Changes in chromatin accessibility are not concordant with transcriptional changes for single-factor perturbations

Karun Kiani1, Eric M Sanford2, Yogesh Goyal3,4,5 & Arjun Raj3,6,*

Abstract

A major goal in the field of transcriptional regulation is the mapping of changes in the binding of transcription factors to the resultant changes in gene expression. Recently, methods for measuring chromatin accessibility have enabled us to measure changes in accessibility across the genome, which are thought to correspond to transcription factor-binding events. In concert with RNA-sequencing, these data in principle enable such mappings; however, few studies have looked at their concordance over short-duration treatments with specific perturbations. Here, we used tandem, bulk ATAC-seq, and RNA-seq measurements from MCF-7 breast carcinoma cells to systematically evaluate the concordance between changes in accessibility and changes in expression in response to retinoic acid and TGF-β. We found two classes of genes whose expression showed a significant change: those that showed some changes in the accessibility of nearby chromatin, and those that showed virtually no change despite strong changes in expression. The peaks associated with genes in the former group had lower baseline accessibility prior to exposure to signal. Focusing the analysis specifically on peaks with motifs for transcription factors associated with retinoic acid and TGF-β signaling did not reduce the lack of correspondence. Analysis of paired chromatin accessibility and gene expression data from distinct paths along the hematopoietic differentiation trajectory showed a much stronger correspondence, suggesting that the multifactorial biological processes associated with differentiation may lead to changes in chromatin accessibility that reflect rather than driving altered transcriptional status. Together, these results show many gene expression changes can happen independently of changes in the accessibility of local chromatin in the context of a single-factor perturbation.

Keywords chromatin accessibility; gene regulation; multi-omics integration; RNA-seq and ATAC-seq concordance; signal response

Subject Category Chromatin, Transcription & Genomics

Introduction

Transcription factors regulate gene expression by binding to specific DNA sequences, facilitating transcription through the recruitment and activation of the transcriptional machinery. Deciphering the combinatorial logic underlying which transcription factors bind to what portions of DNA and in what contexts are a central challenge in creating a complete model of transcriptional regulation. Sequencing-based methods have enabled the measurement of transcript levels for all genes and the putative binding profiles of transcription factors across the genome. However, the precise mapping between changes in these putative binding profiles and the changes in transcriptional activity remains the subject of debate.

A key component of decoding the relationship between transcription factor activity and the resultant changes in transcription is the measurement of transcription factor binding to DNA. Recently, the combination of biochemical binding assays with sequencing-based readouts has led to a cornucopia of methods for making such measurements. One workhorse method is chromatin immunoprecipitation sequencing (ChIP-seq), which characterizes the binding of transcription factors and other DNA-protein interactions genomewide (Barks et al., 2007; Robertson et al., 2007; Ma & Zhang, 2020) by using immunoprecipitation of proteins that bind to chromatin and subsequently sequencing the coprecipitated DNA. However, ChIP-seq is limited in that each experiment can only interrogate the binding profile of one transcription factor at a time.

An alternative approach that circumvents that issue is the measurement of changes in the accessibility of DNA to infer changes in the binding of all transcription factors at once. Accessible regions of DNA (i.e., those regions depleted of nucleosomes) represent only 3% of the genome but often participate in the regulation of gene...
expression (Weintraub & Groudine, 1976; Wu et al, 1979; Lee et al, 2004; Thurman et al, 2012). These regions can be detected genome-wide by combining the enzymatic activity of nucleases with high-throughput sequencing using techniques such as DNase I hypersensitive site sequencing (DNase-seq; Boyle et al, 2008) and assay for transposase accessible chromatin with sequencing (ATAC-seq; Buenrostro et al, 2013). The interpretation of these accessibility methods leans heavily on the assumption that changes in regulatory factor binding are reflected in changes in chromatin accessibility. Certainly, there are many examples in which the correspondence between changes in accessibility strongly corresponds to changes in transcriptional output. For instance, the summation of ChIP-seq signal for 42 transcription factors mapped by encoding in K562 chronic myelogenous leukemia cells paralleled the signal from accessible sites revealed by DNase-seq (Thurman et al, 2012). Moreover, computational methods to infer transcription factor footprints from accessibility measurements have been shown to recapitulate ChIP-seq binding well (Pique-Regi et al, 2011). Accessibility methods can also be used to look for changes in accessibility across various perturbations and cell types. Changes in accessibility generally seem to correspond to changes in transcription in the sense that large changes in transcriptional output are reflected in broad changes in the accessibility of several loci in the surrounding chromatin (González et al, 2015a; de la Torre-Ubieta et al, 2018).

However, it is unclear how well these accessibility-based methods capture the activity of all transcription factors. It is possible that some transcription factors’ binding and activity do not result in corresponding changes in accessibility and vice versa. Such a lack of correspondence could manifest itself as a lack of correlation between changes in accessibility and changes in transcription. Given the underlying assumption that a change in transcription must be mediated by the change in some transcription factor activity, then such a lack of correspondence would suggest that changes in the activity of transcription factors could change expression without changing accessibility near its binding site. Indeed, previous work has demonstrated that the glucocorticoid receptor binds almost exclusively to pre-existing accessible chromatin prior to small-molecule stimulation (John et al, 2011) and that activator protein 1 (AP-1) establishes these binding patterns for the glucocorticoid receptor by maintaining chromatin accessibility (Biddie et al, 2011). Similarly, the lineage-defining transcription factor Foxp3 binds to preformed accessible sites established by its structural homolog, Foxo1, to establish regulatory T cell identity (Samstein et al, 2012). Of note, this process of regulatory T cell specification via Foxp3 is considered a “late differentiation” process, as the precursor cell state, the mature naïve CD4⁺ T cell is considered mature. While reports from the literature generally show a strong correspondence (González et al, 2015a; Ampuja et al, 2017; de la Torre-Ubieta et al, 2018; Starks et al, 2019), it is worth noting that the comparisons in such studies are often across rather different cell types. In such cases, it is possible that the changes in accessibility are not driven by regulation per se but rather reflect the consequences of sequential exposure to multiple regulatory factors that characterize the differentiation process. Such accessibility changes could, in principle, signify the reinforcement of genes that are already transcriptionally active genes or could even just appear around actively transcribed genes without any functional role. Disentangling such possibilities could be revealed with the use of single-factor perturbations that more directly affect an individual pathway; however, few such data are available.

Here, we used tandem bulk RNA-seq and ATAC-seq data from MCF-7 breast carcinoma cells exposed to multiple doses of retinoic acid or TGF-β to determine the degree of concordance between changes in chromatin accessibility and changes in gene expression. We demonstrate that while some differentially expressed genes have a high concordance between gene expression and chromatin accessibility changes, many other genes are differentially expressed without changes in their local chromatin accessibility. We show that these results hold across multiple parameters and definitions of accessibility change and that it does not depend on the type of transcription factor per se. We evaluated another published dataset of hematopoietic differentiation, which has much deeper and multifactorial differences, that showed much stronger concordance. We finally compare differences in pre-existing accessibility between concordant and non-concordant genes prior to single-factor perturbation. Our results provide a systematic evaluation of the concordance between changes in gene expression and local chromatin accessibility.

