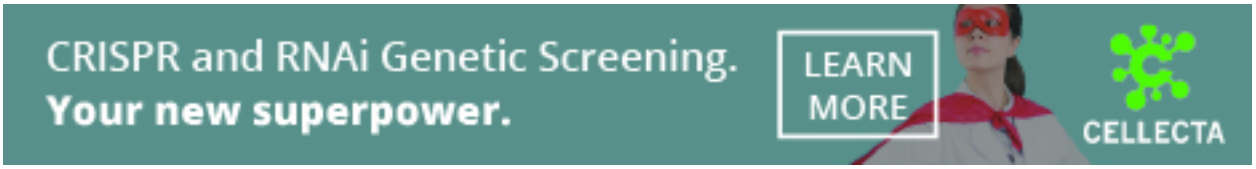
Kinetic networks identify TWIST2 as a key regulatory node in adipogenesis

Arun B Dutta, Daniel S Lank, Roza K Przanowska, et al.

Genome Res. published online February 21, 2023
Access the most recent version at doi:10.1101/gr.277559.122



Kinetic networks identify TWIST2 as a key regulatory node in adipogenesis

Arun B. Dutta[a], Daniel S. Lank[b], Roza K. Przanowska[c], Piotr Przanowski[c], Lixin Wang[c], Bao Nguyen[a], Ninad M. Walavalkar[a], Fabiana M. Duarte[d], and Michael J. Guertin[e,f]

[a]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia, United States of America

[b]Department of Pharmacology, University of Virginia, Charlottesville, Virginia, United States of America

[c]Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States of America

[d]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, United States of America

[e]Center for Cell Analysis and Modeling, University of Connecticut, Farmington, Connecticut, United States of America

[f]Department of Genetics and Genome Sciences, University of Connecticut, Farmington, Connecticut, United States of America

ABSTRACT

Adipocytes contribute to metabolic disorders such as obesity, diabetes, and atherosclerosis. Prior characterizations of the transcriptional network driving adipogenesis overlook transiently acting transcription factors (TFs), genes, and regulatory elements that are essential for proper differentiation. Moreover, traditional gene regulatory networks provide neither mechanistic details about individual RE-gene relationships nor temporal information needed to define a regulatory hierarchy that prioritizes key regulatory factors. To address these shortcomings, we integrate kinetic chromatin accessibility (ATAC-seq) and nascent transcription (PRO-seq) data to generate temporally resolved networks that describe TF binding events and resultant effects on target gene expression. Our data indicate which TF families cooperate with and antagonize each other to

regulate adipogenesis. Compartment modeling of RNA polymerase density quantifies how individual TFs mechanistically contribute to distinct steps in transcription. Glucocorticoid receptor activates transcription by inducing RNA polymerase pause release while SP and AP-1 factors affect RNA polymerase initiation. We identify *Twist2* as a previously unappreciated effector of adipocyte differentiation. We find that TWIST2 acts as a negative regulator of 3T3-L1 and primary preadipocyte differentiation. We confirm that *Twist2* knockout mice have compromised lipid storage within subcutaneous and brown adipose tissue. Previous phenotyping of *Twist2* knockout mice and Setleis syndrome *Twist2*$^{-/-}$ patients noted deficiencies in subcutaneous adipose tissue. This network inference framework is a powerful and general approach for interpreting complex biological phenomena and can be applied to a wide range of cellular processes.

INTRODUCTION

Mature adipocytes contribute to a multitude of metabolic processes by regulating energy balance, producing hormones, and providing structural and mechanical support (Rosen and Spiegelman, 2006). Adipocyte hyperplasia downstream of increased adipogenesis is associated with pathogenesis of obesity, type 2 diabetes, and cardiovascular disease (Unamuno et al. 2018; van Kruijsdijk et al. 2009). Adipogenic factors represent opportunities for intervention and possible mitigation of obesity-related sequelae (Ahmad et al. 2020; Ghaben and Scherer, 2019). Adipocyte maturation is a tightly regulated process involving many chromatin and transcriptional changes downstream of TF binding (Madsen et al. 2020; Rauch et al. 2019; Siersbæk et al. 2011; Thompson et al. 2016; Tsankov et al. 2015). While prior studies have extensively characterized the TFs and gene expression changes required for adipogenesis (Lefterova and Lazar, 2009; Rosen and Spiegelman, 2006; Siersbæk et al. 2012), this work relied on measurements taken hours or days apart on cells undergoing adipogenesis. Molecular events, such as TF binding, chromatin remodeling, and redistribution of RNA polymerase, occur on a time scale of seconds to minutes (Chen et al. 2014; Duarte et al. 2016; McNally et al. 2000). Therefore, previous examinations of adipogenic signaling likely omitted multiple waves of signaling and potential regulatory factors that may be critical to the process.

Molecular genomics assays can query transcriptional events with extremely high temporal resolution. While each assay delivers a tremendous amount of information, each is limited in the biology that it measures. ChIP-seq directly quantifies chromatin occupancy of proteins, but the assay is dependent upon availability of antibodies and limited to a single factor at a time. ATAC-seq and DNase-seq assays quantify chromatin accessibility, which is an indirect measure of regulatory element activity (Boyle et al. 2008; Buenrostro et al. 2015). Combining accessibility data with TF motif analyses can accurately infer TF binding without the need for factor and species-specific antibodies (Vierstra et al. 2020; Wu et al. 1979). Kinetic experiments can further increase the sensitivity of inferring dynamic TF binding, since changes in TF binding modulate local chromatin structure and accessibility (Guertin and Lis, 2013, 2010; Siersbæk et al. 2014, 2011). However, these assays do not directly inform on changes in transcription and RNA polymerase dynamics. While RNA-seq is a popular approach for measuring transcription, the assay relies on accumulation of mature RNA species over hours, making it inappropriate for rapid measurements. In addition, it is difficult to deconvolve mechanistic insights from RNA-seq data, which measures secondary and compensatory transcription as well as long-lived RNA species predating initial measurements. Alternatively, nascent transcription profiling with PRO-seq captures RNA polymerase density genome-wide at high spatial and temporal resolution (Kwak et al. 2013). PRO-seq, like RNA-seq, is limited in its ability to identify potential upstream regulatory elements (REs) and regulatory transcription factors. Only by combining multiple approaches can one fully capture the signaling dynamics driving transcription regulatory cascades.

