that high transcription at loop interiors might antagonize loop extrusion. Indeed, several other studies provide evidence that transcription can serve as a barrier to and/or interfere with loop extrusion<sup>34-36</sup>.

### Gained loops are associated with increased accessibility at AP-1 motifs

To assess which TFs were involved in loop based regulation, we mapped chromatin accessibility using ATAC-seq in K562 cells treated with PMA for 0, 6, and 72 h. Differential chromatin accessibility analysis with DESeq2 revealed 37,609 differential peaks (adjusted p-value, LFC  $> 2$ ), which we grouped into 4 clusters based on whether the peak reached maximal or minimal normalized counts at 6 or 72h (**Fig 2E**). As was the case with gene expression, differential chromatin accessibility peaks were highly concordant at the anchors of gained, but not lost, loops (**Fig 2F-G**).

To understand the mechanisms driving differential looping we performed TF motif enrichment on various sets of ATAC peaks at loops anchors (**Fig 2H**). As expected CTCF was the most enriched motif at ATAC peaks overlapping all loop anchors, which is consistent with its known role in loop formation and maintenance. In contrast, peaks of gained chromatin accessibility at gained loop anchors were highly enriched for Activator Protein 1 (AP-1) family members. This is consistent with our previous work showing the enrichment of AP-1 at the anchors of gained loops during macrophage development<sup>30</sup>. Peaks of decreased chromatin accessibility at lost loop anchors were enriched for Gata family members, albeit to a far lesser degree compared to the enrichments observed at all or gained loops. GATA TFs have a well established role in mediating cell fate decisions in myeloid cell development<sup>37-39</sup>.

### Differential looping is associated with chromatin features both at and between loop anchors

To gain further insight into the mechanisms driving differential looping, we generated matched CUT&RUN data-sets for multiple TFs and histone modifications. Due to the enrichment of AP-1 motifs at the anchors of gained loops, we performed CUT&RUN for JUN, a member of the AP-1 family. We also targeted histone H3 K27 acetylation, which is commonly used to identify putative enhancers, and CTCF and RAD21, proteins with known roles in DNA

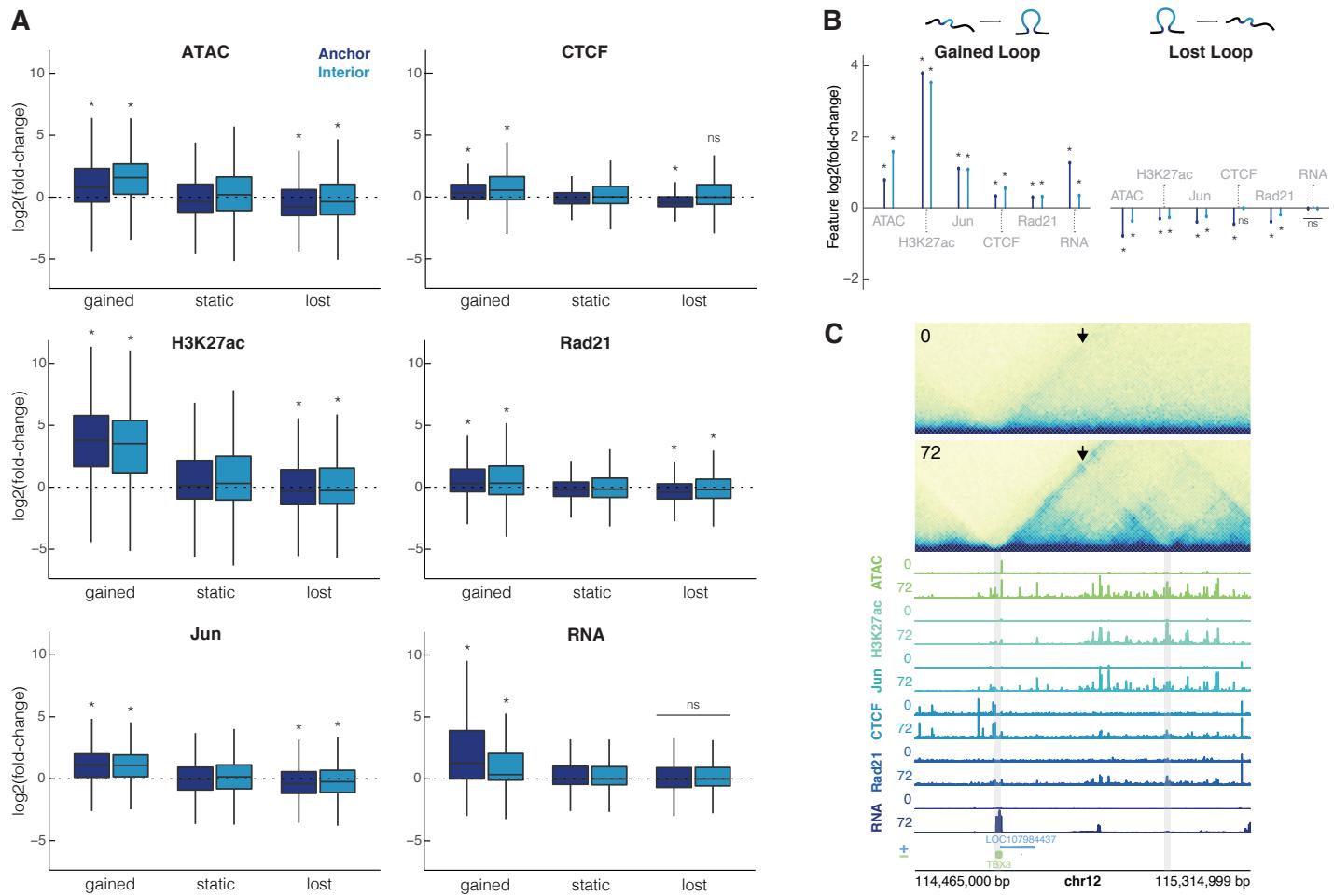
looping. For each assay, we performed differential analysis and compared how each differential feature corresponds to differential looping (**Fig S3A-D**). For all features, we observed statistically significant concordance at gained loop anchors (**Fig 3A-B**). With the notable exception of RNA, all features were also concordant at lost loop anchors, albeit to a lesser degree.

Investigating these trends more closely revealed two prominent features. First, while all features increased at gained loop anchors, some exhibited bigger changes than others. Surprisingly, the canonical mediators of chromatin looping, CTCF and RAD21, increased by only 1.26- and 1.23-fold respectively at gained loop anchors. In contrast, histone H3 K27ac increased nearly 14-fold at gained loop anchors (median = 13.86-fold). Chromatin accessibility and JUN binding increased by modest amounts, 1.72- and 2.16-fold respectively. Second, while increased signals were observed at gained loop anchors, they were also observed between loop anchors. For example chromatin accessibility and histone H3 K27ac peaks within the boundaries of gained loops increased by 2.99- and 11.52-fold respectively.

Examples of these trends are evident at the *TBX3* locus (**Fig 3C**). Gained enhancer-promoter looping is associated with increases in CTCF and RAD21 occupancy at loop anchors as well as large gains in chromatin accessibility and JUN occupancy at, between, and even beyond loop boundaries. Expression of *TBX3* itself, a TF known to regulate developmental transitions<sup>40</sup>, increases by over 50 fold. Taken together these results suggest that differential looping may involve more than just alterations to CTCF and RAD21 occupancy. They may also be mediated by chromatin modifying proteins and condition-specific TF binding events that act both at loop anchors and within the loop interior.

### Changes in chromatin features predict changes in chromatin looping

To explore this further and determine which chromatin features are the most predictive of changes in chromatin looping, we investigated how changes in each feature correlated with changes in looping (**Fig 4A**). Despite the relatively small fold change of CTCF peaks at differential loop anchors, we found that CTCF occupancy changes had the highest correlation ( $R^2 = 0.189$ ) with loop log2 fold-change (**Fig 4A**), consistent with the role of CTCF in the formation and maintenance of loops. However, changes to multiple features in the loop “interior”—the region in between the two anchors—also strongly exhibited strong correlations with differential looping. Interior chromatin accessibility, histone H3 K27 acetylation, and JUN occupancy had correlations of 0.175, 0.154, 0.127, followed by anchor RAD21 occupancy ( $R^2 = 0.117$ ). Several features