Results

Genome-wide expression and chromatin accessibility changes reflect known biology of two perturbations

To measure the correspondence between changes in chromatin accessibility and changes in gene expression, we used MCF-7 breast carcinoma cells due to their previously described transcriptional responses to all-trans retinoic acid (Hua et al, 2009; referred to here on as retinoic acid) and transforming growth factor beta (TGF-β; Mahdi et al, 2015). We used paired, bulk accessibility (ATAC-seq) and expression data (RNA-seq) from these cells (Sanford et al, 2020a; Data ref: Sanford et al, 2020b) collected 72 h after continuous exposure to three different doses of each signal (Fig 1A). We chose this timescale because previous work with MCF-7 cells showed more transcriptional changes at 72 h compared with 24 h after exposure to retinoic acid (Hua et al, 2009), and chromatin accessibility changes may not be detectable until 24 h after perturbation (Ramirez et al, 2017a).

Differential gene expression and differential peak accessibility analysis showed a dose-dependent response to both signals compared with ethanol control (Fig 1A and Appendix Fig S1A, bar plots). The ethanol “vehicle” controls comprise three different densities of cells, and the transcriptomes of control conditions globally were similar regardless of cell density (Appendix Fig S1B). To confirm that global gene expression and chromatin accessibility patterns were similar between signals and dosages, we performed a principal component analysis. For both RNA-seq and ATAC-seq data, all samples exposed to the same signal or ethanol control clustered together, indicating that their gene expression and chromatin accessibility were more similar to each other than to other conditions, supporting the quality of these data.

To validate that changes in gene expression were consistent with the known biology of these signaling pathways, we performed an over-representation analysis on the upregulated genes in response to high dose retinoic acid or TGF-β against curated gene sets from the molecular signatures database (Liberzon et al, 2011, 2015). The
Figure 1.
top 10 gene sets based on false discovery rate (FDR)-adjusted P-values were processes canonically associated with retinoic acid (morphogenesis, organ development, anterior-posterior patterning) and TGF-β (extracellular matrix, endopeptidase activity), respectively (Fig 1B). Gene set enrichment analysis (Subramanian et al., 2005) showed that genes that were differentially expressed in response to high dose retinoic acid were significantly enriched for genes associated with skeletal system morphogenesis, and genes that were differentially expressed as a result of exposure to high dose TGF-β were significantly enriched for genes associated with epithelial-to-mesenchymal transition (Appendix Fig S1B). Thus, the differentially expressed genes generally reflected the known biology of the signals the cells were exposed to.

We next wondered if the changes in chromatin accessibility in response to signal were associated with the activity of specific transcription factors, in particular, those associated with the biology of these signaling pathways. We used a modified version of the chromVAR package along with its curated database of transcription factor motifs, cisBP, to identify the transcription factors with the largest predicted change in activity (Schep et al., 2017). We used the set of differential peaks to determine the set of the top 150 transcription factors associated with the greatest magnitude of change. These included the binding motifs of transcription factors that are canonical effectors of retinoic acid (RAR-α, HOXA13) and TGF-β signaling (SMAD3, SMAD4, and SMAD9). For each of these transcription factor motifs, we calculated a motif enrichment score for each condition based on the bias-uncorrected deviation score from chromVAR. The motif enrichment score represents the percentage change in ATAC-seq fragment counts in all peaks that contain a given transcription factor’s motif (Fig 1B). For example, the enrichment score of 28% for SMAD3 in the TGF-β condition meant that peaks containing the SMAD3 motif on average saw a 28% increase in fragment counts after exposure to TGF-β. We pooled together the low, medium, and high doses for each condition together in order to decrease the variability of motif enrichment scores estimates. Thus, our data recapitulated expected changes in accessibility, presumably due to the activity of transcription factors well-known to be activated by the signals used. Thus, of the changes in accessibility we did detect, they made sense based on a model of transcription factor activity leading to changes in accessibility. However, it was still possible that the activity of many transcription factors was not captured by changes in accessibility.

The relationship between changes in chromatin accessibility and gene expression varies on a gene-by-gene basis

We next wondered whether genes that were differentially expressed were more likely to have differentially accessible peaks nearby, i.e., was there concordance between gene expression and chromatin accessibility changes at the level of individual genes? To initially characterize the extent of concordance between these data, we looked at the overlap between genes that were differentially expressed in response to high dose signal and genes with differentially accessible peaks nearby after exposure to signal (Fig 1C). We assigned each accessible peak to the nearest transcriptional start site (“nearest approach”). Using this approach, the majority of genes had fewer than 20 peaks assigned to them (Appendix Fig S2A). We found that of the over 2,000 genes upregulated in response to high dose retinoic acid, more than half of them had at least one differential peak (irrespective of the direction of peak change) assigned to them in response to TGF-β, approximately 75 and 81% of genes had all peaks either not differentially accessible or differentially accessible in the same direction of gene expression changes, respectively (Appendix Fig S2B). Thus, using the “nearest” approach, genes that are differentially expressed are more likely than random chance to have a nearby peak that is differentially accessible in response to retinoic acid or TGF-β.

While using this overlap-based approach showed correspondence between genes that are differentially expressed and their nearby peaks in response to signal, aspects of the nature of the concordance of these changes were not captured by this analysis. For example, the overlap-based method counted all differentially accessible genes
that had at least one differentially accessible peak assigned to them as concordant but did not take into account the proportion or degree to which those nearby peaks change. Moreover, we did not take into account the relationship between the directionality of changes in gene expression and chromatin accessibility. The underlying assumption at the basis of this relationship is that when peaks become more accessible than the nearby gene increases its expression, and the overlap-based approach does not take this correspondence of the direction of change into account. To better characterize these facets of concordance, we first individually examined the changes in chromatin accessibility nearby two genes whose expression was upregulated in response to retinoic acid.

After optimizing parameters for calling peaks and determining differentially accessible peaks (Appendix Fig S3), we found that while a large number of peaks are differentially accessible near the HOXA1 locus, very few peaks are differentially accessible near the SLC5A5 locus (Fig 1D and E, track view middle, accessibility plot, right; Appendix Fig S4A and B). HOXA1 and SLC5A5 induction are associated with exposure to retinoic acid (Schmutzler et al., 1997; Kogai et al., 2000; Glover et al., 2006), and both genes showed a dose-dependent increase in expression in response to retinoic acid (Fig 1D and E leftmost panels; Appendix Figs S4A and B). Therefore, genes with high expression change in response to signal can show a large degree of accessibility changes or show very little accessibility changes, suggesting that changes in transcription factor activity may or may not be reflected in changes in accessibility.