Differential TF activity defines cell identity and drives cellular responses to environmental stimuli by enforcing gene regulatory programs (Takahashi and Yamanaka, 2006). Sequence-specific TFs bind to conserved motifs (Ptashne, 1967) in REs within promoters and enhancers to regulate different mechanistic steps in transcription (Fuda et al. 2009). TFs recruit cofactors such as chromatin modifying enzymes and general transcription machinery to REs. TFs are generally characterized as activators or repressors based upon their interaction partners, and recent studies more specifically describe TFs based upon their molecular function and which mechanistic steps they regulate (Danko et al. 2013; Duarte et al. 2016; Hah et al. 2011; Neumayr et al. 2022; Sathyan et al. 2019; Scholes et al. 2017). In addition to chromatin opening and RNA

polymerase recruitment, many transcription steps are precisely regulated, such as RNA polymerase pausing, elongation, and termination. RNA Polymerase II (PolII) pauses ~30-50 base pairs downstream of the transcription start site (TSS) (Rasmussen and Lis, 1993; Rougvie and Lis, 1988) and the vast majority of genes exhibit promoter-proximal PolII pausing (Core et al. 2008; Muse et al. 2007; Zeitlinger et al. 2007). Further modifications to the PolII complex triggers pause release and productive elongation (Marshall and Price, 1995). Defining the steps regulated by TFs is necessary to understand how TFs coordinate with one another productively or antagonistically to regulate complex gene expression programs.

Transcriptional networks consist of multiple rapid waves of signaling through time with potential regulatory feedback and signal propagation through activation and repression of regulatory factors. These complex regulatory cascades are not captured in traditional gene regulatory networks. Differentiating one wave from the next requires observations at multiple, closely spaced time points. In this study, we perform ATAC-seq and PRO-seq on 3T3-L1 cells at seven time points within the first four hours of adipogenesis. We incorporate accessibility and transcription changes into a multi-wave signaling network and identify TF families driving the regulatory cascade.

RESULTS

TFs from at least 14 families are associated with dynamic chromatin accessibility in 3T3-L1 differentiation

TFs bind promoters and enhancers to modify chromatin structure and influence transcription of nearby genes. To identify dynamic REs and potential TFs that regulate adipogenic differentiation, we induced adipogenesis in 3T3-L1 mouse preadipocytes (see Methods), harvested samples at 8 time points, and performed genome-wide chromatin accessibility assays (ATAC-seq) (Fig. 1A). Chromatin accessibility is a molecular measurement used to infer TF binding and RE activity. We identified over 230,000 accessibility peaks and found that differentiation time is the major driver of variation among the samples (Supplemental Fig.

S1A). To address whether the sequencing libraries are saturated we called peaks on subsets of the total reads and found that the number of called peaks had not reached saturation (Supplemental Fig. S1B). The fraction of reads in peaks (FRiP) varied between 0.2 and 0.3 for most of our ATAC-seq libraries (Supplemental Fig. S1C). However, we note that this score is depressed because we use our total peak set which includes dynamically accessible peaks that may not be accessible at all time points. Approximately 13% of all peaks change significantly over the time course (Supplemental Fig. S1D). We clustered dynamic peaks based on kinetic profiles (Supplemental Fig. S1E), which resulted in five general response classes (Fig. 1B). To identify candidate sequence-specific TFs that drive RE dynamics, we performed *de novo* motif analysis on dynamic peaks (Bailey et al. 2015). This approach yielded 14 potential TF family motifs including CEBP, TWIST, SP, KLF, AP-1, and the steroid hormone receptor motif (Fig. 1C & S1F). TF families comprise multiple proteins containing paralogous DNA-binding domains that recognize very similar sequence motifs (Fig. 1C). For example, multiple factors including androgen receptor, mineralocorticoid receptor, progesterone receptor, and glucocorticoid receptor (GR) bind to the steroid hormone receptor motif. However, GR is the only factor gene that is expressed in 3T3-L1 cells (Supplemental Fig. S1G). Therefore, we refer to the steroid hormone receptor binding consensus sequence as the GR motif. We identified AP-1, CEBP, and GR, which are known positive effectors of adipogenesis (Distel et al. 1987; Flodby et al. 1996; Freytag et al. 1994; Moitra et al. 1998; Ramji and Foka, 2002; Rubin et al. 1978; Siersbæk et al. 2011; Steger et al. 2010; Tanaka et al. 1997; Wang et al. 1995; Yeh et al. 1995). Members of the KLF and SP families are known to be associated with both pro-adipogenic (Birsoy et al. 2008; Inuzuka et al. 1999; Li et al. 2005; Mori et al. 2005; Pei et al. 2011) and anti-adipogenic functions (Banerjee et al. 2003; Kawamura et al. 2006; Sue et al. 2008; Tang et al. 1999). The TWIST family of TFs have previously unappreciated roles in adipogenesis, but have been shown to be important for differentiation of other mesenchymal cell types, such as osteoblasts (Bialek et al. 2004; Yousfi et al. 2001). Members of all these factor families are expressed in 3T3-L1 cells (Supplemental Fig. S1G).

TF binding or dissociation from DNA leads to enrichment of cognate motifs in dynamic peaks. The biological functions of the TFs determine whether binding or dissociation results in increased or decreased accessibility. Binding of TFs that recruit activating cofactors, such as histone acetyltransferses or remodeling

enzymes that eject nucleosomes, can increase accessibility; dissociation of these factors decreases chromatin accessibility. Likewise, binding and dissociation of factors that recruit deacetylases, repressive methyltransferases, or DNA methyltransferases can affect accessibility. We found that the majority of peaks containing CEBP, KLF, GR, or AP-1 motifs increase accessibility, while peaks containing TWIST or SP motifs decrease accessibility (Fig. 1D). We performed the reciprocal analysis and plotted the density of motif instances relative to the summits of increased, decreased, and nondynamic peak classes to confirm the classification (Fig. 1E & S1H). AP-1, GR, and CEBP motifs are strongly enriched around summits of increased peaks, while TWIST and SP motifs are enriched around summits of decreased peaks. SP and KLF families have paralogous DNA-binding domains and recognize similar motif sequences; however, we confidently associate chromatin decondensation to KLF factors and chromatin condensation to SP factors (Supplemental Fig. S1I). The SP family is composed of canonical activators (McKnight and Kingsbury, 1982), therefore SP TFs are likely dissociating from the chromatin to reduce accessibility. Although we ascribe opening and closing functions to the KLF and SP families, it is impossible to determine the relative contribution of KLF and SP factors at any individual motif. We believe that the dual enrichment of KLF motifs at both increased and decreased peak summits is due to erroneous classification of SP-bound REs as KLF-bound REs. This complication is not limited to closely related motifs, as many dynamic peaks contain multiple factor binding motifs, making it difficult to isolate the contribution of individual factors. To address this complication, we plotted the changes in accessibility at dynamic peaks that contain only a single motif (Fig. 1F). This confirmed that the majority of isolated AP-1, GR, CEBP, and KLF motif-containing peaks increase in accessibility, while TWIST and SP motif-containing peaks decrease. The biological interpretation of these results is that the adipogenic cocktail activates members of the AP-1, GR, CEBP, and KLF TF families both directly and through transcriptional activation of family member genes, leading to RE binding and chromatin decondensation. SP and TWIST motifs are associated with decreased accessibility. TFs can act as repressors by binding to chromatin and recruiting chromatin modifiers such as deacetylases. Alternatively, dissociation of an activating TF can lead to gene repression. These results confirm the importance of several TF families and suggest that previously unappreciated TF families, such as TWIST, contribute to adipogenesis.