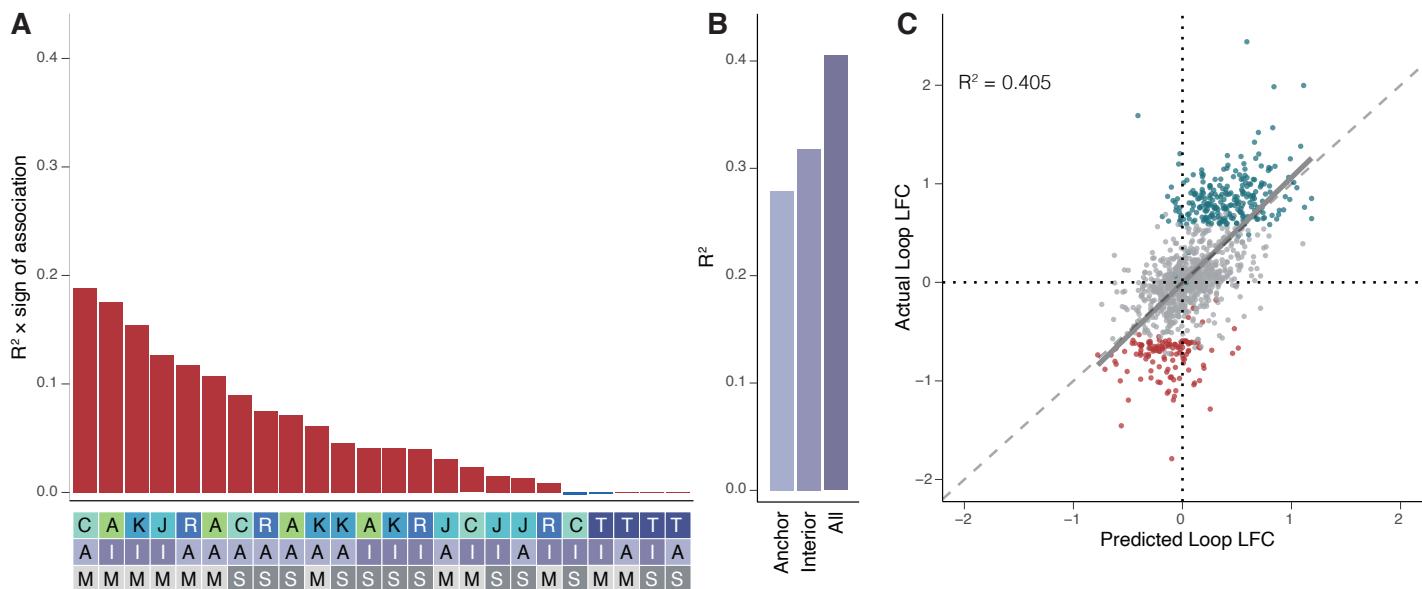
**Figure 3. Changes in chromatin features are correlated with changes in looping.**

**(A)** Intersections of each feature at the lotop anchors and interior for gained, static, and lost loops (dark blue = anchor, light blue = interior). All plots are on the same scale for the y-axis, showing  $\log_2(\text{fold-change})$ . Wilcoxon rank sum test was performed for each feature to compare gained/lost anchors to static anchors and gained/lost interiors to static interiors, asterisks represent  $p < 0.05$ . **(B)** Median unshrunken  $\log_2(\text{fold-change})$  of each dataset at gained loops (left), and lost loops (right). Asterisks represent  $p < 0.05$ , dots represent  $p > 0.05$ . **(C)** 600 kb region around a loop at the TBX3 locus at 10 kb resolution. The arrow is pointing to the gained loop. Signal tracks for ATAC-seq, H3K27ac, Jun, CTCF, Rad21, and RNA for K562s at 0h and differentiated megakaryocytes at 72h show increased occupancy for all features. Gray bars indicate 10 kb loop anchors.

exhibited very slight negative correlations including interior gene expression and CTCF occupancy, which is consistent with each of these having the ability to antagonize loop extrusions as previously described<sup>13,34,41</sup>. In addition to correlations with changes in looping, many of the features are highly correlated with each other (**Fig S4**).

Given the relationships between changes in looping and multiple features both at and between loop anchors, we next asked if we could build a better model to predict differential looping using multiple features simultaneously. We used the caret package to perform LASSO regression<sup>42,43</sup>. We selected 1,127 differential loops and 2,254 non-differential loops that were matched for distance and contact frequency as our training set and held out 376 differential loops and 752 non-differential loops as a testing set to evaluate our model. We then used LASSO with 10 cross-validations to generate a predictive mod-

el using various sets of features (**Fig 4A**) and evaluated it by applying it to our test set. Using all anchor features improved the correlation to  $R^2 = 0.279$ , far higher than using the best single feature alone (CTCF anchor max,  $R^2 = 0.189$ ; **Fig 4B**). Surprisingly, using only interior features yielded an even more accurate model with an  $R^2$  of 0.317 (**Fig 4B**). Combining all features at both the anchors and interiors produced the strongest predictions with an  $R^2$  of 0.405 (**Fig 4B-C**). Further, the model accurately predicted the signs of 91% of gained loops and 72% of lost loops. The predictive power of histone H3 K27 acetylation, chromatin accessibility, and JUN occupancy is surprising, and especially so given that the highest correlations were observed within loop boundaries rather than at anchors themselves. This may suggest that epigenetic changes between anchor sites play a significant role in modulating loop strength.



**Figure 4. Changes in chromatin features predict changes in chromatin looping.**

**(A)**  $R^2$  multiplied by the sign of association for all possible features correlated individually against changes in all loops in the model (top). Heatmap (bottom) is a legend where the feature, position, and measure for each bar are reported. (features: A = ATAC, K = H3K27ac, J = JUN, C = CTCF, R = RAD21, T = RNA; position: A = anchor, I = interior; measure: M = max, S = sum). **(B)**  $R^2$  values calculated for the anchor only, interior only, or all feature models. **(C)** Scatterplot showing the predicted loop fold-change vs actual loop fold-change for the testing dataset for looping. (Gray = static loops, teal = gained loops, maroon = lost loops,  $R^2$  calculated for all loops included in the testing dataset).

### Changes in and histone acetylation and chromatin structure predict changes in gene expression

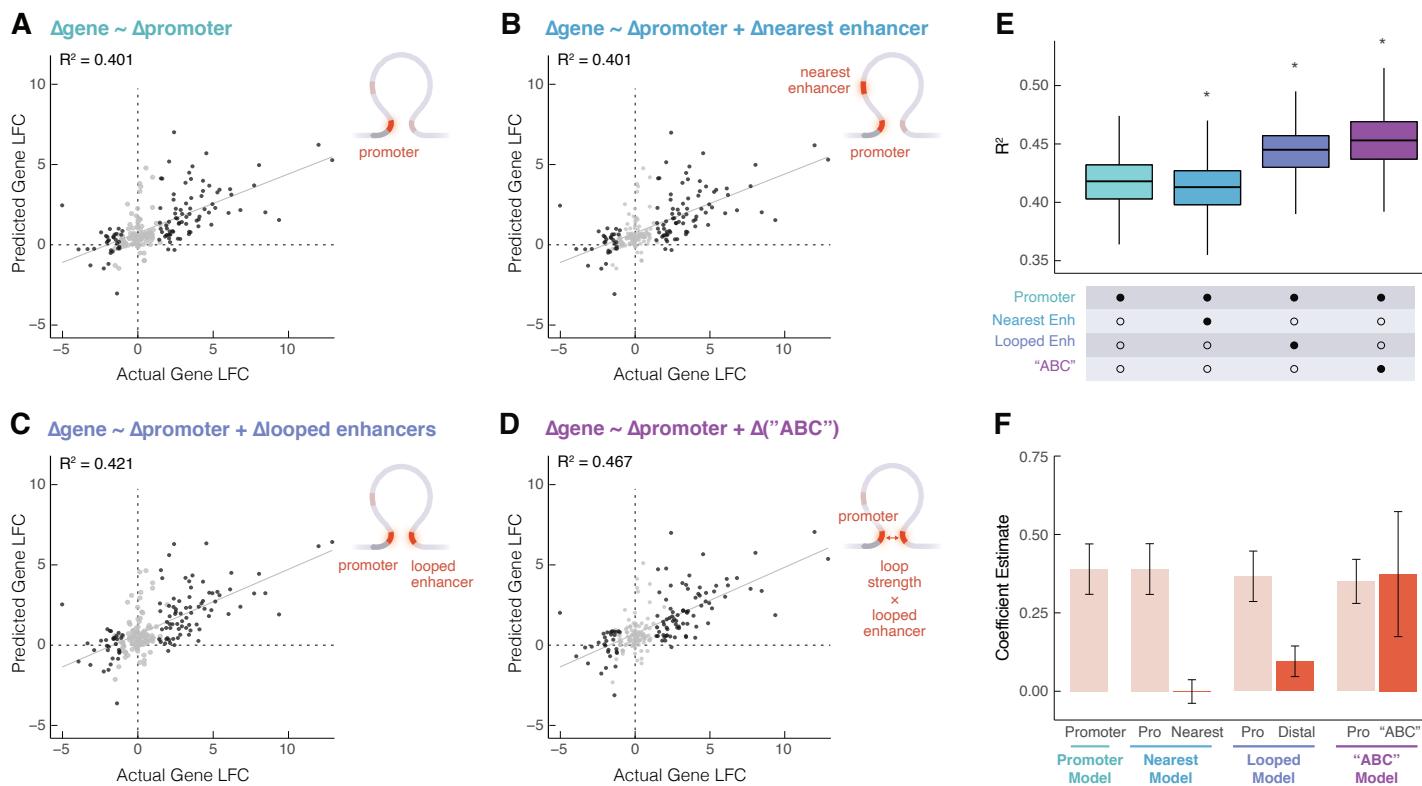
Predicting gene expression patterns from chromatin features is a long standing and difficult problem in the field of gene regulation. Recent advances have been made by incorporating both 2D (e.g. histone H3 K27 acetylation) and 3D (e.g. Hi-C contacts) features into the Activity-By-Contact model of gene regulation<sup>44</sup>. This model has been successful in assigning enhancers to their target genes but has typically been applied to resting cells rather than biological transitions. A notable exception is Beagan et al. that correlated changes to ABC score to changes in gene expression<sup>45</sup> albeit at only a handful of genomic loci. We leveraged this approach to determine if chromatin dynamics could help predict changes in gene expression in a genome-wide fashion.