**Chromatin accessibility changes are less concordant with large changes in gene expression in signaling compared with hematopoietic differentiation**

Next, we evaluated the concordance between accessibility and gene expression genome-wide while also factoring in the directionality of changes and the relative proportion of peaks that are changing on a gene-by-gene basis. As a point of comparison, we used previously published gene expression and chromatin accessibility data from hematopoietic differentiation (González et al., 2015a; Data ref: González et al., 2015b) that demonstrated that large changes in gene expression were typically associated with gains or losses (depending on the direction of expression change) of cell type-specific enhancers when comparing the expression and accessibility of hematopoietic stem and progenitor cells (HSPCs) to monocytes.

Before using this dataset as a comparison to ours for measuring concordance between chromatin accessibility and gene expression changes, we verified that the hematopoietic differentiation data was similar to our own by a variety of metrics. First, we wanted to compare whether the number of differentially expressed genes and differentially accessible peaks between HSPCs and monocyes in the hematopoietic differentiation data was similar to the numbers from MCF-7 cells exposed to retinoic acid or TGF-β. We found that both HSPC and monocyte populations had greater than 2,000 genes that were specifically expressed in their respective cell types compared with the approximately 2,000 and 1,500 genes differentially expressed in MCF-7 cells in response to high dose retinoic acid and TGF-β, respectively (Fig 1A). Moreover, HSPC and monocyte populations had more than 6,000 differentially accessible peaks (Appendix Fig S5A) compared with the approximately 15,000 and 6,000 differentially accessible peaks in MCF-7 cells in response to high dose retinoic acid and TGF-β, respectively (Fig 1A).

Next, we annotated the location of peaks based on where in the genome they were located relative to gene bodies and quantified what proportion of peaks fell into annotation categories such as promoter, intergenic, exonic, intronic, etc. ATAC-seq peaks from MCF-7 cells had a larger proportion of peaks at gene promoters (within 3 kilobases upstream or downstream of the transcription start site) whereas a greater proportion of the DNase I hypersensitive sites in the HSPC and monocyte populations were from distal intergenic regions compared with promoters (Appendix Fig S5B). This finding could be the result of inherent differences in the assays or could reflect biological differences. Moreover, the MCF-7 data had a greater proportion of peaks located at gene promoters, which could in principle bias our results toward having a larger degree of concordance because accessibility changes at promoters were more strongly correlated with gene expression changes than distal accessible. Despite this bias, our data demonstrate less concordance.

Given the different assays used to determine genome-wide chromatin accessibility, we realigned the DNase-seq data to the hg38 reference and examined the peaks at a “housekeeping gene” (GAPDH), hematopoietic differentiation-specific genes (CD34, CD14) and retinoic acid and TGF-β-related genes (DHRS3, SERPINA11) to spot-check that the accessibility data were similar. Indeed, there were similar accessibility profiles for GAPDH, and appropriate differences in accessibility given the cell type of signal for the other sites, indicating the accessibility data were comparable (Appendix Fig S6A–E). Moreover, to look at similarities in accessibility genome-wide, we calculated the intersection of the consensus peak sets from hematopoietic differentiation and MCF-7 signal response datasets, which included both peaks that were differentially accessible and those that were not. We observed that approximately 55% of peaks from hematopoietic differentiation data (DNase-seq) overlapped with peaks from the MCF-7 signal response dataset (ATAC-seq). These results show that the datasets do not have systematic qualitative differences in either expression or accessibility, enabling us to compare the degree of concordance across these two systems.

In the original analysis of hematopoietic differentiation, the authors found that regulatory complexity (defined as the number of accessible regions closest to a gene’s transcriptional unit) was an important discriminating factor for whether changes in accessibility corresponded to changes in expression, with areas of high complexity showing more correspondence than those of low complexity. Hence, we similarly grouped genes from our MCF-7 dataset into high and low complexity for our comparisons. We categorized genes with more than 7 peaks assigned to them using the “nearest approach” as “high complexity,” while genes with 7 or fewer peaks were categorized as having “low complexity” (Fig 2A, top panel). The cutoff for loci complexity was calculated by taking a tertile-based approach (González et al., 2015a) and calling any number of peaks above the highest tertile cutoff as high and any peak below that as low complexity (Fig 2B, solid line, lower plot). Because high complexity genes on average had higher levels of expression in the hematopoietic differentiation data, we sought to determine whether there was any difference in expression between high and low complexity genes in our MCF-7 data. The median expression of high complexity loci was similarly higher than low complexity loci in
A

3/4 peaks are differentially more accessible

Control

+Signal

Gene A

2/4 peaks are differentially less accessible

Control

+Signal

Gene B

1/4 peaks are differentially more accessible

Control

+Signal

Gene C

González et al., 2015 (differentiation)

B

Hematopoietic stem and progenitor cell (HSPC) (CD34+ population)

Monocyte (CD14+ population)

Top 50 upregulated HSPC transcripts

Next 50 upregulated HSPC transcripts

peaks more differentially accessible in HSPCs

CCRF1

Top 50 upregulated monocyte transcripts

Next 50 upregulated monocyte transcripts

peaks more differentially accessible in Monocytes

Sanford et al., 2020 (signal: retinoic acid)

C

Top 50 upregulated ethanol transcripts

Next 50 upregulated ethanol transcripts

peaks more differentially accessible before addition of high dose retinoic acid

STRAG6

peaks more differentially accessible after addition of high dose retinoic acid

WNT11

Sanford et al., 2020 (signal: TGF-β)

D

Top 50 upregulated ethanol transcripts

Next 50 upregulated ethanol transcripts

peaks more differentially accessible before addition of high dose TGF-β

PMENP1

peaks more differentially accessible after addition of high dose TGF-β

COL4A3

Proportion of high complexity DHS differentially accessible

Figure 2.
response to both exposure to high dose retinoic acid (23.30 versus 13.27 TPM) and high dose TGF-β (24.06 vs. 13.05 TPM) (Appendix Fig S7C, P-value < 2.2 x 10^-16 for both, Kolmogorov–Smirnov test) demonstrating that high complexity genes are more highly expressed as in the hematopoietic differentiation data. Despite this difference in expression, the distributions of peak widths for peaks of high and low complexity genes were similar (Appendix Fig S7D).