SP, NRF, E2F6, KLF, and AP-1 factor motifs are associated with bidirectional transcription at regulatory elements

Coordinate TF binding ultimately results in the recruitment of RNA polymerases and initiation of transcription. In mammals, core promoters and enhancers often lack sequence information that consistently orient initiating RNA polymerases (Core et al. 2014). Therefore we sought to identify bidirectional transcription signatures as a complement to chromatin accessibility assays to identify REs (Core et al. 2008; Danko et al. 2013; Seila et al. 2008). We captured the short lived divergent transcripts found at active REs with PRO-seq in parallel with the ATAC-seq adipogenesis time points (Fig. 1A). We used discriminative regulatory-element detection (dREG) to identify peaks of bidirectional transcription from our PRO-seq data (Wang et al. 2019). We identified over 180,000 dREG peaks (Fig. 2A & B) and 18% change significantly over the time course (Supplemental Fig. S2A). ATAC-seq and PRO-seq measure distinct but related biological phenomenon, therefore they identify different but overlapping sets of REs. Approximately 22% of dynamic dREG peaks overlap with dynamic ATAC-seq peaks, compared to 20% of dynamic ATAC-seq peaks in the inverse comparison. To further analyze the two classes of REs, we separated the dynamic dREG and ATAC-seq peaks into intragenic, intergenic, and promoter regions (Fig. 2C). Both methods effectively identify REs within promoters (Supplemental Fig. S2B). We find PRO-seq more sensitively detects intragenic REs relative to the other categories, while ATAC-seq efficiently detects intergenic REs. We closely evaluated the overlap between ATAC and dREG peaks by plotting PRO-seq signal at ATAC peaks and vice versa (Supplemental Fig. S2C). We observe the distinctive bidirectional transcription signature at ATAC peaks irrespective of whether or not the ATAC peaks intersect dREG peaks. The signature is less intense at ATAC-seq peaks that do not overlap dREG peaks. Likewise, ATAC-seq signal is enriched at dREG-peaks that do not overlap ATAC-seq peaks (Supplemental Fig. S2D). Moreover, dREG peaks within intergenic, intragenic, or promoter regions that do not overlap with ATAC-seq peaks have less bidirectional transcription (Supplemental Fig. S2E). Although we find that bidirectional transcription and accessibility do not perfectly correlate, we are likely underestimating the extent of accessibility and bidirectional transcription overlap.

We sought to identify TFs that drive bidirectional transcription by further characterizing PRO-seq-defined REs. We hypothesized that different sets of TF motifs are enriched within REs defined by ATAC-seq and PRO-seq. For instance, the cognate motifs of TFs that recruit initiation machinery may be preferentially enriched at dREG-defined REs. We performed *de novo* motif analysis on dynamic dREG peaks and found enrichment of AP-1, SP, KLF, NRF, and E2F6 motifs (Fig. 2D). Of these, only the E2F6 motif was not also enriched in ATAC-seq peaks. We plotted motif density around the summits of either dynamic ATAC or dREG peaks to further differentiate ATAC and dREG-defined REs (Fig. 2E & Supplemental Fig. S2F). Of the motifs found *de novo* in dREG peaks, only E2F6, NRF, and SP were more enriched in dynamic dREG versus dynamic ATAC-seq peaks. We hypothesize that these three factor families regulate bidirectional transcription in adipogenesis. The SP motif is found in over 25% of human and mouse promoters, making the SP motif the most enriched *cis*-regulatory element (RE) within promoters (Benner et al. 2013). To determine whether divergent transcription signatures found at SP motifs are dominated by SP factors within promoters, we plotted plus and minus strand nascent transcription at all SP motif instances (Fig. 2F top). Indeed, when SP motifs within promoters are removed from the composite input, divergent transcription peaks collapse (Fig. 2F bottom). We also observe this phenomenon with E2F6 and NRF motifs (Supplemental Fig. S2G), implying that these factors and SP preferentially regulate divergent transcription at promoters. Next, we wanted to determine whether SP, NRF, and E2F6 motifs within promoters associate with increasing or decreasing divergent transcription. We find that bidirectional transcription tends to decrease in REs with dREG-enriched motifs as opposed to those without dREG-enriched motifs (Fig. 2G). This further supports the previous conclusions that SP and NRF motifs associate with decreases in RE activity (Fig. 1E & Supplemental Fig. S1H). We find distal regulatory elements are more likely to exhibit accessibility changes while promoters are more likely to exhibit bidirectional transcription changes.

Defining predicted TF binding events as *trans*-edges in the network

We determined candidate functional TFs within the set of REs by searching for over-represented sequence motifs and determining the expression levels of TF family members. However, inferring TF binding

from accessibility, motif, and expression data at any individual site remains a challenge (Guertin et al. 2012; Li et al. 2019). In addition to chromatin accessibility, expression of the TF, and presence of the TF's cognate motif, we leverage the change in accessibility over the time course to infer TF binding and dissociation events in adipogenesis. We term these predicted changes in TF occupancy, which are directed linkages from TFs to REs, as *trans*-edges in our networks. For simplicity, we refer to *trans*-edges as factor binding or dissociation events.