To evaluate our ability to predict changes in gene expression we built and evaluated four linear models. In the first model, changes in gene expression were predicted based solely on changes in promoter H3 K27ac which performed well with an  $R^2$  of 0.401 (**Fig 5A**). In the second model, adding information from the nearest enhancer slightly decreased the predictive power (Wilcoxon rank sum test on permutations,  $p = 1.99 \times 10^{-6}$ ; median  $R^2 = 0.401$ ; **Fig 5B**). In contrast, in the third model, by combining H3 K27ac information from both the promoter and enhancers that were physically looped to the promoter, the  $R^2$  increased significantly to 0.421 (Wilcoxon rank sum test on permutations,  $p = 1.99 \times 10^{-122}$ ; **Fig 5B**). We

then calculated a modified ABC score (see methods) by taking the product of loop strength and distal enhancer activity for each looped enhancer-promoter pair, summing across all enhancers that were looped to each gene, and then calculating the fold-change. Finally, the last linear model built using both promoter histone H3 K27ac and change in ABC score increased the  $R^2$  even further to 0.467 (**Fig 5D**), a significant improvement compared to all other models (Wilcoxon rank sum test on permutations,  $p = 1.95 \times 10^{-182}$ ; **Fig 5E-F**). The improvement was even more drastic when building and applying the model to specifically differential genes and differential loops. For the models built on differential genes and loops,  $R^2$  values improve from 0.674 (promoter only model) to 0.763 for the promoter plus ABC score model (**Fig S5A-E**). These findings suggest that alterations to both enhancer activity and contact frequency can tune transcriptional programs during cellular differentiation.

### DISCUSSION

Collection and integration of deeply sequenced Hi-C and other genomic data characterizing the differentiation of K562 cells into a megakaryocyte-like state strengthened previous findings and provided novel observations into the mechanisms and impacts of changes to 3D chromatin structure. Our results confirmed canonical roles for CTCF and RAD21 in loop establishment and were consistent



**Figure 5. Changes in gene expression are explained by combined proximal and distal enhancer activity and loop strength.**

Scatter plots showing predicted gene-fold change vs actual gene fold-change based on (A) promoter H3K27 ac LFC alone, (B) promoter H3K27ac LFC and the nearest enhancer to the promoter's LFC, (C) promoter H3K27ac LFC and distal looped H3K27ac, and (D) promoter H3K27ac LFC and the modified ABC LFC (dark gray = differential gene, light gray = static gene). (E)  $R^2$  for each model calculated based on 1000 permutations of splitting data into training and testing sets. Wilcoxon rank sum test was performed to compare each group to the promoter only model, asterisks represent  $p < 0.05$ . (F) Estimates for each term in each model were calculated based on 1000 permutations of splitting the data into training and testing sets.

with chromatin looping playing a role in transcriptional activation and/or enhancement; however, these results also revealed strong correlations between differential looping and other regulatory features including chromatin accessibility, histone H3 K27 acetylation, and AP-1 occupancy and found a lack and/or anti-correlation of gene expression at lost loops. Intriguingly, differential looping correlated with transcriptional and regulatory features both at and between loop boundaries. In many cases, the correlations with internal features were stronger than the correlation with anchor signals. These results raise further questions regarding the mechanisms driving differential looping and point to previously underappreciated roles of molecules other than CTCF and cohesin.

The correlation between gained loops and the expression genes at their anchors agrees with many previous studies and supports the role of chromatin looping in gene activation. The increased predictive power when incorporating acetylation dynamics of looped enhancers and the dynamics of loop strength itself further empha-

size this point. In contrast, loss of looping does not coincide with decreased expression of anchor genes. In fact, more genes were increasing in expression than decreasing at the anchors of lost loops. This is consistent with our previous studies of macrophage development and activation, neither of which identified a decrease in expression at lost loop anchors<sup>30,31</sup>. Taken together, this suggests that while loops may be involved in transcriptional activation, loss of looping alone may not be sufficient for transcriptional repression or decreased expression. This agrees well with previous studies involving rapid depletion of cohesin and subsequent global loss of loop extrusion-driven loops. Rao et al<sup>14</sup> found that loop elimination did not substantially alter gene transcription—as measured directly using PRO-seq—in human colorectal cancer (HCT-116) cells. In contrast, loop disruption does inhibit activation of proinflammatory transcription in macrophages treated with LPS<sup>46</sup> (a bacterial cell wall component commonly used as a model for inflammatory activation), again suggesting that loops do play a role in gene activation. While

it is difficult to fully reconcile these findings, the evidence seems to be mounting that while increased looping can play a role in gene activation, events beyond loop disruption are required to decrease expression levels. One possible explanation is that looping imparts some sort of regulatory memory that is not erased as soon as the loop is disrupted. Similar mechanisms (e.g. the kiss-and-run mechanism)<sup>47</sup> have been suggested before. While these trends are coming into focus, more functional studies are required to understand the exact role of chromatin looping in gene activation and repression.

Our data also support the theory that loop loss may be a result, rather than cause, of changes in gene transcription. Differential genes within lost loops are significantly biased towards increased expression. And those increased genes tend to be fairly highly expressed. We observed this exact same phenomena in macrophages responding to LPS<sup>31</sup>. This is consistent with transcription occurring at a high level between loop boundaries being antagonistic to loop extrusion, something that is easy to imagine given that both involve fairly large molecular complexes traversing the same stretch of DNA. Indeed several previous studies have suggested that transcription can act as a molecular barrier to the loop extrusion process<sup>34–36</sup>.

Interestingly, we revealed that changes to multiple other chromatin features (i.e. chromatin accessibility, histone H3 K27 acetylation, and AP-1 occupancy) are roughly as predictive as changes to known loop extrusion-related proteins CTCF and RAD21. The correlation between looping and these other features was strongest for features within the loop interior which may provide clues into the nature of this relationship. One intriguing possibility is that increased accessibility, histone H3 K27 acetylation, and AP-1 binding might play a role in increasing the efficiency or rate of cohesin loading. This seems consistent with previous findings that suggested that cohesin loading takes place preferentially at active gene promoters<sup>4</sup> which are also associated with histone acetylation, TF binding, and chromatin accessibility. It is important to acknowledge that we can not rule out the possibility that all of these internal changes are the result, rather than cause, of differential looping events.

Despite the intriguing findings presented here, several limitations of this study must be considered when interpreting the data and speculating about their meaning. First, this study is largely correlative. While intersecting the results of this study can provide important mechanistic insights, several of the results raise new questions that must be addressed by future functional experiments. Second, in contrast to our previous work<sup>31</sup>, this time-course lacked the temporal resolution to put regulatory and transcriptional events in temporal order, which makes it difficult

to infer the direction of causality between any two features. This was a conscious decision as the current cost of sequencing makes it unfeasible to acquire deeply sequenced data sets across deeply sampled time courses; however, this is likely to change soon as sequencing costs continue to decrease<sup>48,49</sup>. Finally, while this study encompassed a broad number of regulatory features—including ATAC-seq, which when combined with motif analysis can provide insights into the binding patterns of hundreds of TFs—there are a vast number of other features that likely influence looping (e.g. DNA methylation) for which we are not measuring nor explicitly accounting for.

Despite these limitations, these findings improve our understanding of how different trans regulators and epigenetic features govern changes in looping, as well as our understanding of the relationship between looping and gene expression. The deeply sequenced nature of this differential Hi-C analysis offers a uniquely well-powered dataset with which to explore a pressing number of biological questions. Moreover, this data was acquired in one of the most widely studied human cell lines (K562) for which hundreds of publicly available genome-wide data sets are already available. As such, this study provides a valuable new resource for future studies of chromatin biology.

## DATA AVAILABILITY

Raw and processed data for Hi-C (GSE213909), RNA-seq (GSE213386), ATAC-seq (GSE213295), and CUT&RUN (GSE213908) are publicly available on GEO and SRA under SuperSeries 213909.

## FUNDING

This work was supported by NIH grant R35-GM128645 to D.H.P.. M.L.B was supported by NIH training grant T32-GM135128. E.S.D. was supported by NIH Training Grant T32-GM067553. I.Y.Q. was supported by the Bright-Focus Postdoctoral Fellowship 911831. M.I.L. was supported by NIH grant R01-MH118349. H.W. was supported by NIH grant DP2MH122403.

## ACKNOWLEDGMENTS

We thank Erika Deouedes for graphic design of figures and preparation of the manuscript. We thank Sam Pattenden for allowing us to use the Covaris LE220 instrument. We thank Katie Reed for assistance with project design and intellectual discussions. We thank Nicole Kramer for the development of plotgardener and help with generating figures and data visualization.