We began our analysis by focusing on the high complexity genes. To determine the concordance between gene expression changes and chromatin accessibility changes, we used the “nearest approach” to assign peaks to genes. For each gene, we compared the log2 of the fold change in expression between conditions versus the proportion of peaks that were differentially accessible in the same direction (i.e., peaks that increase in accessibility for genes that increase in expression after exposure to signal and vice versa). We observed that for hematopoietic differentiation, the 100 most highly expressed high complexity genes in the HSPC and monocyte populations had a high proportion of peaks, which were differentially accessible in the concordant direction, reproducing the conclusions of González et al (2015a) that large changes in expression were consistently associated with concordant changes in chromatin accessibility (Fig 2C). Next, we used this approach on our data to compare expression and accessibility changes between ethanol vehicle control and high dose retinoic acid or TGF-β. For both signals, we observed two distinct groups of genes within the top 100 most differentially expressed genes. One group of genes (“accessibility-concordant genes”) behaved similarly to those in the hematopoietic differentiation data, demonstrating a concordance between expression and accessibility changes (Fig 2C and D). However, the other group of genes (“accessibility-nonconcordant genes”) had large expression changes with little to no peaks nearby changing in accessibility, creating a skew in the distribution toward a lower proportion of peaks being differentially accessible in a concordant manner compared with the hematopoietic differentiation data (Fig 2B–D, density plots).

Adjusting the minimum peak coverage parameter changes the number of differential peaks and the proportion of differential peaks that change in the corresponding direction of expression. We wondered if a lower minimum coverage threshold changed the qualitative result we noticed before and thus conducted the same analysis using a lower minimum peak coverage threshold for determining differential peaks (see methods). We observed that a similar pattern occurred in high complexity genes with this set of parameters (Appendix Fig S8).

González et al (2015a) showed that for some low complexity genes, large changes in expression were not accompanied by concordant changes in accessibility (González et al, 2015a). We similarly wanted to confirm whether this decreased correspondence was the case in our data in response to retinoic acid and TGF-β. Using the same approach as before, we compared the log2 of the fold change in expression of low complexity genes to the proportion of peaks with differential accessibility in the concordant direction. The hematopoietic differentiation and signaling data for low complexity all qualitatively had genes whose expression increased without concordant changes in accessibility (Appendix Fig S9A–C). The distribution of the proportion peaks that was differentially accessible in the concordant direction for the top 100 up and downregulated genes was roughly uniform when comparing HSPCs to monocytes (Appendix Fig S9A, density plot on right). By comparison, the distribution was skewed toward more genes having a lower proportion of peaks being differentially accessible in the concordant direction in response to signals in MCF-7 cells, especially in the case of TGF-β (Appendix Fig S9B and C, density plots on right). However, it is also possible that these differences may owe to systematic differences between the datasets, given that the hematopoietic differentiation data have twice as many differentially accessible peaks as the MCF-7 data. Thus, while both the signaling in MCF-7 and hematopoietic data demonstrated large gene expression changes without concordant changes in chromatin accessibility with low complexity genes, a greater proportion of genes did so in the signaling data.

Figure 2. Signaling shows less concordance between highly differentially expressed genes and chromatin accessibility changes compared with hematopoietic differentiation data for high complexity genes.
A Schematic demonstrating how data from chromatin accessibility and gene expression data are used to create a proportion-based concordance plot.
B Concordance between expression and accessibility changes between hematopoietic stem and progenitor cells and monocytes. Left: plot showing changes in gene expression in CD34+ hematopoietic stem and progenitor cells (blue) and CD14+ monocytes (orange) from González et al, 2015a (schematic, top). For the plots, each dot is a gene, and on the x-axis is a log2 fold change in expression and on the y-axis the proportion of differentially accessible ATAC-seq peaks for each gene. The top 100 most highly expressed genes in ethanol vehicle control and high dose retinoic acid are colored in shades of gray and red, respectively. Middle: density plot of the distribution of the proportion of high complexity ATAC-seq peaks associated with the top 100 expressed genes in ethanol vehicle control and high dose retinoic acid with median value marked by a vertical black line. Right: example ATAC-seq tracks of STRA6 and WNT11.
C Concordance between expression and accessibility changes between cells exposed to ethanol vehicle control and high dose retinoic acid. Left: plot showing changes in gene expression and chromatin accessibility between ethanol vehicle control and high dose retinoic acid. Each dot is a gene, and on the x-axis is the log2 fold change in expression and on the y-axis the proportion of differentially accessible ATAC-seq peaks for each gene. The top 100 most highly expressed genes in ethanol vehicle control and high dose retinoic acid are colored in shades of gray and red, respectively. Middle: density plot of the distribution of the proportion of high complexity ATAC-seq peaks associated with the top 100 expressed genes in ethanol vehicle control and high dose retinoic acid with median value marked by a vertical black line. Right: example ATAC-seq tracks of PMEP1 and COL4A3.
D Concordance between expression and accessibility changes between cells exposed to ethanol vehicle control and high dose TGF-β. Left: plot showing changes in gene expression and chromatin accessibility between ethanol vehicle control and high dose TGF-β. Each dot is a gene, and on the x-axis is the log2 fold change in expression and on the y-axis the proportion of differentially accessible ATAC-seq peaks for each gene. The top 100 most highly expressed genes in ethanol vehicle control and high dose TGF-β are colored in shades of gray and blue, respectively. Middle: density plot of the distribution of the proportion of high complexity ATAC-seq peaks associated with the top 100 expressed genes in ethanol vehicle control and high dose retinoic acid with median value marked by a vertical black line. Right: example ATAC-seq tracks of PMEP1 and COL4A3.
Peaks nearby genes with high concordance have lower accessibility prior to exposure to signal

We wondered what the differences were between genes that were differentially expressed and had large accessibility changes versus those that were differentially expressed and had low accessibility changes. First, for high dose retinoic acid and TGF-β, we split genes into four groups based on whether they were differentially expressed and the proportion of peaks assigned to them using the “nearest” method that was differentially accessible in the appropriate direction. These four groups were (i) genes with differentially upregulated expression and concordant accessibility changes (ii) genes with differentially upregulated expression nonconcordant accessibility changes (iii) genes with differentially downregulated expression and concordant accessibility changes, and (iv) genes with differentially downregulated expression and nonconcordant accessibility changes (Fig 3A and B). We quantified the distribution of peak complexity across these groups and observed that genes with any concordant peaks for both conditions in both directions tended to have a higher degree of locus complexity (Appendix Fig S10A and B).

We first asked whether the change in accessibility between these two gene groups was due to differences in the pre-existing accessibility of peaks for these genes. Indeed, we found the baseline accessibility of peaks for genes with concordant increases in expression and accessibility in ethanol vehicle conditions was lower than those of peaks of genes that increase in expression without a commensurate change in chromatin accessibility (Fig 3C). This relationship was also recapitulated for concordant peaks that increase in expression and accessibility in response to high dose TGF-β (Fig 3D). Similarly, when comparing genes that are differentially downregulated in expression a similar pattern holds true in the opposite direction (Fig 3C and D; Appendix Fig S10C and D). One explanation may be that genes whose nearby chromatin was already accessible were permissive toward the action of the appropriate transcription factors to modulate expression. An alternative explanation is that the ATAC-seq assay itself had saturated in its ability to measure chromatin accessibility. By contrast, the difference in accessibility decreased between genes with a low proportion of peaks that were differentially accessible and genes with a high proportion of accessible peaks after exposure to signal (Appendix Fig S10C and D). Thus, the difference in the proportion of accessible peaks nearby the two groups of genes was partially explained by the pre-existing chromatin accessibility.