We define the following rules for *trans*-edge inference: 1) The RE must first be defined as an ATAC-seq peak at any time point. 2) The binding motif of the upstream TF must be present in RE. 3) Chromatin accessibility must change significantly between two time points to infer binding or dissociation. 4) The direction of accessibility changes must match with the molecular function of the TF as defined in Fig. 1. 5) Members of the TF family must be expressed at the appropriate time point; for example, the TF must be expressed at the later time point for binding and the earlier time point for dissociation. 6) GR, AP-1, and CEBP are directly activated by the adipogenic cocktail, so we infer edges from expressed family members to REs from 0-20 minutes. We necessitate that the nascent RNA expression of the other TFs changes significantly to infer *trans* edges from their genic node to an RE node. Mechanistically, TFs have short residency times on DNA and they are continually binding and dissociating from their sites *in vivo* (Chen et al. 2014; McNally et al. 2000). When we refer to inferred binding and dissociation within the network, we are strictly referring to overall changes in occupancy at a genomic site within the population of cells.

The following examples highlight implementations of these rules. The *Nr3c1* gene, which encodes GR, decreases expression immediately upon treatment (Supplemental Fig. S3A). We suggest that the rapid transcriptional repression of *Nr3c1* is the reason GR-associated increases in accessibility are transient. Therefore, we restrict binding edges attributed to GR to the first 40 minutes of the time course. We attribute any significant decreases in accessibility at inferred GR binding REs observed at later time points to dissociation of GR. We label these edges with a *dissociation* attribute. In the case of SP, we find that *Sp1*, *Sp3*, and *Sp4* are all repressed early in the time course (Supplemental Fig. S3B). We hypothesize that the delayed

accessibility decrease associated with SP motifs is due to transcriptional repression and natural turnover of the SP pool, which results in overall dissociation of SP on chromatin (Fig. 1F). We restrict *trans*-edges for SP to the later part of the time course. Conversely, we observe *Twist2* gene activation early in the time course (Supplemental Fig. S3C). Therefore, we predict that TWIST-associated repression is a result of increased TWIST binding and recruitment of negative cofactors. *Twist2* expression levels have returned to baseline in mature adipocytes (Supplemental Fig. S3D), suggesting that TWIST's effects are transient and can only be captured with an early, high-resolution time course. By focusing only on the REs that change accessibility and integrating with transcription data, we infer TF binding and dissociation events that drive adipogenesis.

We integrated publicly available TF ChIP-seq datasets to assess the performance of *trans*-edge inference. Specifically, we incorporated ChIP-seq profiling of AP-1 (cJun and JunB), KLF (KLF4 and KLF5), CEBPβ, and GR (Siersbæk et al. 2014, 2011). All these experiments were performed in 3T3-L1 preadipocytes at 4 hours of differentiation. Furthermore, we integrated TWIST2 ChIP-seq data from myoblasts overexpressing 3x-Ty1 tagged TWIST2 protein (Li et al. 2019). Both the difference in cell type and the disruption of normal TWIST2 function due to overexpression renders this dataset a poor comparison for our system. Nevertheless, we found that 55-70% of our inferred binding events for these factors overlap called ChIP-seq peaks (Supplemental Fig. S3E). The one exception was for GR, which exhibited a much lower degree of overlap (35%). This is consistent with our network, which suggests that GR binds and dissociates rapidly from the chromatin at many sites and would not be expected to be bound at the 4 hour ChIP-seq time point. Furthermore, CEBP, GR, and KLF regulatory elements that exhibit sustained high accessibility (i.e. nonattenuated) displayed a higher degree of overlap with ChIP-seq peaks than those that did not, suggesting that our ATAC-seq dynamics captured fluctuations in factor binding. In addition, we plotted composite ChIP signal at our predicted binding sites and found a strong enrichment in signal at REs that overlap with ChIP-seq peaks (Supplemental Fig. S3F). We also observed a weaker enrichment of ChIP signal around inferred binding events that do not overlap with ChIP-seq peaks, suggesting weaker binding events at these locations were overlooked in ChIP-seq peak calling. However, it is also possible that the enrichment in ChIP-seq signal at these regions reflects underlying accessibility rather than actual factor binding. We note that ChIP assays are

limited by antibody specificity, control data sets, and peak thresholding. Regardless, these ChIP-seq data validate the predictive power of using dynamic ATAC signal and the presence of sequence motifs to infer factor binding.

Proximal changes in accessibility are tightly linked to transcription

Chromatin accessibility positively correlates with local gene transcription. We confirmed this assertion by quantifying transcription of genes within 10 kilobases (kb) of dynamic ATAC-seq peak sets that exclusively increase or decrease accessibility (Fig. 3A). The majority of genes (63%) with one proximal increasing ATAC-seq peak are activated, likewise 68% of genes proximal to a single decreasing ATAC-seq peak are repressed. Genes near two or more increased accessibility peaks are much more likely to be associated with transcription activation, and vice versa (Fig. 3A). To further validate this association and explore the relationship between RE and target gene distance, we focused on all genes near one dynamic peak and stratified gene/peak pairs based on distance between the TSS and peak summit (Supplemental Fig. S3G-I). The closer the peak and the gene, the more likely gene transcription and peak accessibility correlate in the same direction. This result indicates that proximal REs have a greater impact on gene expression than distal elements. Moreover, we plotted change in gene transcription against distance-scaled local accessibility changes and observed the expected positive correlation between transcription and accessibility at both early and late phases of the time course (Supplemental Fig. S3J & K). These findings indicate that both accessibility dynamics and distance are important factors when considering the relationship between REs and genes.

We incorporated the distance between REs and genes as well as covariation in their accessibility and transcription to infer functional links, termed *cis*-edges, in our networks. We define *cis*-edges as predicted regulatory relationships between REs and genes. For example, if GR binds a regulatory element within a gene's promoter and induces gene activation, we draw a *cis*-edge between the RE and the gene. Within the network, we assign GR as an attribute to the edge. To confidently infer and annotate *cis*-edges, we must assess whether a class of TFs is associated with increasing or decreasing transcription. Since the distance