## AUTHOR CONTRIBUTIONS

M.L.B. designed and performed most of the experiments, performed all of the computational analysis, and wrote the manuscript.

E.S.D. developed some of the software used and assisted with computational analysis.

I.Y.Q. prepared ATAC-seq libraries.

M.I.L. oversaw data analysis and linear modeling

H.W. acquired funding, supervised data analysis, and helped write the manuscript.

D.H.P. acquired funding, conceptualized the project, supervised experiments and data analysis, and helped write the manuscript.

## METHODS

### K562 culture and differentiation

K562 were cultured in RPMI media (Corning, cat # 10-040-CV) with 10% fetal bovine serum (FBS) (Gibco, cat # 26140079) and 1% penicillin-streptomycin (PS) (Gibco, cat # 15140122). For megakaryocyte differentiation, K562 were plated in either 6-well plates (RNA-seq, ATAC-seq) or T-175 flasks (Hi-C, CUT&RUN) at a density of  $1 \times 10^5$  cells/mL and treated with 25 nM PMA (Sigma-Aldrich, cat # P1585-1MG). After 24h, the cells become semi-adherent. Cells were provided with fresh media and PMA after 24h and 48h. Cells were collected without treatment or after 6 or 72h. For all treatments and library preparations, K562s were thawed and immediately split into two T-25 flasks to create biological replicates.

### Crosslinking

Cells were cultured in T-175 flasks containing  $10 \times 10^6$  cells for Hi-C and  $5 \times 10^6$  cells for CUT&RUN at  $1 \times 10^5$  cells/mL. The collection protocol differs for where the cells are at during the differentiation. For suspended cells (0h and 6h): Media was centrifuged at 300 x g for 5 min. Cells are in suspension for 0h and 6h. Suspended cells were collected and spun down 300 x g for 5 min. Pellets were resuspended in 10 mL of 1% formaldehyde (Thermo Fisher Scientific, cat # 28908) in RPMI for 10 min with rotation. For semi-adherent cells (72h): The suspended cells were harvested and centrifuged for 300 x g for 5 min. Pelleted cells were resuspended in 10 mL of 1% formaldehyde, and added to the adherent cells in the T-175 flask to crosslink all cells at once, where they were put on the shaker for 10 min (from here, the protocol is the same for all time points). Cells were quenched with cold glycine for 5 min (Invitrogen, cat # 15527013) to a final concentration of 2.0 M for 5 min. Cells were then centrifuged at 562 x g, resuspended in cold PBS (Corning, cat #

21040CV), and split into 3 tubes of approximately  $3 \times 10^6$  cells each (HiC) or 10 tubes of approximately  $5 \times 10^5$  each (CUT&RUN). Cells were spun again at 562 x g for 5 min and washed again with cold PBS, then aspirated and flash frozen in liquid nitrogen, and stored at -80°C.

### In situ Hi-C library preparation

Four treatments (biological replicates) were performed. For each treatment, either two or three frozen pellets ( $3 \times 10^6$  each) were used to generate technical replicates (3 technical replicates for the first two biological replicates and 2 technical replicates for the last two biological replicates). Libraries were generated according to the protocol as described in Rao et al<sup>15</sup>. Briefly, crosslinked we lysed the cells, isolated nuclei, and used MboI (New England Biolabs, cat # R0147L) to digest chromatin overnight. The fragment ends were biotinylated, proximity ligated, and reverse crosslinked. Quantification of shearing DNA was achieved with Qubit (dsDNA Broad Range (BR) assay) (Thermo Fisher Scientific, cat # Q32850). We then sheared the samples on a Covaris LE 220 (duty factor 25, PIP 500, 200 cycles/burst, 90 seconds). 2% of each sample was run on a 2% agarose gel to confirm fragmentation. Size selection with AMPure XP beads (Beckman Coulter, cat # A63881) was then performed to select for DNA fragments between 300 and 500 bp. Biotinylated chromatin was pulled down with streptavidin beads. Biotin was then removed from unligated ends and the libraries were end repaired. We added the Illumina TruSeq Nano (Set A) (Illumina, cat # 20015960) indices to each sample in a combination appropriate for pooling and amplified using 9 cycles of PCR. Final quantification was achieved using Qubit (dsDNA High Sensitivity (HS) assay) (Thermo Fisher Scientific, cat # Q32851) and Tapestation (D1000 screentape) (Agilent, cat # 5067-5584). Libraries were pooled to 10 nM and sequenced across 7 Illumina NovaSeq S4 lanes (Novogene, 150-bp paired-end).

### RT-qPCR

We extracted RNA from  $5 \times 10^5$  cells using the QIA-GEN RNeasy Mini kit (Qiagen, cat # 74014) with DNase I treatment (Qiagen, cat # 79254) and quantified with a Qubit Broad Range assay (Thermo Fisher Scientific, cat # Q32850). Reverse transcription into cDNA was performed with the iScript cDNA synthesis kit (Bio-Rad, cat # 1708891). qPCR was performed with the TaqMan reagents using probes for ITGB3, KLF1, and GAPDH (Thermo Fisher Scientific, cat # Hs01001469, Hs00610592, Hs02786624).

### RNA-seq library preparation

We extracted RNA from  $5 \times 10^5$  cells using the QIAGEN RNeasy Mini kit with DNase I treatment. To confirm qual-

ity of libraries, we checked RNA integrity numbers with a Tapestation RNA screentape (Agilent, cat # 5067-5577) and confirmed them all to be above 9.7. We determined the concentration of all RNA samples with the Qubit Broad Range assay (Thermo Fisher Scientific, cat # Q10211).

The KAPA RNA HyperPrep kit with RiboErase (HMR) (Kapa Biosciences, cat # KK8560) was used for library preparation. Illumina TruSeq adapters (Illumina, cat # 20015960) were diluted and 0.0075 nmol was added to each sample. We determined library concentration and fragments size with Qubit (dsDNA HS assay) and Tapestation (D1000 screentape). Libraries from each timepoint were pooled at 10 nM and each biological replicate was sequenced on an Illumina NextSeq 500 (75-bp paired-end, high output kit).

### ATAC-seq library preparation

We used the Omni ATAC-seq protocol as described in Corces et al<sup>50</sup> with some adjustments to perform ATAC-seq. Two treatments (biological replicates) were performed. For untreated and 6h, cells were harvested and centrifuged at 500 x g for 5 min. For semi-adherent 72h cells, all of the floating cells were harvested. Adherent cells in each well were washed with 2 mL PBS, lifted with 500  $\mu$ L 0.5 M EDTA for 5 min, and quenched with 3 mL of RPMI before combining with the floating cells. 5 x 10<sup>5</sup> cells were used for library preparation. Illumina Nextera XT indices (Illumina, cat # FC-131-1001) (3.75  $\mu$ L/sample) were used for PCR.

After 5 PCR cycles, 5% of each sample was used in qPCR to determine how many more cycles were necessary. We found that 4-7 cycles were sufficient for final amplification. AMPure XP beads were used to perform a final cleanup (0.5X followed immediately by 1.3X) and quantified with the Qubit (ds DNA HS assay). The concentration in molarity of samples was determined by the KAPA Library Quantification Kit (Kapa Biosystems, cat # 4854). Each replicate was pooled to 4 nM and sequenced separately on an Illumina NextSeq 500 (75-bp paired-end, high output kit).

### CUT&RUN library preparation

We generated CUT&RUN libraries following existing protocols<sup>51</sup>, but modified for the use of crosslinked cells. Cells were centrifuged at 500 x g at 4°C for 10 min. For H3 K27ac, 0.5  $\mu$ L of 1:10 diluted antibody (Abcam, cat # ab4729) was added to each sample. For CTCF, 0.5  $\mu$ L of 1:10 diluted antibody (Thermo Fisher Scientific, cat # MA5-31344) was added to each sample. For JUN, 2.08  $\mu$ L of stock antibody (Thermo Fisher Scientific, cat # MA5-15172) was added to each sample. For RAD21, 0.625  $\mu$ L of stock antibody (Abcam, cat # ab992) was added to each sample. We then added 5  $\mu$ L of KAPA Unique Dual-Indexed Adapters (Roche, cat # 08861919702) diluted

to 750 nM. Libraries from each timepoint for H3 K27ac, CTCF, and JUN were pooled to 6 nM, and sequenced on an Illumina NextSeq 500 (75-bp paired-end, high output kit). RAD21 libraries from each timepoint were pooled to 9 nM, and were sequenced on an Illumina NextSeq 500 (75-bp paired-end, high output kit).

### Hi-C data processing and calling compartments, domains, and loops

We processed our Hi-C data using a modified version of the Juicer pipeline (version 1.9.8)<sup>15</sup>. Hi-C contact maps were generated at 5, 10, 25, 50, 100, 200, 250, 500, 1000, and 2500 kb resolution for each individual technical replicate that was sequenced. This was for 4 biological replicates, 3 timepoints, and 2-3 technical replicates each, totaling 30 unique samples. Additionally, all of the samples for each timepoint were merged to create merged Hi-C maps. All samples and replicates across all timepoints were also merged to create a "Mega" map.