Multiple approaches to integrating chromatin accessibility and gene expression changes show a low degree of concordance during signaling

Finally, we measured to what degree the change in the accessibility of chromatin nearby a gene is reflected in the change in gene expression. Because linear distance is not always a good predictor of what accessible regions interact with what genes, we used multiple approaches to assign peaks to genes. First, we used the “nearest approach” to create a one-to-one mapping between accessible sites and genes by assigning them to the nearest transcriptional start site (Li et al., 2012; Nair et al., 2021), again comparing our signaling dataset to the hematopoietic differentiation dataset. Because many genes have multiple peaks assigned to them, we used two methods for collapsing peak values per gene: either the median accessibility of peaks across genes or the maximum (Fig 4A, schematic). We observed a stronger correlation between accessibility and expression changes in differentiation data (median approach Pearson’s \( r = 0.34 \), maximum approach Pearson’s \( r = 0.26 \)) than in MCF-7 in response to signal (retinoic acid: median approach Pearson’s \( r = 0.27 \), maximum approach Pearson’s \( r = 0.10 \); TGF-β: median approach Pearson’s \( r = 0.27 \), maximum approach Pearson’s \( r = 0.10 \); Fig 4A, right side).

Next, we used a window-based approach where there was the possibility of many-to-one mapping of peaks to genes. We assigned all peaks within a 100 kilobase window (Sanford et al., 2020a) in order to maximize the number of differential peaks assigned to a gene (Appendix Fig S11A and B). Similar to the “nearest” approach, we collapsed values using median accessibility change across all peaks assigned to a gene and maximum accessibility per gene (Fig 4B, schematic). We observed a similar effect using this approach where there was a stronger correlation between change in accessibility and change in expression between HSPC versus monocyte versus MCF-7 cells exposed to signal (Fig 4B). Of note, the correlation coefficients were similar between both methods of assigning peaks. Additionally, we used the window-based method to subset promoter-proximal peaks (i.e., within 1.5 kilobase pairs up or downstream from the transcription site) and distal peaks (greater than 20 kilobase pairs from the transcriptional site). This approach similarly did not demonstrate a strong relationship in the concordance between chromatin accessibility changes and gene expression changes in response to retinoic acid or TGF-β (Appendix Fig S12A and B). We also used a variable window approach by restricting our analyses to all peaks within the same topologically associating domain (TAD), which also did not demonstrate a strong correlation between changes in accessibility and changes in gene expression (Appendix Fig S12C).

We also wondered if the correlation between the extent of chromatin accessibility changes and gene expression changes would be different at the two lower doses. We used both the median and maximum peak value per gene while assigning peaks to genes using the nearest and window approaches. We observed a similarly weak correlation as high dose signal using all methods at both low and medium doses (Appendix Fig S11C and D). Consequently, the correlation between the magnitude of change in gene expression and chromatin accessibility was modest across the range of doses of signals.

To see whether peaks in specific genomic regions (promoters, parts of the gene body, downstream and intergenic areas) had unique relationships between change in chromatin accessibility and change in gene expression, we subsetted our correlation analysis. We annotated peaks using ChiPseeker (Yu et al., 2015) to categorize them as being at promoters, within the gene body (5‘ UTR, 3‘ UTR, intronic, and exonic sequences), downstream of the gene end, or at intergenic sequences. We used peaks assigned to genes using the “nearest” approach and took the median change in accessibility per gene. The strongest correlation between changes in accessibility and gene expression across sets of comparisons was at promoter peaks (Fig 4C). While promoter correlation is quantitatively stronger, the overall qualitative conclusion remains the same. We also quantified changes in intronic reads and compared them with changes in
exonic reads in order to determine whether there is a stronger relationship between more nascent RNA changes and accessibility changes. However, the quantitative relationship was no better than those using other methods (Appendix Fig S13). Thus, despite using a variety of approaches for both assigning peaks to genes and collapsing the accessibility of all peaks for a given gene to a single
Figure 4.
RAR-α peaks could be facilitating transcription factor binding, we measured changes and expression changes, even for pioneer transcription factors that show no more concordance between peak accessibility (top column) or maximum peak change (bottom column). Right: scatterplots showing the change in peak accessibility (median or maximum) versus log₂ fold change in expression on y-axis for hematopoietic differentiation data from González et al (2015a) (left column) and for high dose retinoic acid and high dose TGF-β (right two columns). Pearson’s correlation coefficients were reported with 95% confidence interval from bootstrapping with 10,000 replicates in parentheses.

C Using the “nearest” approach to look for correlation between accessibility and gene expression changes based on annotations of peak location. First two columns showing correlation for hematopoietic differentiation data from González et al (2015a), and right four columns showing correlation for high dose retinoic acid and high dose TGF-β, respectively. Pearson’s correlation coefficients were reported with 95% confidence interval from bootstrapping with 10,000 replicates in parentheses.

value, we failed to appreciate a strong relationship between changes in accessibility and changes in gene expression.

Finally, we wondered if peaks that contained the motifs of transcription factors that are associated with retinoic acid and TGF-β signaling only (as opposed to all peaks) would show a stronger correlation between the changes in chromatin accessibility and gene expression. We annotated peaks with a log-likelihood score of a given motif being found in that peak and subsetted those peaks with a nonzero log-likelihood score to examine the correlation between changes in accessibility and gene expression. Using this approach, we examined log-likelihood scores for motifs associated with retinoic acid signaling (RAR-α, HOXA13, and FOXA1) and motifs associated with TGF-β (SMAD3, SMAD4, and SMAD9). We observed that focusing on peaks annotated with peaks we would a priori expect to be involved in modulating gene expression in response to signal showed a limited correlation between changes in chromatin accessibility and changes in gene expression (Fig 5A and B). We also used this approach to look for peak-expression correlations for transcription factors downstream of IL-1α (NFKB1, REL, and RELA, of which NFKB1 has pioneer activity) and similarly found little correlation between chromatin accessibility and gene expression changes (Appendix Fig S14B). These results suggest that particular transcription factors show no more concordance between peak changes and expression changes, even for pioneer transcription factors. To exclude the possibility that “subregions” within unchanging peaks could be facilitating transcription factor binding, we measured RAR-α and SMAD footprints within these peaks. The number of reads between control and exposure conditions did not change for these footprints, indicating that there are no more accessible subregions that could mediate transcription factor binding within peaks that were not differentially accessible (Appendix Fig S15A and B).