between a RE and a gene influences the likelihood that accessibility and transcription will covary, we classified the function of a TF class within the context of adipogenesis by determining if peaks with a cognate TF motif are closer to activated or repressed gene classes. We expected that factor families associated with decreases in accessibility, like SP, would be closer to repressed genes on average. To test this hypothesis, we first categorized genes as significantly activated, repressed, or unchanged for each pairwise comparison within the time course. For example, *Klf5* (Fig. 3B left panel) is one of a subset of 4225 genes immediately activated from 0 to 20 minutes (Fig. 3B right panel). Plotting the cumulative distribution function (CDF) for genes against the distance between the closest peak summit and the gene TSS shows that GR peaks tend to be closer to the 4225 activated genes compared to the repressed or unchanged genes (Fig. 3C). To estimate the maximum range that a factor can act, we plotted the difference between the repressed gene class CDF against the unchanged gene class CDF against distance between gene and peak (Fig. 3C inset). We find that the difference in the CDFs plateaus at around 114 kb, meaning that inferred GR binding events accumulate at the same rate for activated and unchanged genes at distances greater than 114 kb. This distance constraint represents an empirical observation that suggests a maximum regulatory distance in this system. It is possible that we would detect a different constraint in another cellular context due to differences in genomic architecture and regulatory environment. Immediately repressed genes like *Stat1* are closer to SP peaks than control gene sets (Fig. 3D & E). This analysis indicates that SP acts very proximal to its target genes, with an actionable range of less than 2 kb (Fig. 3E inset). This finding is consistent with our previous conclusions that decreased SP peaks are primarily found in promoters (Fig. 2F). Our observed maximal distances represent a hypothesized maximal actionable distance for factor activity in 3T3-L1s. However, we anticipate that optimal distance for most factors is much closer to target genes than these maxima. Therefore, we apply a closer distance threshold when inferring regulatory relationships between individual REs and genes as described in the next section. The lower thresholds increase our confidence in our predicted *cis*-edges.

Linking REs to target genes

We incorporate these biological principles into logical rules to define *cis*-edge predictions within an adipogenesis network. We develop our rules to maximize confidence in our predicted regulatory interactions. First, a gene and regulatory element must be within 10 kb to infer a *cis*-edge. Second, RE accessibility and gene transcription must covary over the same time range. These two logical rules provisionally link REs and genes, then we employ additional rules that reflect the biology of individual TFs. For instance, the 10 kb distance metric is made more strict for factors like SP, for which the functional distance constraint as determined in Fig. 3E is less than 10 kb. Temporal rules also influence edge predictions. For instance, GR-bound REs are only significantly closer to genes activated in comparison to the 0 minute time point such as the 20 vs. 0 comparison (Fig. 3C), meaning that genes activated later in the time course cannot be directly activated by GR binding in the network. Therefore, as with *trans*-edges, we only infer *cis*-edges between GR-bound REs and genes that change early in the time course. Incorporating these observations into our *cis*-edge rules we infer direct functional relationships between regulatory elements, bound TFs, and changes in target gene expression.

Constrained networks identify genes regulated combinatorially or by individual TF families

Quantifying nascent transcription with PRO-seq maps the position and orientation of RNA polymerase with base-pair resolution. Nascent transcriptional profiling captures engaged RNA polymerase species throughout the genome, including intragenic features such as the proximal promoter and gene body. We can infer regulatory mechanisms of gene sets by quantifying relative changes in RNA polymerase density within the pause region and gene body. For instance, if the rate of RNA polymerase pause release increases between conditions, we expect that the signal in the pause region to decrease and the gene body signal to increase. Previous studies focus on biological systems where one TF dominates the response, such as ER, HSF, and NF-kB (Danko et al. 2013; Duarte et al. 2016; Hah et al. 2011). In these systems, the composite RNA polymerase signals at activated genes highlight differences in densities between pause and gene body compartments (Danko et al. 2013; Duarte et al. 2016; Hah et al. 2011; Sathyan et al. 2019). A complication in our system is that multiple TFs cooperate to drive transcription changes, making it difficult to identify the target

steps (i.e. initiation, pause release) that TFs regulate. In order to address this complication, we identified genes that are predominantly regulated by a single TF in our network.

We constructed a bipartite network inferring changes in TF binding (*trans* edges) that regulate downstream changes in transcription (*cis* edges). Genes and REs can be regulated or bound by either one or a combination of TFs. For example, we constructed a constrained network with RE and gene nodes downstream of individual TFs, including AP-1. In this network, 1224 genes are solely activated by AP-1 and 1847 genes activated by AP-1 and at least one other factor (Fig. 4A). Most REs downstream of AP-1, both individually and combinatorially bound, are not linked to any downstream genes (12608 v. 4829). This network highlights a paradigm in the transcription field that a minority of TF binding events lead to changes in gene expression (Spradling et al. 1975; Westwood et al. 1991). We constructed similar networks for GR (Fig. 4B), SP (Fig. 4C), CEBP (Supplemental Fig. S4A), KLF (Supplemental Fig. S4B), and TWIST (Supplemental Fig. S4C). These networks illustrate the interconnectivity of gene regulation, while simultaneously identifying genes that are predominantly regulated by individual factors.

To extract mechanistic information from genes regulated by only one TF, we plotted composite RNA polymerase density from our PRO-seq data around pause peak summits at different time points for the isolated genes (Fig. 4D, E, F & Supplemental Fig. S4D, E, F). The resulting traces show the characteristic pause peak centered around 0 followed by release of the RNA polymerase into the gene body. Examining RNA polymerase density traces of genes only activated by AP-1 at 60 and 40 minutes show an increase in density in the pause region, suggesting increased RNA polymerase recruitment to AP-1-activated genes (Fig. 4D). These time points were chosen because AP-1 peaks are closest to genes activated between 60 and 40 minutes, suggesting that AP-1 exerts the most transcriptional control during this time range. At first glance we see a similar result for GR when comparing traces from 0 and 20 minutes (Fig. 4E). However the situation becomes more complex when considering the ratio of pause density to gene body density, or pause index (PI). The PI for genes regulated solely by AP-1 increases on average from 40 minutes to 60 minutes (Fig. 4D inset). Conversely, the PI for 71% of genes regulated solely by GR decreases on average between 0 and 20 minutes.

This suggests that GR primarily activates transcription by inducing pause release (Fig. 4E inset). The affected step is unique to the factor, as isolated AP-1 genes don't exhibit the decrease in pause index from 0 to 20 minutes observed with isolated GR genes (Supplemental Fig. S4G inset). Isolated GR genes show increases in pause index later in the time course, likely due to GR dissociation from the genome after the early phase of the time course and an associated decrease in pause release rate (Supplemental Fig. S4H). As for repressed genes, we find a decrease in pause peak and gene body intensity in predicted SP and TWIST target genes (Fig. 4F & Supplemental Fig. S4F). We find that SP target genes demonstrate a more symmetrical distribution of pause indices (Fig. 4F inset). While it is likely that these factors affect PolII recruitment, we sought to develop a more rigorous approach to determine how changes in initiation and pause release rates can account for observed changes in PolII density.