Compartments were identified using the EigenVector R package at a 10 kb resolution<sup>25</sup>.

TADs were identified using the arrowhead command within the Juicer pipeline at 25 kb resolution. Cell type specific TADs were identified by merging with the mariner R package and using the denovo function.

Loops were called from the merged timepoint Hi-C files and Mega map with SIP<sup>26</sup> (version 1.6.1). The settings "-g 2 -5 2000 -fdr 0.05" were used both on the timepoint and the Mega map. Loops were merged in R with mariner using the mergeBedpe function, providing a list of 33,914 loops.

A count matrix was prepared using mariner<sup>52</sup>, where unnormalized counts at each loop pixel from each technical replicate were extracted.

The compartment, TAD, and loop-level Hi-C maps were SCALE normalized and visualized with plotgardener<sup>53</sup> at 100, 10, and 5kb resolutions respectively.

### Differential loop and aggregate peak analysis

DESeq2 was used to identify differential loops using the count matrix prepared as described above. Loops with a median count of 5 counts or less were filtered out. Counts from the technical replicates from each biological replicate were summed together and the design "rep + time" was used, with a reduced design of "rep" used to form a likelihood ratio test (LRT). Apeglm was used to calculate log-2 fold changes for each loop, comparing both 6 and 72h to 0h. Loops were deemed significant if they had an adjusted p-value < 0.05 and a log-2 fold change > 1.5.

Aggregate peak analysis (APA) was performed with mariner. For all, gained, and lost loops, the loop pixel and 10 pixels around the loop were extracted with SCALE normalization at 10 kb resolution.

## Hi-C Power Analysis

Power analysis was performed with the RNAPower package<sup>27</sup>. Dispersion was calculated from the differential loop analysis in DESeq2 as described above, where the minimum dispersion value was used. Power was modeled across various theoretical sequencing depths and replicates for identifying a log-2 fold change of 2 with a p-value of 0.05. The rnapower function was used with an alpha of 0.05/33914 and a cv of the square root of the dispersion value.

We subsampled our Hi-C data from the merged\_nodups files to approximate sequencing depths of 100, 300, 500, and 700M per biological replicate. We then repeated our differential loop analysis using the subsampled data for either 2, 3, or 4 replicates using all of the same parameters and loop calls.

## RNA-seq processing

Fastq quality was assessed using the FastQC and MultiQC tools (FastQC version 0.11.5, MultiQC version 1.5)<sup>54,55</sup>. Fastq files were trimmed with Trim Galore! (version 0.4.3) and quantified with Salmon (version 1.4.0) to the hg38 genome<sup>56,57</sup>. Alignment was performed using HISAT2 (version 2.1.0), which generated BAM files that were indexed using samtools (version 1.9)<sup>58,59</sup>. The BAM files for each timepoint's two biological replicates were merged using samtools and converted to bigwigs using deeptools (version 3.0.1)<sup>60</sup> for easy visualization of signal tracks. Reads were summarized into a format compatible with DESeq2 using tximport (r version 3.3.1, tximport version 1.2.0)<sup>29,61</sup>.

## Differential gene analysis

DESeq2 was again used to identify differential genes. The txi file was used as input, and the DESeqDataSetFromTximport was used with “~rep + time” as the design. A reduced design of “~rep” was used to form an LRT, as previously for peak analysis. Shrunken log-2 fold change values were calculated for each gene by comparing the counts at each time point to 0h with apegelm<sup>62</sup>. Significant genes had an adjusted p-value < 0.05 and a log-2 fold change > 2.

The DESeq2 dataset was normalized with variance stabilized transformation. We then filtered for differential genes and calculated Z-scores based on standard deviation and mean. Replicates were then averaged, and kmeans clustering was used to identify 6 temporal clustered based on the vectors of Z-scores.

## ATAC-seq processing and peak calling

Adapters were trimmed and low quality reads were filtered out using Trim Galore! (version 0.4.3)<sup>56</sup>. BWA mem (version 0.7.17) was used to align reads to the hg38 genome and sorted using Samtools (version 1.9)<sup>59</sup>. PicardTools

(version 2.10.3) was used to remove duplicate reads<sup>63</sup>. Mitochondrial reads were filtered out with Samtools idxstats<sup>59</sup>. For each timepoint, biological replicates were merged and indexed with Samtools. We called peaks were called on the merged files using MACS2 with the following parameters: -f BAM -q 0.01 -g hs --nomodel --shift 100 --extsize 200 --keep-dup all -B --SPMR (version 2.1.1.20160309)<sup>64</sup>. A comprehensive peak list was generated by merging peaks across all time points (181,136 peaks). For each peak across all biological replicates independently, counts were extracted with bedtools multicov, which was the input for differential peak analysis<sup>65</sup>. Signal tracks were generated from merged time points with deeptools (version 3.0.1) for visualization<sup>60</sup>.

## CUT&RUN processing peak calling

Adapters were trimmed off and low quality reads were filtered out with Trim Galore! (Version 0.4.3)<sup>56</sup>. BWA mem (version 0.7.17) was used to align reads to the hg38 genome, sorted with Samtools (version 1.9), and filtered for duplicates with PicardTools (version 2.10.3)<sup>59,63</sup>. Samtools was again used to index all BAM files. For each timepoint, biological replicates were merged and indexed with Samtools. We called peaks on the merged files using MACS2 with the following parameters: -f BAM -q 0.01 -g hs --nomodel --shift 0 --extsize 200 --keep-dup all -B --SPMR (version 2.1.1.20160309)<sup>64</sup>. A comprehensive peak list was generated for each antibody by merging peaks across all time points (X, Y, Z peaks for X, Y, Z datasets). For each biological replicate independently, counts were extracted with bedtools multicov, which was the input for differential peak analysis<sup>65</sup>. Signal tracks were generated from merged time points with deeptools (version 3.0.1) for visualization<sup>60</sup>.

## Differential ATAC-seq and CUT&RUN Peak Analysis

DESeq2 was again used to identify differential peaks from ATAC-seq and all CUT&RUN data. The count matrix generated in the ATAC-seq and CUT&RUN processing was used as the input with the function DESeqDataSetFromMatrix. We used “~rep + time” as the design and a reduced design of “rep” to form an LRT. Apeglm was used to calculate shrunken log-2 fold-changes for each peak for each dataset<sup>62</sup>. Significant peaks had an adjusted p-value of < 0.05 and an absolute log-2 fold change > 2. Peaks were clustered into up early, up mid, up late, down early, down mid, and down late by determining whether their max/min value was at 0, 6, or 72h.

## Gene Ontology and KEGG Pathway Enrichment Analysis

We used the HOMER function findMotifs.pl on each of our 6 gene clusters to identify enriched gene ontology terms and KEGG pathways with default settings<sup>66</sup>. For GO

Terms, the biological\_process.txt file was used and for KEGG pathways, the kegg.txt file was used.

## Motif Enrichment Analysis

We used the HOMER function findMotifsGenome.pl for all motif enrichment<sup>66</sup>. For motifs at the anchors of all loops, we intersected all ATAC peaks with all loop anchors. For motifs at the anchors of gained loops, we intersected all differential gained ATAC peaks with gained loop anchors. For motifs at the anchors of lost loops, we intersected all differential lost ATAC peaks with lost loop anchors. For each motif enrichment, all ATAC peaks were used as the background. The default parameters were used with the following adjustments: -size given.

## Genomic Intersections

The GenomicRanges and InteractionSet R packages were used to perform all genomic intersections<sup>67</sup>. Loop bedpe files were converted into GInteractions objects and were intersected with the coordinates for ATAC, H3 K27ac, JUN, CTCF, RAD21 peaks and genes with subset-ByOverlaps. The unshrunken fold-changes as calculated by DESeq2 analysis were extracted from each of the peaks that overlapped a loop anchor.

## Chromatin Looping Linear Model

The count matrices from previous analysis were again used to generate the data for the linear model. We used loop counts, peak counts from ATAC-seq and all CUT&RUN data, and transcripts per million (TPM) per 10 kb bin from RNA-seq. For all loops, each of these counts was extracted from both the anchors and each anchor individually. In the case of multiple peaks intersecting with loops, the sum of all peaks was recorded. We also recorded the maximum values from extracting counts from bigwig files instead. A final anchor measure was calculated by taking the product of signal (either sum or max) at both of the anchors for each loop. All interior measures were normalized to the length of the loop.

We added a pseudocount of 1000 to the entire data-frame (which is roughly 0.04 times the average count value) and then calculated the log(fold-change) between 72h and 0h. This "delta" matrix was then scaled, and DESeq2 log-2 fold-changes were used for looping. Our model consisted of all differential loops and twice as many static loops matched for distance and contact. Matched static loops were generated from the matchRanges<sup>68</sup> function within the nullranges Bioconductor package. 75% of the dataset was used for training and the remaining 25% was reserved for testing.