Discussion

Here, we integrated tandem, genome-wide chromatin accessibility, and transcriptomic data to characterize the extent of concordance between them in response to inductive signals. We demonstrated that while certain genes have a high degree of concordance of change between expression and accessibility changes, there is a large group of differentially expressed genes whose local chromatin remains unchanged. By comparison, data from cell types along the hematopoietic differentiation trajectory had a much higher degree of concordance between genes with large gene expression changes and chromatin accessibility changes.

What might explain the lack of concordant changes in chromatin accessibility? One explanation could be that pre-existing chromatin accessibility dictates the de novo binding of transcription factors, but that the binding of transcription factors to those regions does not result in further changes to accessibility. Such effects have been reported in the context of glucocorticoid signaling, in which the glucocorticoid receptor almost exclusively binds to chromatin that is already accessible in response to dexamethasone (John et al, 2011). Indeed, we demonstrated that genes that lacked concordance between changes in chromatin accessibility and gene expression were more likely to have nearby chromatin that was already accessible (Fig 3C and D). It is possible that in MCF-7 cells, the transcriptional effects of RA and TGF-β do not lead to a significant change in the activity of pioneer transcription factors, which are able to bind directly to condensed or inaccessible chromatin to facilitate its opening (Zaret, 2020). Also, implicit in our approach is the assumption that an increase in accessibility is associated with an increase in expression, which is not necessarily the case if a genomic locus becomes accessible to a repressive factor or a bound repressive factor is displaced by a nucleosome.

We looked at MCF-7 cells exposed to retinoic acid and TGF-β because these two signals induce a robust transcriptional response through distinct mechanisms. RAR-α remains bound to DNA and interacts with transcriptional activators in response to retinoic acid binding, while SMAD family members require TGF-β to bind to surface receptors to translocate to the nucleus. Yet, despite these differences, we observed that many genes changed expression independent of changes in chromatin accessibility for both signals. It is, however, possible that signaling molecules that exert their effects through very different types of transcription factors may have a different profile of concordance between changes in accessibility and gene expression. It is possible that other types of factors in a different context (e.g., different cell lines) may yield a stronger correspondence (Appendix Fig S14). Indeed, some acute stimuli can lead to more concordance (Appendix Fig S16A). Potential reasons for this difference are the ability of transcription factors such as NFKB to rapidly decondense heterochromatin to quickly mediate inflammatory responses (Jurida et al 2015a; Weiterer et al, 2020a). Further, systematic studies of a number of signaling pathways and timescales will be required to make a complete determination of the degree of concordance in various contexts.
Our data characterized molecular changes resulting from a single input (retinoic acid or TGF-β) in a clonal cell line, whereas the majority of work reported a stronger concordance between simultaneous measurements of accessibility and transcription compared with entirely different cell types or cells undergoing a directed differentiation protocol. What we have observed in the case of a single perturbation applied to cells that are not thought to change type per se is increased or decreased transcription with less concomitant nearby change in accessibility. How can one reconcile these observations? One possibility is that if we were to leave the signal on for longer, or combine it over time with the effects of several other signals, that we eventually would observe many further changes in accessibility proximal to a gene, concordant with the aforementioned results from comparisons between cell types. Whatever the source, these further

Figure 5.
changes in accessibility do not seem to occur randomly, given that
they largely reflect the direction of change in transcription (increased
accessibility for upregulation, decreased for downregulation). It may
be that these subsequent changes in accessibility do not explicitly
change transcription but rather alter the underlying regulatory logic
of the gene, i.e., the removal of a signal may not lead to a decrease in
the gene’s transcription, or the gene’s transcription may be sensi-
tized or desensitized to some other set of transcription factors.

Materials and Methods

PCA of RNA and ATAC-sequencing samples

Principal component analysis and visualization of RNA-seq and
ATAC-seq samples were performed using raw counts and perform-
ing a variance stabilizing transform. Results were visualized using
functions from the R DESeq2 package (Love et al., 2014).

RNA-sequencing analysis

Initial RNA-sequencing analysis was performed as previously
(preprint: Goyal et al., 2021). Briefly, reads were aligned to the
hg38 assembly using STAR v.2.7.1a and counted uniquely
mapped reads with HTSeq v0.6.1 and hg38 GTF file from
Ensemb1 (release 90). We used DESeq2 v1.22.2 in R 3.5.1 using
a minimum absolute-value log-fold change of 0.5 and a q value
of 0.05. For genes with multiple annotated transcriptional start
sites, we used the “canonical” transcription start site from the
knownCanonical table from GENCODE v29 in the UCSC Table
Browser.

We performed functional over-representation and gene set
enrichment analysis (Subramanian et al., 2005) of upregulated tran-
scripts in the high dose retinoic acid and high dose TGFB using
clusterProfiler v4.0.5 and enrichplot v1.12.3 (Wu et al., 2021).
P-values for the over-representation analysis were adjusted using a
false discovery rate approach. We used the C5 ontology and H hall-
mark curated gene sets from the Molecular Signatures Database
(MSigDB) v7.4 (Liberezon et al., 2011, 2015) as reference gene sets
to compare our upregulated genes.

ATAC-sequencing analysis

ATAC-seq alignment and peak calling were performed as previously
(Sanford et al., 2020a). We aligned peaks to the hg38 assembly using
bowtie2 v2.3.4.1 and filtered out low-quality alignments with sam-
tools v1.96, removed duplicate read pairs with picard 1.96, and used
custom Python scripts along with bedtools v2.25.0 to create
alignment files with inferred Tn5 insertion points. We called peaks
using MACS2 (Zhang et al., 2008) v2.1.1.20160309 with the
command, “macs2 callpeak --nomodel --nolambda --keep-
dup all --call-summits -B --SPMR --format BED -q 0.05
--shift 75 --extsize 150.”

Since we had three biological replicates per condition, we used
a majority rule approach to retain only summits that
were found in at least two replicates (Yang et al., 2014). Using
these condition-specific peak files, we used bedtools to create a
consensus peak file by merging each individual condition’s peak
summit file together in a manner that disallowed overlapping
peaks. We used bedtools merge command “bedtools merge -d 50” to combine features within 50 base pairs of
each other into a single peak after testing multiple merge
distances. We used the number of ATAC-seq fragment counts at
each peak in this merged consensus peak file for differential
peak analysis.