Modeling changes in regulatory transcription steps

We developed a mathematical model to further characterize how TFs target specific steps in the transcription cycle. Our model breaks up the gene unit into two compartments: a pause region and gene body region. PRO-seq directly measures RNA polymerase density within these regions for each gene. We define a series of differential equations to model polymerase density as measured by PRO-seq within the two compartments (Fig. 5A). We establish rate constants representing different transcriptional steps, namely RNA polymerase recruitment / transcription initiation (k_{init}), premature termination (k_{pre}), pause release (k_{rel}), and elongation (k_{elong}) The values of the rate constants determine the predicted density within the two compartments. We vary the rate constants for k_{init}, k_{pre}, and k_{rel} over two orders of magnitude and vary k_{elong} rate from 600 to 6000 bases per minute to determine the effect on pause and gene body density and how the model compares to observed changes. We make the assumption that k_{elong} remains constant between time points. Since k_{rel} and k_{init} are opposing rates in the model, we cannot distinguish an increase in one rate from a decrease in another. To simplify the model, we keep k_{pre} constant between time points. We determine how changes in k_{init} combined with k_{rel} changes can account for the average density changes for the 174 isolated GR-regulated genes from Fig. 4E. A wide range of rate parameters can describe the initial pause and gene

body densities, but regardless of the initial rates, a narrow fold-change in these rates can account for the observed changes between time points (Fig. 5B). We find that a ~1.07-fold increase in recruitment / initiation and a ~1.50-fold increase in pause release explain the changes in compartment occupancy between 0 and 20 minutes (Fig. 5B left). We calculated the absolute rate of initiation and residency time of PolII in the pause region based on the models and plotted a simulated PolII profile (Fig. 5C). For this simulation, we chose the parameter set with an elongation rate closest to the established consensus rate of approximately 2500 bases per minute (Ardehali and Lis, 2009; Jonkers and Lis, 2015). Estimated pause residency time drops from 29 seconds to 19 seconds between 0 and 20 minutes as a result of the rate constant changes. Taking a similar approach, an ~0.78-fold decrease in recruitment / initiation rate with a ~0.94-fold change in pause release rate produces observed changes in PolII occupancy between 60 and 120 minutes for the 1127 isolated SP genes (Fig. 5B middle). This corresponds to an initiation / recruitment rate reduction from 15.1 to 11.9 polymerase molecules per minute (Fig. 5D). If SP factors normally stimulate initiation, then mass action would explain dissociation of SP factors upon transcriptional repression of SP genes. Previous studies link SP1 to transcriptional initiation through interaction with the TFIID general TF (Gill et al. 1994). The observed changes in RNA polymerase composite profiles between 40 and 60 minutes at AP-1 target genes are explained by 1.27 to 1.39-fold increases in initiation rate and 0.85 to 0.93-fold decreases in pause release rate (Fig. 5B right). These relative changes in k_{init} and k_{rel} for AP-1 targets do result in gene activation, but it was unexpected that the profiles are explained by a decrease in k_{rel}. Since composite profiles represent the average of all included genes, it is possible that the composite represents a diverse set of genes that are regulated by different AP-1 family members. We speculate that we could gain a more clear insight if we were able to deconstruct the AP-1 targets and identify gene targets of specific AP-1 factors. The above analyses indicate that we can deconvolve complex transcriptional networks to identify gene targets of individual TFs and determine which steps in the transcription cycle each TF preferentially regulates.

We applied this model and approach to a separate PRO-seq dataset of the C7 B cell line treated with dexamethasone to determine whether GR regulates pause release within a different system in which GR is specifically activated. We identified 70 genes activated by dexamethasone treatment. The pause index of 80%

of these genes decrease between 0 and 60 minutes (Supplemental Fig. S5A). Compartment modeling of these genes showed that a ~1.33-fold increase in pause release rate explained the observed changes in PolII density at the activated genes (Supplemental Fig. S5B & C). These validation results support the role of GR regulating pause release and highlight the power of predicting the molecular function of transcription factors within complicated regulatory cascade networks generated from kinetic PRO and ATAC data.

TFs cooperate to bind REs and activate gene expression

AP-1, CEBP, GR, and KLF bind REs either individually or in combination in order to activate expression. We classify REs based on the combination of factors that bind and drive accessibility changes. Likewise, we classify genes based on which TFs are immediately upstream in the network. Genes activated by the same combination of factors can be downstream of different classes of REs. We use genes activated by all 4 of AP-1, CEBP, GR, and KLF to illustrate potential regulatory scenarios. These genes may be downstream of a single RE that binds all factors (Fig. 6A orange). Alternatively, the gene may be downstream of a pair of REs each binding 2 factors (Fig. 6A purple), 3 and 1 (Fig. 6A blue), or more complicated regulatory schemes (Fig. 6A green). All 15 possible classes of REs contribute to activation of the 82 genes downstream of AP-1, CEBP, GR, and KLF (Fig. 6B & S6A). The largest population of RE classes is isolated AP-1 peaks with 8794, while peaks bound by all activating factors is the smallest category with 74. The distribution of gene classes generally mirrors the distribution of RE classes, with isolated AP-1 genes being the largest class. As discussed above, all factors activate more genes in combination than in isolation. There are comparatively few combinatorially-regulated genes without AP-1 contribution (1924 with AP-1 v. 149 without). This finding, along with the high number of genes regulated by AP-1, underscores the importance of the AP-1 family in the network. While the bulk of negatively regulated genes are downstream of either SP or TWIST, approximately 20% are affected by both TWIST-mediated repression and SP-mediated attenuation (Supplemental Fig. S6B).

We did not observe a significant relationship between magnitude of RE accessibility change and number of regulatory factors (Supplemental Fig. S6C). We found that the relative change in transcription

positively correlates with the number of immediate upstream activators in the network (Supplemental Fig. S6D). Normalizing transcriptional change by local accessibility change eliminates the observed correlation between transcription and number of regulatory factors (Supplemental Fig. S6E). We confirmed this observation by plotting transcription of all predicted target genes against total local accessibility stratified by the number of regulatory factors and peaks (Supplemental Fig. S7). We find that more local regulatory peaks, which corresponds to greater total local accessibility, correlates with a greater magnitude of transcription. However, the number of regulatory factors largely does not affect transcription. Therefore, we find that transcription is positively correlated with total local accessibility change, regardless of the number of factors effecting that change. We conclude that if a gene is regulated in the network, the magnitude of expression change is independent from the number of upstream TFs.