Each feature was tested against loop LFC with the base R function lm to determine R<sup>2</sup> values. The sign of correlation was determined with the cor function. LASSO regression was used to find a sparse model combining features, calling glmnet<sup>42</sup> within the caret R package<sup>43</sup>. We trained the LASSO model on the training set, using anchor features only. We evaluated selected LASSO models on the test set using R<sup>2</sup>. This was repeated again for all interior features, and with all anchor and interior features combined. The R<sup>2</sup> was calculated with the cor function. This was repeated for all interior features, and again repeated with all anchor and interior features combined.

## Gene Expression Linear Model

We used the lm function within the stats R package to model how gene expression changes correlate with changes in changes in proximal and distal acetylation, and looping. We used the GenomicRanges function subsetByOverlaps and linkOverlaps to determine which differential genes had promoter H3 K27ac and were looped to a distal H3 K27ac peak<sup>67</sup>, identifying 332 genes, and included 332 static genes matched for expression. We also identified the nearest enhancer with the nearest function of GenomicRanges. For proximal and distal H3 K27ac, we extracted the counts and calculated log2(-fold-change). For genes that had multiple enhancers, we took the sum of the counts at 0h and 72h and then calculated log2(fold-change). For the ABC score, we scaled all enhancer and loop counts to be between 1 and 100 to ensure that both factors were contributing equally to the interaction despite differences in sequencing depth. For each enhancer-promoter pair, we multiplied the normalized distal enhancer counts by the normalized loop strength counts, then calculated log2(fold-change). For genes with multiple enhancer-promoter pairs, we summed the multiplied score at 0h and 72h and then calculated log2(fold-change).

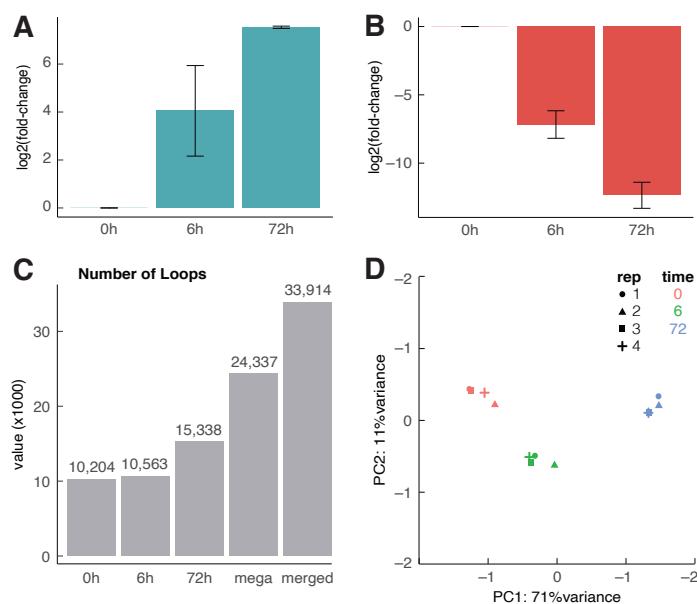
The first model only used changes in promoter acetylation to predict changes in gene expression. The following three models used promoter acetylation in addition to either nearest enhancer, distal enhancer, or the interaction between distal enhancers and loops. We trained on 60% of the data and tested on the remaining 40%. The predict function was used to predict on the testing data-set from the trained model and cor was used to calculate R<sup>2</sup> values. For R<sup>2</sup> and coefficient estimate calculations, we performed 1000 permutations of splitting the data into testing and training datasets.

## REFERENCES

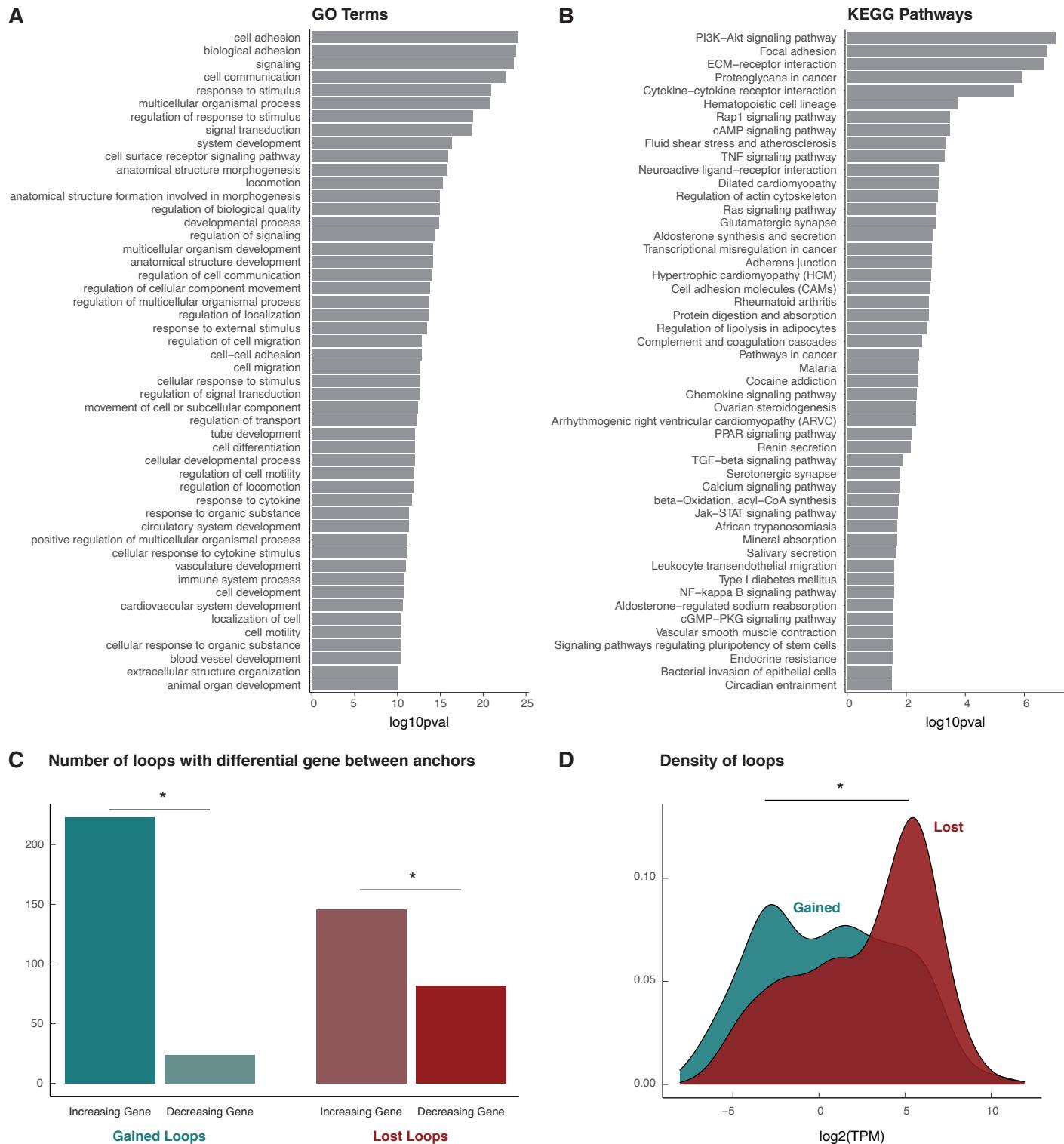
1. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
2. Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737–743 (2020).
3. Kim, A. & Dean, A. Chromatin loop formation in the  $\beta$ -globin locus and its role in globin gene transcription. *Mol. Cells* **34**, 1–5 (2012).
4. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
5. Panarotto, M., Davidson, I. F., Litos, G., Schleiffer, A. & Peters, J.-M. Cornelia de Lange syndrome mutations in NIPBL can impair cohesin-mediated DNA loop extrusion. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2201029119 (2022).
6. Paliou, C. *et al.* Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12390–12399 (2019).
7. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
8. Ahn, J. H. *et al.* Phase separation drives aberrant chromatin looping and cancer development. *Nature* **595**, 591–595 (2021).
9. Kon, A. *et al.* Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.* **45**, 1232–1237 (2013).
10. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6456–65 (2015).
11. Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448–453 (2019).
12. Krivega, I. & Dean, A. LDB1-mediated enhancer looping can be established independent of mediator and cohesin. *Nucleic Acids Res.* **45**, 8255–8268 (2017).
13. Jeppsson, K. *et al.* Cohesin-dependent chromosome loop extrusion is limited by transcription and stalled replication forks. *Sci Adv* **8**, eabn7063 (2022).
14. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24 (2017).
15. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
16. Hu, B. *et al.* Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat. Commun.* **12**, 3968 (2021).
17. Sey, N. Y. A. *et al.* Chromatin architecture in addiction circuitry identifies risk genes and potential biological mechanisms underlying cigarette smoking and alcohol use traits. *Mol. Psychiatry* **27**, 3085–3094 (2022).
18. Deng, W. *et al.* Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849–860 (2014).
19. Rupon, J. W. *et al.* Using Forced Chromatin Looping To Overcome Developmental Silencing Of Embryonic and Fetal  $\beta$ -Type Globin Genes In Adult Erythroid Cells. *Blood* **122**, 433 (2013).
20. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* **8**, (2019).
21. Huang, R. *et al.* Megakaryocytic differentiation of K562 cells induced by PMA reduced the activity of respiratory chain complex IV. *PLoS One* **9**, e96246 (2014).
22. Whalen, A. M., Galasinski, S. C., Shapiro, P. S., Nahreini, T. S. & Ahn, N. G. Megakaryocytic differentiation induced by constitutive activation of mitogen-activated protein kinase kinase. *Mol. Cell. Biol.* **17**, 1947–1958 (1997).
23. Pencovich, N., Jaschek, R., Tanay, A. & Groner, Y. Dynamic combinatorial interactions of RUNX1 and cooperating partners regulates megakaryocytic differentiation in cell line models. *Blood* **117**, e1–14 (2011).
24. Kuvardina, O. N. *et al.* RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. *Blood* **125**, 3570–3579 (2015).
25. *EigenVector*. (Github).
26. Rowley, M. J. *et al.* Analysis of Hi-C data using SIP effectively identifies loops in organisms from *C. elegans* to mammals. *Genome Res.* **30**, 447–458 (2020).
27. Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A. & Kocher, J.-P. Calculating sample size estimates for RNA sequencing data. *J. Comput. Biol.* **20**, 970–978 (2013).
28. Cohen, J. A power primer. *Psychol. Bull.* **112**, 155–159 (1992).
29. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
30. Phanstiel, D. H. *et al.* Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Mol. Cell* **67**, 1037–1048.e6 (2017).
31. Reed, K. S. M. *et al.* Temporal analysis suggests a reciprocal relationship between 3D chromatin structure and transcription. *bioRxiv* 2022.05.05.490836 (2022) doi:10.1101/2022.05.05.490836.
32. Villevie, J.-L. & Vainchenker, W. Megakaryocytes tame erythropoiesis with TGF $\beta$ 1. *Blood* vol. 136 1016–1017 (2020).
33. Wang, H. *et al.* Decoding Human Megakaryocyte Development. *Cell Stem Cell* **28**, 535–549.e8 (2021).
34. Brandão, H. B. *et al.* RNA polymerases as moving barriers to condensin loop extrusion. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20489–20499 (2019).
35. Gu, B. *et al.* Opposing Effects of Cohesin and Transcription on CTCF Organization Revealed by Super-resolution Imaging. *Mol. Cell* **80**, 699–711.e7 (2020).
36. Heinz, S. *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**, 1522–1536.e22 (2018).
37. Noh, J.-Y. *et al.* Inducible Gata1 suppression expands megakaryocyte-erythroid progenitors from embryonic stem cells. *J. Clin. Invest.* **125**, 2369–2374 (2015).
38. Stachura, D. L., Chou, S. T. & Weiss, M. J. Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. *Blood* **107**, 87–97 (2006).
39. Tijssen, M. R. & Ghevaert, C. Transcription factors in late megakaryopoiesis and related platelet disorders. *J. Thromb. Haemost.* **11**, 593–604 (2013).
40. Weidgang, C. E. *et al.* TBX3 Directs Cell-Fate Decision toward Mesendoderm. *Stem Cell Reports* **1**, 248–265 (2013).
41. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop-extruding cohesin complexes. *bioRxiv* 2022.01.07.475367 (2022) doi:10.1101/2022.01.07.475367.
42. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat.*