We used the custom peak analysis algorithm from Sanford et
al (2020a) that took advantage of additional ethanol control condi-
tions to estimate the false discovery rate in ethanol controls to then
identify differential peaks. Briefly, reads were quantified for each
peak in the master consensus file and fragments at each peak were
normalized to correct for differences in total sequencing depth using
the equation:

\[
\text{estimated FDR} = \frac{\text{(no. of conditions)}(\text{est. number of false positive peaks per condition})}{\text{total number of differential peaks in experimental conditions}}
\]

After calculating the estimated FDR in each cell of the 50 × 50 grid,
we then pooled together differential peaks contained in any cell
with an FDR of less than 0.25%.
We performed motif analysis on our set of differential peaks using chromVAR v1.8.0 (Schep et al, 2017), its associated cisBP database of transcription factor motifs, and the motifmatchR package from bioconductor. To decrease the variance of the transcription factor motif deviation scores, we pooled together the different dosages of retinoic acid or TGF-β. The chromVAR code was modified to extract an internal metric that equals the fractional change in fragment counts at motif-containing peaks for a given motif. For the calculation of motif enrichment scores, the motifs we used were derived from four distinct groups of motif recognition sequences: RARA, HOX, SMAD, and AP-1.

Footprint was performed by subsetting on peaks that were not differentially accessible but met the minimum normalized fragment count threshold. These peaks were used as the basis for the HINT (Li et al, 2019) subroutine of the Regulatory Genomics Toolbox suite (http://www.regulatory-genomics.org/).

Hematopoietic differentiation data

We used pre-existing RNA- and DNase I-seq data (aligned to genome assembly hg19) of hematopoietic differentiation (Data ref: González et al, 2015b) to compare against our data. We used data from the website provided in the paper (http://cbio.mskcc.org/public/Leslie/Early_enhancer_establishment/) to download annotations of peaks (peaksTable.txt), counts of DNase-seq (DNaseCnts.txt), and RNA-seq counts (RNAseqCnts.txt). Counts presented in these data files were quantile normalized and averaged when biological replicates were available. We filtered peaks with “CD14” or “CD34” under the “accessPattern” annotation to choose from peaks that were relevant for comparing HSPCs to monocytes. We used a log2 fold change of greater than or equal to 2 as a cutoff for assigning differential peaks. We used the pre-existing annotations of genes for each peak for peak-gene mapping. For determining the log2 fold change in gene expression we discarded genes whose maximum expression value across the two conditions was fewer than 5 quantile-normalized units.

For visualization of this dataset with our own accessibility data, we realigned raw fastq files DNase-seq files to the hg38 assembly using bowtie v2.3.4.1 and filtered out low-quality alignments with samtools v1.1 to generate new bam alignment files. The alignment files were combined using samtools merge in a single .bam file per cell type. Bam files were converted to .bigWig format using deepTools 3.5.1 (Ramírez et al, 2016) “bamCoverage -- normalizeUsing CPM” to create a “consensus” .bigWig for visualization. Peaks for CD34+ and CD14+ samples were made by filtering peaks annotated for these populations in the “accessPattern” column and creating separate .bed files using a custom script. The peak location in these .bed files was then lifted over from hg19 to hg38 using UCSC hgLiftOver. For comparing the overlap of peaks between data sets, we created consensus peak sets across all sample types and used the bedtools intersect function to quantify the proportion of peaks that intersected between the hematopoietic differentiation and signaling data.

Peak annotation

Peaks were annotated using ChIPseeker (Yu et al, 2015) to determine the relative proportion of features in the data from González et al (2015a) (DNAseq-seq) and Sanford et al, 2020a (ATAC-seq). For ease of visualization, certain categories like the three promoter categories were collapsed into one. ChIPseeker was also used to identify the nearest transcriptional start site to a gene used for the nearest integration approach described below. For making scatterplots of change in accessibility versus change in expression annotated by peak feature, a custom script was used to combine annotations from ChIPseeker into four categories: downstream, gene body, intergenic, and promoter.

For each of the top 150 most variable transcription factor motifs we identified using differential accessibility analysis, we used the R bioconductor motifmatchR package to annotate both the number of motif matches and a log-likelihood match score for each peak.

IL-1α stimulation and myelocyte differentiation data

Pre-existing RNA- and ATAC-seq data from KB epithelial cells before and after 1 h of stimulation with IL-1α (Data ref: Jurida et al, 2015b; Data ref: Weiterer et al, 2020b) and differentiation data of cells along the myelocyte lineage (Data ref: Ramirez et al, 2017b) were used for further comparison. Both types of data were aligned and normalized as mentioned above and peaks were annotated accordingly.

For the IL-1α data, there was only one biological replicate and the data were not collected concurrently. For the myelocyte data, we compared three replicates of HL-60 promyelocytes and three replicates of monocyte-derived macrophages, which were the result of a directed differentiation over the course of 168 h that involved exposure to vitamin D3 and phorbol-12-myristate-13-acetate (PMA). For determining the log2 fold change in gene expression we discarded genes whose maximum expression value across the conditions was fewer than 5 TPM. We used a log2 fold change of greater than or equal to 0.58 as a cutoff for assigning differential peaks.

ATAC-seq footprinting

To examine whether possible “subpeaks” within nondifferentially accessible peaks could explain concordance, be used “samtools merge” to combine aligned and filtered .bam files across conditions to make a single .bam file per condition. We then used a custom script to make .bed files of the peaks that were not differentially accessible and used HINT-ATAC (Li et al, 2019) from the regulatory genomics toolbox using “rgt-hint footprinting,” “rgt-motifanalysis matching,” and “rgt-hint differential” commands to measure the difference in reading accessibility for motifs across ethanol control and either high dose retinoic acid or high dose TGF-β conditions.

RNA and ATAC data integration

We employed a variety of methods for assigning peaks to genes. In the “nearest” approach, we used annotation from ChIPseeker to assign each peak to the nearest transcriptional start site. With this method, each peak is uniquely mapped to a single gene. In the “window” approach, we used a window of 50 kilobases on either side of the transcriptional start site (100 kilobases in total) to assign peaks to a gene, which could result in a peak being assigned to
multiple genes. To look at promoter-proximal peaks, we used a window of 1.5 kilobases on either side of the transcriptional start site (3 kilobases in total) to assign peaks to a gene. To examine promoter distal peaks, we took all peaks within the 100 kilobase total window and subsequently all peaks within 20 kilobases up or downstream of the transcriptional start site. For creating windows based on topologically associating domains (TADs), we used available MCF-7 TAD data (Barutcu et al., 2015) and lifted coordinates from hg19 to hg38 using UCSC hgliftOver. We then used bedtools intersect to assign out peaks within TADs near transcriptional start sites.

For the initial overlap analysis for Fig 1, peaks were assigned to genes using the “nearest method.” Overlap was considered if any of the differentially expressed genes had a differentially accessible peak, regardless of whether the expression and accessibility changes were in the same direction. Further analyses integrating gene expression and chromatin accessibility data took the magnitude and direction of change into account.

Intron/exon analysis

In order to look at the possibility of more nascent transcripts being more concordant with chromatin accessibility changes, we performed a secondary alignment of RNA-seq reads to an intron-formatted .gff file and quantified counts using HT-seq with the option “--type intron.” Raw intron and exon counts were quantile normalized and used to create scatterplots to compare the change in counts across conditions versus change in accessibility.