Since the degree of activation and repression are unrelated to the number of upstream factors, we asked if having multiple TFs upstream in the network influences whether a gene is dynamic. To determine whether two TFs cooperate with one another we considered genes close to dynamic peaks with either a single TF motif or both TF motifs. We determine if the fraction of dynamic and nondynamic genes proximal to a TF is influenced by the presence of another TF. We define TF-proximal genes as genes that are close to dynamic ATAC peaks containing the TF motif. We find that there is no difference between the fraction of GR-proximal activated genes in the absence of AP-1. However, there is an increase in the fraction of activated genes proximal to GR in the presence of AP-1 (Fig. 6C). The reciprocal analysis shows that AP-1 is a more effective activator in the presence of GR (Fig. 6D). These results support the model that AP-1 and GR coordinate with one another to increase the likelihood of gene activation. The repressive factors TWIST and SP do not seem to work together in this way. The fraction of repressed genes proximal to TWIST increases regardless of the presence of SP (Fig. 6E). This suggests that TWIST functions largely independently of SP, supporting our hypothesis that the two TFs result in gene repression through unrelated mechanisms (Supplemental Fig. S3B & C). We find a lower proportion of repressed genes proximal to SP motifs both in the presence and absence of TWIST (Fig. 6F). We speculate that these genes tolerate dissociation of SP and maintain their expression levels despite local decreases in chromatin accessibility. In support of this explanation, we find higher basal

transcription and lower magnitude of repression in genes proximal to SP, suggesting these genes are more actively transcribed before loss of SP (Supplemental Fig. S6F & G). These results highlight the complexity of gene regulatory control and how kinetic networks reveal coordinate and independent relationships between transcription factors.

Multi-wave networks incorporate molecular dynamics and kinetic information

We further interrogate the adipogenesis gene regulatory network by leveraging temporal information to infer multiple waves of accessibility and transcriptional changes throughout the time course. The importance of TFs can be inferred by the number of predicted direct target genes (Fig. 6) or the total number of connected downstream genes. The latter is captured by temporal multi-wave network depictions. We assembled a representative multi-wave deep network (Fig. 7A). The differentiation cocktail induces AP-1 and GR binding to thousands of REs to activate thousands of genes; binding at 4 of these REs results in activation of the *Twist2* gene (Fig. 7A & B). The resulting TWIST2 protein returns to the nucleus and binds hundreds of REs and represses its target genes. Among the hundreds of repressed TWIST2 target genes are the late (40+ minutes) acting factors *Sp1* and *Sp3* (Fig. 7C). The decreased occupancy of SP transcription factors from the genome leads to decreases in RE accessibility and attenuation of gene expression. Our network suggests that SP1/3 dissociation and TWIST2 binding leads to repression of *Srf* (Fig. 7D). We hypothesize that if we were to extend the time course, we would identify the SRF binding motif in REs decreasing in accessibility beyond 4 hours as result of attenuated transcription.

Many genes activated in the early phase of the time course are repressed later on, either through active repression or factor dissociation. We detect these negative feedback loops for each activating TF (Supplemental Fig. S8A). About 63% of AP-1, 74% of CEBP, and 80% of GR *cis*-edges are transient. Only 27% of KLF *cis*-edges are attenuated, suggesting that KLF-mediated activation is less transient and less dependent on the extracellular stimuli found in the adipogenic cocktail. Similarly, a minority of TWIST *cis*-edges and no SP *cis*-edges are attenuated, indicating that SP and TWIST factors mediate sustained repression. A

much smaller proportion of *trans*-edges are attenuated, implying that accessibility changes downstream of factor binding and dissociation are more stable (Supplemental Fig. S8B) than changes in nascent transcription.

We find that regulatory potential for each TF varies greatly throughout the time course. AP-1, CEBP, and GR activate the most genes during the initial phase of the time course, indicating that these TFs precipitate the initial wave of signaling during the first 20 minutes. Transcriptional activation of TWIST and KLF family genes by the initial factors leads to the next wave of signaling after 20 minutes (Fig. 7B and Supplemental Fig. S8C). We begin to detect changes in accessibility at KLF- and TWIST-bound REs as early as 20 minutes (Supplemental Fig. S8D), however these presumptive binding events do not manifest as detectable changes in nascent transcription until 40 minutes (Fig. 7E). Although we had originally expected changes in accessibility and transcription to be observed concomitantly, these data show that we have the sensitivity to detect changes in RE accessibility before changes in transcription.

In addition to the TFs whose activity is stimulated by the adipogenesis cocktail, we identify transcriptionally regulated TF genes that are highly connected nodes within the network. The *Twist2* gene is the most highly connected node and directly affects accessibility and transcription of thousands of downstream nodes by binding REs and repressing proximal genes (Fig. 7F). TWIST2 acts through intermediate factors, such as SP, AP-1, GR, to repress thousands of additional genes. In the case of SP, TWIST2 mediated repression of *Sp1* and *Sp3*, results in SP dissociation and activation attenuation of downstream genes. TWIST2-mediated repression of AP-1 factors causes AP-1 dissociation and attenuation of AP-1-mediated activation. The cumulative result from both direct TWIST2 action and indirect dissociation / attenuation of TWIST2-targeted TF families affects accessibility at 12662 REs and 4574 genes. We believe that TWIST2 may have been overlooked as an important adipogenic TF because *Twist2* is only transiently activated (Supplemental Fig. S3C), but this kinetic network implicates TWIST2 as a critical intermediary in the adipogenesis cascade.

TWIST2 represses predicted target genes

We tested whether inferred TWIST2-repressed target genes from the network increase expression upon *Twist2* depletion. We used two different shRNA sequences (v1 & v2) to knockdown *Twist2* and harvested RNA for RNA-seq at 0, 1, 2, and 4 hours after switching cells into differentiation media. We observed a ~50%-75% reduction of *Twist2* expression prior to differentiation (Fig. 8A & Supplemental Fig. S9A). Unlike PRO-seq, RNA-seq requires mature mRNA accumulation above baseline signal to detect activation and RNA degradation to detect repression. Therefore many observed transcriptional changes at nascent RNA level take much longer to be detected at the mature RNA level. To this point, we find 32094 total genes expressed in the control RNA-seq data set compared to 13122 in the PRO-seq data set. This suggests that of all the genes detected by RNA-seq, only ~40% are actively transcribed. Furthermore, we identify 285 dynamic genes over the first hour of differentiation by RNA-seq versus 9201 by PRO-seq, underscoring the sensitivity of PRO-seq for detecting transcription changes over short time intervals. For RNA-seq analysis we only focused on the 520 predicted TWIST2 targets that are significantly repressed in both the PRO-seq and RNA-seq time courses. The other 547 predicted TWIST2 target genes from the network are not detected as repressed by conventional RNA-seq, so we would have no power to detect derepression upon *Twist2* depletion. Approximately 66% of the examined genes were expressed at higher levels at baseline in the *Twist2* knockdown as compared to the control, supporting our hypothesis that TWIST2 directly represses the majority of our predicted targets (Fig. 8B & Supplemental Fig. S9B). This result is likely an underestimate of the specificity of our network, because the cells can compensate for chronic RNAi-mediated depletion of *Twist2*.