- Softw. **33**, 1–22 (2010).
43. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **28**, 1–26 (2008).
44. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
45. Beagan, J. A. et al. Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nat. Neurosci.* **23**, 707–717 (2020).
46. Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010).
47. Karr, J. P., Ferrie, J. J., Tjian, R. & Darzacq, X. The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer-promoter communication. *Genes Dev.* **36**, 7–16 (2022).
48. Almogy, G. et al. Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform. *bioRxiv* 2022.05.29.493900 (2022) doi:10.1101/2022.05.29.493900.
49. November, J. More than Moore's Mores: Computers, Genomics, and the Embrace of Innovation. *J. Hist. Biol.* **51**, 807–840 (2018).
50. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
51. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).
52. Davis, E. *mariner: Explore the Hi-C's.* (Github).
53. Kramer, N. E. et al. Plotgardener: Cultivating precise multi-panel figures in R. *Bioinformatics* (2022) doi:10.1093/bioinformatics/btac057.
54. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
55. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
56. Babraham Bioinformatics - Trim Galore! [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).
57. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
58. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
59. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
60. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).
61. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).
62. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
63. Picard. <https://broadinstitute.github.io/picard/>.
64. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
65. bedtools: a powerful toolset for genome arithmetic—bedtools 2.30.0 documentation. <https://bedtools.readthedocs.io/en/latest/>.
66. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
67. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
68. Davis, E. S. et al. matchRanges: Generating null hypothesis genomic ranges via covariate-matched sampling. *bioRxiv* 2022.08.05.502985 (2022) doi:10.1101/2022.08.05.502985.

## SUPPLEMENTAL FIGURES

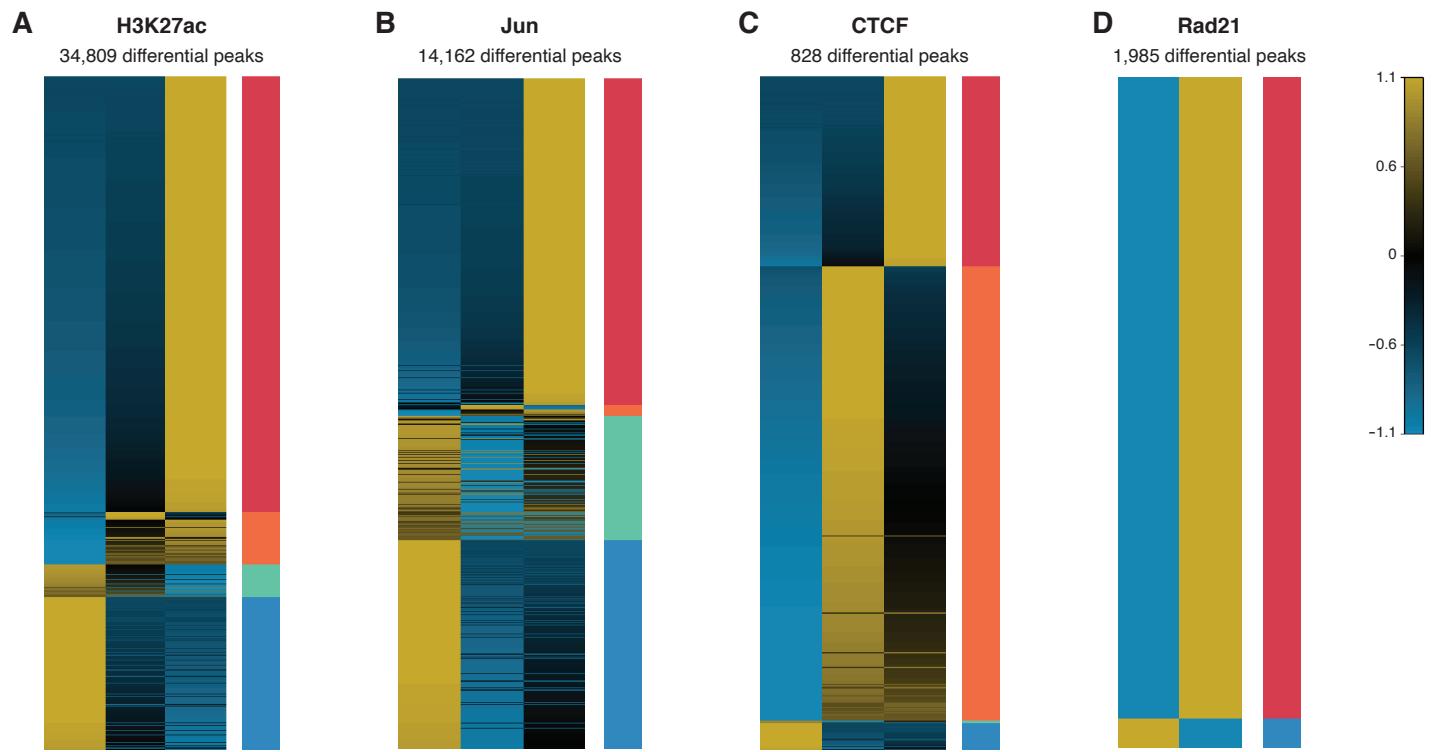


**Figure S1. Confirmation of megakaryocyte differentiation and detection of loops.**  
qPCR analysis of **(A)** ITGB3 and **(B)** KLF1 over megakaryocyte differentiation at 0, 6, and 72h. Two biological replicates and 4 technical replicates were collected, normalized to GAPDH levels, and log2(fold-change) was calculated relative to 0h. **(C)** Number of loops identified with SIP after 0, 6, or 72h of differentiation, after merging all timepoints together into the Mega map, and merging all timepoints together with mariner. **(D)** PCA plot showing similarities in loop counts between replicates and timepoints.