Track visualization

We visualized accessibility data using the web-based version of integrative genomics viewer (IGV) (Robinson et al., 2011, 2020). We prepared accessibility data for visualization by taking consensus files and converting them to .bigWig file format with either fragments per million or counts per million normalization. Bed files for identifying peaks were created using custom scripts.

Statistics and software

Unless otherwise stated, we performed analyses using R v4.1.0 with data manipulation and visualization done with tidyverse v1.3.1 (Wickham et al., 2019) and ggpubr v0.4.0. We used a Kolmogorov–Smirnov test to compare means. Unless otherwise stated, 95% confidence intervals for Pearson’s r were calculated by bootstrapping using 10,000 replicates.

Data availability

The datasets and computer code produced in this study are available in the following databases:

- Raw and processed data: BioStudies (S-BSST886; https://www.ebi.ac.uk/biostudies/studies/S-BSST886).
- Intermediate data types, processing, and plotting scripts: GitHub (https://github.com/kdhkiani/concordancePaper).

- Links to pre-existing data from other studies can be found in Appendix Table S1.

Expanded View for this article is available online.

Acknowledgements

We are greatly indebted to Professor Christina Leslie and Alvaro González for many insightful discussions and for assistance in working with their datasets. We also thank the members of the Raj lab for valuable feedback, especially Ally Coté, Lee Richman, and Ryan Boe. KK acknowledges support from NIH T32 GM008216; ESM acknowledges support from F30 HG010986; YG holds a Career Award at the Scientific Interface from BWF; and AR acknowledges support from NIH Director’s Transformative Research Award R01 GM137425, NIH R01 CA238237, NIH R01 CA232256, NIH P30 CA06520, NIH SPORE P50 CA174523, and NIH U01 CA227550.

Author contributions

Karun Kiani: Conceptualization; data curation; software; formal analysis; visualization; methodology; writing – original draft; writing – review and editing.

Eric M Sanford: Conceptualization; data curation.

Yogesh Goyal: Conceptualization; methodology.

Arjun Raj: Conceptualization; resources; supervision; funding acquisition; investigation; methodology; writing – original draft; project administration; writing – review and editing.

Disclosure and competing interests statement

AR receives royalties related to Stellaris RNA FISH probes. All other authors declare that they have no conflict of interest.

References

Ampuja M, Rantapero T, Rodriguez-Martinez A, Palmroth M, Alarimo EL, Nykter M, Kallioniemi A (2017) Integrated RNA-seq and DNase-seq analyses identify phenotype-specific BMP4 signaling in breast cancer. BMC Genomics 18: 68

Baskri A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823–837

Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, Browne G, van Wijnen AJ, Lian JB, Stein JL et al (2015) Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. Genome Biol 16: 214

Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schultz RL, Miranda TB, Sung M-H, Trump S, Lightman SL et al (2011) Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Mol Cell 43: 145–155

Boyle AP, Davis S, Shulha HP, Melzter P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132: 311–322

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10: 1213–1218

Glover JC, Renaud J-S, Rijli FM (2006) Retinoic acid and hindbrain patterning. J Neurobiol 66: 705–725

González AJ, Setty M, Leslie CS (2015a) Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. Nat Genet 47: 1249–1259

© 2022 The Authors

Molecular Systems Biology 18 e10979 | 2022 15 of 17
Ramírez RN, El-Ali NC, Mager MA, Wyman D, Conesa A, Mortazavi A (2017a) Dynamic gene regulatory networks of human myeloid differentiation. Cell Syst 4: 416–429.e3
Ramírez RN, El-Ali NC, Mager MA, Wyman D, Conesa A & Mortazavi A (2017b) Gene Expression Omnibus GSE79046 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79046). [DATASET]
Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A et al (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4: 651–657
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29: 24–26
Robinson JT, Thorvaldsdóttir H, Turner D, Mesirov JP (2020) igvjs: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). bioRxiv https://doi.org/10.1101/2020.05.03.075499
Sanstein RM, Arvey A, Josefovicz SZ, Peng X, Reynolds A, Sandstrom R, Nepo S, Sabo P, Kim JM, Liao W et al (2012) Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. Cell 151: 153–166
Sanford EM, Emert BL, Coté A, Raj A (2020a) Gene regulation gravitates toward either addition or multiplication when combining the effects of two signals. Elife 9: e59388
Sanford EM, Emert BL, Coté A & Raj A (2020b) Gene Expression Omnibus GSE152749 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152749). [DATASET]
Scheper AN, Wuu B, Buenrostro JD, Greenleaf WJ (2017) chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 14: 975–978
Schmutzler C, Winzer R, Meissner-Weigl J, Köhrle J (1997) Retinoic acid increases sodium/iodide symporter mRNA levels in human thyroid cancer cell lines and suppresses expression of functional symporter in nontransformed FRTL-5 rat thyroid cells. Biochim Biophys Res Commun 240: 832–838
Starks RR, Biswas A, Jain A, Tuteja G (2019) Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. Epigenetics Chromatin 12: 16
Subramanian A, Tamayo P, Mootha VK, Nanda K, Lempicki RA, Eisen MB, Spellman PT, Brown PO, Botstein D et al (2005) Gene set enrichment analysis: knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545–15550
Thurman RE, Rynes E, Humbert K, Vierstra J, Tanay A, Spivak JL, Miska RE, Hartwell L, Rinn JL et al (2011) The transcriptional landscape of the mouse embryonic stem cell. Nature 471: 45–52
Van de Rijn M, Dudoit S, Luu P, Speed T (2000) A decision-theoretic approach to 2-color microarray gene expression analysis. Bioinformatics 16: 465–476
Vanderveld AE, Tzelepis V, Oehlerking R, Koster W, Zhao M, Nassif C, Hofman J, van der Burg S, van de Winkel J et al (2019) Distinct FOXA2 activity profiles predict aggressiveness of human breast cancer. Cell Rep 29: 778–794
Wang Y, Wang H, Weinberg R, Iyer V, Hinrichs AS, Nettles JA, Stamatoyannopoulos JA, Zhang M, Lander ES et al (2009) Should we call it chromatin immunoprecipitation or chromatin immunoprecipitation sequencing? Nat Biotechnol 27: 1162–1162
Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J et al (2019) Welcome to the tidyverse. *J Open Source Softw* 4: 1686

Wu C, Wong YC, Elgin SC (1979) The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* 16: 807–814

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L et al (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovations* 2: 100141

Yang Y, Fear J, Hu J, Haecker I, Zhou L, Renne R, Bloom D, McIntyre LM (2014) Leveraging biological replicates to improve analysis in ChIP-seq experiments. *Comput Struct Biotechnol J* 9: e201401002

Yu G, Wang L-G, He Q-Y (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31: 2382–2383

Zaret KS (2020) Pioneer transcription factors initiating gene network changes. *Annu Rev Genet* 54: 367–385

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137

License: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.