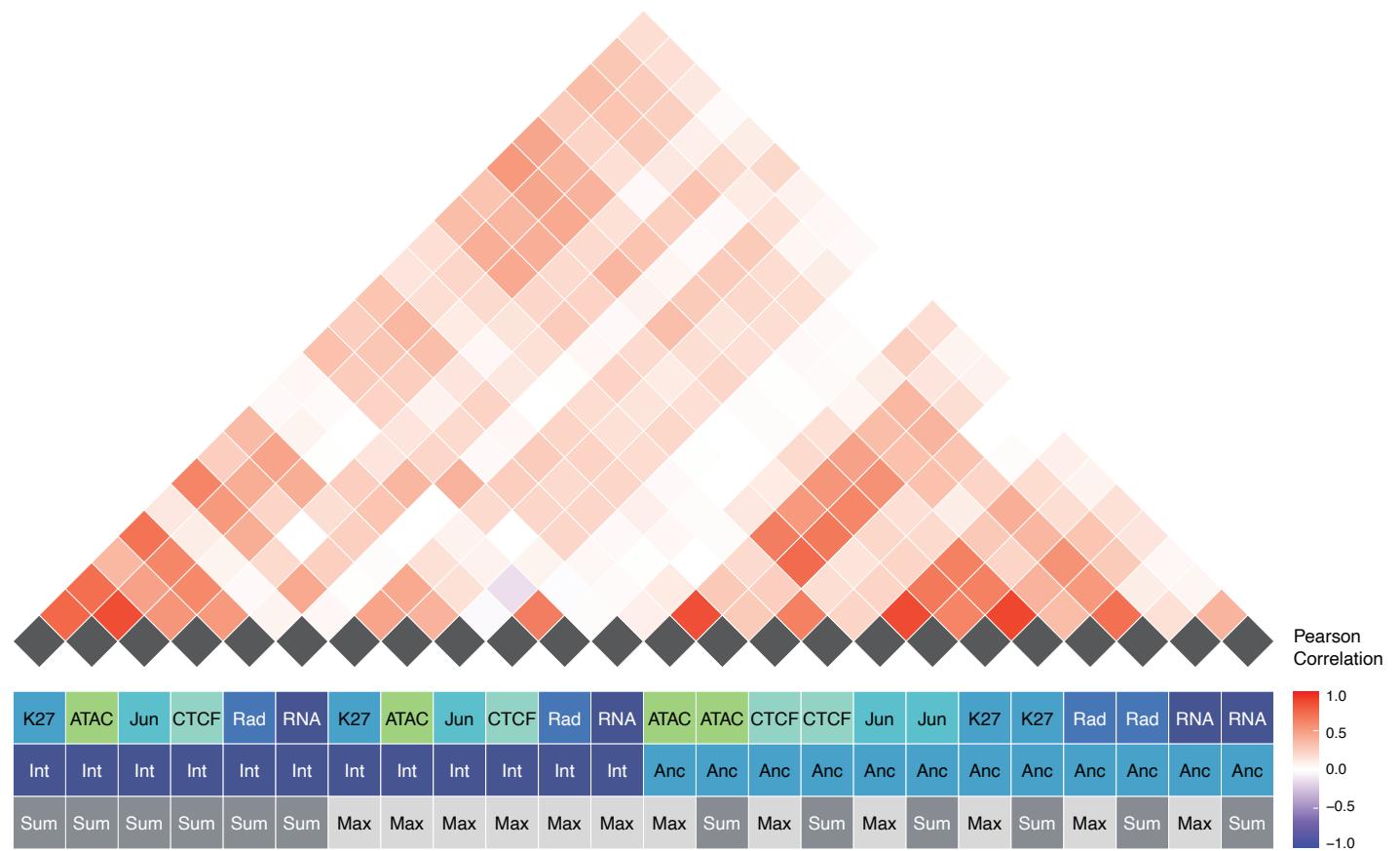
**Figure S2. Megakaryocyte pathway enrichment and interior gene expression.**

**(A)** Top 50 GO terms for upregulated genes from RNA-seq. **(B)** Top 50 KEGG Pathways for upregulated genes from RNA-seq. **(C)** Concordance analysis for the 475 differential loops that had a differential gene promoter between their anchors. Binomial test performed for each comparison, \*asterisks represent  $p < 0.05$ . **(D)** Expression of genes located between the anchors of gained and lost loops, \*asterisk represents  $p < 0.05$ . (TPM: transcripts per million).



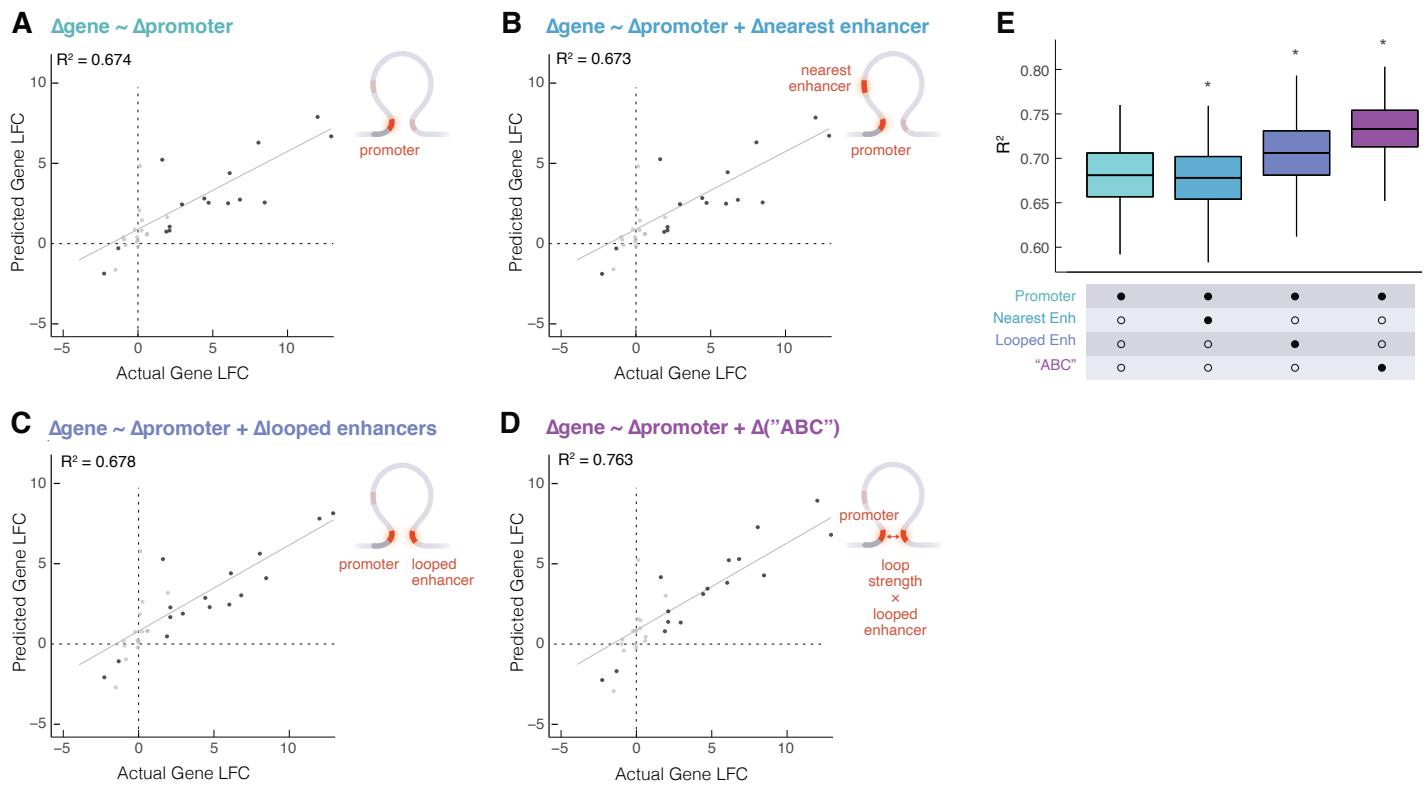
**Figure S3. Differential chromatin and transcription factor binding events.**

Heatmaps showing normalized counts for (A) H3K27ac, (B) Jun, (C) CTCF, and (D) Rad21. Clusters are indicated by the bars on the right side of each heatmap,  $p < 0.05$ ,  $\log_2(\text{fold-change}) > 2$ .



**Figure S4. Correlation of genomic features.**

Correlation heatmap showing the individual correlations of each of the features (top) and legend representing which feature is represented (bottom).



**Figure S5. Changes in gene expression at differential loops are explained by combined proximal and distal enhancer activity and loop strength.** Scatter plots showing predicted gene-fold change vs actual gene fold-change for genes that are at the anchors of differential loops based on (A) promoter H3K27 ac LFC alone, (B) promoter H3K27ac LFC and the nearest enhancer to the promoter's FC, (C) promoter H3K27ac FC and distal looped H3K27ac, and (D) promoter H3K27ac LFC and the LFC of the product of distal looped H3K27ac and loop strength (red = differential gene, gray = static gene). (E)  $R^2$  for each model calculated based on 1000 permutations of splitting data into training and testing sets. Wilcoxon rank sum test was performed to compare each group to the promoter only model, \*asterisk represents  $p < 0.05$ .