We then measured the effect of chronic *Twist2* depletion on differentiation-induced transcription. *Twist2* was activated in both the control and the knockdown samples, supporting our PRO-seq results (Fig. 8A & Supplemental Fig. S3C). *Twist2* was more strongly activated in the knockdown sample than in the control (~3.1-fold to ~2-fold). This is likely because the RNAi machinery cannot keep up with the dynamic accumulation of *Twist2* transcripts in the hours following treatment with the differentiation cocktail. Since we inferred that TWIST2 represses target genes in this system, then we would expect a greater degree of repression over the time course in the knockdown samples due to the relatively greater accumulation of TWIST2 protein. This analysis more directly tests the accuracy of predicting TWIST2-target genes compared to

chronic knockdown. We find that ~75% of the predicted targets are repressed to a greater magnitude in the *shTwist2* samples compared to the control knockdown (Fig. 8C). We observe supporting results with a second, less effective knockdown (Supplemental Fig. S9C). The above findings support the conclusions from the network and indicate that TWIST2 is a transcriptional repressor of predicted target genes in 3T3-L1 differentiation.

TWIST2 influences differentiation of 3T3-L1s and primary preadipocytes

To test TWIST2's effect on differentiation of preadipocytes, we depleted TWIST2, induced differentiation of 3T3-L1s, and measured lipid uptake after 6 days of differentiation. We stained differentiated adipocytes with Oil Red O and measured absorbance at 540nm to quantify lipid uptake. Lipid accumulation is a cellular phenotype that acts as a proxy measurement for adipogenesis. Lipid uptake increased with shTwist2 treatment compared to shControl within each experiment, suggesting that TWIST2 expression negatively regulates differentiation in the 3T3-L1 system (Fig. 8D). As an orthogonal approach, we designed and transduced a tetracycline-inducible 3xFLAG-tagged human *Twist2* construct into 3T3-L1 cells (Supplemental Fig. S9D). After 6 days of differentiation, 3T3-L1s overexpressing *Twist2* exhibited decreased Oil Red O staining (Fig. 8E & Supplemental Fig. S9E), supporting our previous finding that TWIST2 expression reduces 3T3-L1 differentiation.

Next, we extracted preadipocytes from inguinal white adipose tissue (WAT) of 3 day old *Twist2*$^{+/-}$ pups. We induced differentiation in the primary preadipocytes and found that preadipocytes derived from heterozygous mice differentiated to a greater extent than those derived from wild type mice (Fig. 8F & Supplemental Fig. S9F). The 3T3-L1 and primary cultured preadipocyte results indicate that TWIST2 opposes induced differentiation in both *in vitro* and *ex vivo* contexts.

We found that *Twist2*$^{+/-}$ mice have a deficiency of dermal and subcutaneous white adipose tissue in the skin (Fig. 8G). *Twist2*$^{-/-}$ mice have fewer and smaller fat droplets within interscapular brown adipose tissue

(BAT) deposits (Fig. 8H). Other groups have reported loss of subcutaneous fat and a paucity of fat storage in *Twist2*$^{-/-}$ mice (Kim et al. 2022; Šošić et al. 2003; Tukel et al. 2010). We postulate that TWIST2 acts as a 'brake' on adipogenesis, preventing cell exhaustion and apoptosis during the differentiation process. Regulated braking of adipogenesis may be necessary to allow supportive adipose tissues to sufficiently develop in the mouse. Isolated preadipocytes may be able to overcome the additional stress *in vitro*, but not within their native tissue context.

DISCUSSION

Kinetic accessibility and nascent transcriptional profiling of developmental cascades can identify key regulatory nodes that may be transiently active, but are nonetheless necessary for proper cellular differentiation. We present an extremely rapid and precise capture of chromatin and transcription changes induced by an adipogenic cocktail. These changes represent the first few waves of differentiation signaling and precipitate the cellular transition process. RE accessibility and gene transcription change within minutes of initiating adipogenesis. By focusing only on dynamically accessible REs, we can infer TF binding and dissociation events that drive adipogenesis without performing hundreds of genomic ChIP experiments. We find a multitude of enriched TF family motifs, many of which have been previously associated with adipogenic REs including AP-1, GR, KLF, and CEBP (Siersbæk et al. 2014). We do not identify PPARG, the master regulator of adipogenesis (Lefterova et al. 2014; Rosen et al. 2002), as a driver of adipogenic signaling. This agrees with previous conclusions that PPARG does not influence adipogenesis until several days into the process (Nielsen et al. 2008). Stable PPARG activity is indispensable for adipogenesis and maintaining adipocyte identity, but other factors may be critically important and overlooked because their role is transient.

Our method implicates TWIST2 as a novel contributor to adipogenesis. The TWIST subfamily of bHLH TFs homo- and heterodimerize with other bHLH proteins to affect gene expression. Although TWIST family factors all recognize the same DNA motif, different members can act as either activators or repressors. TWIST proteins can repress transcription by non-productive dimerization with TWIST family activators, competing with

TWIST family activators for DNA motifs, or by recruiting chromatin condensers like HDACs to the genome (Bialek et al. 2004; Gong and Li, 2002; Hamamori et al. 1999; Hayashi et al. 2007; Koh et al. 2009; Lee et al. 2003; Šošić et al. 2003). Previous studies have implicated TWIST2's role in targeting corepressors (Fu et al. 2011; Kim et al. 2022). While multiple mechanisms may be at play in our system, we hypothesize that our observed repressive effects are downstream of increased TWIST2 binding. TWIST1 and TWIST2 negatively regulate multiple developmental pathways including myogenesis, osteogenesis, and myeloid differentiation (Bialek et al. 2004; Gong and Li, 2002; Hebrok et al. 1994; Murray et al. 1992; Sharabi et al. 2008; Spicer et al. 1996). The role of the TWIST TF family in adipogenesis is less clear. While TWIST1 and TWIST2 are known regulators of mature adipose tissue homeostasis, TWIST1 does not affect adipogenesis (Dobrian, 2012; Lee et al. 2003; Pan et al. 2009). Homozygous *Twist2* mutations cause Setleis syndrome, a disease characterized by facial lesions lacking subcutaneous fat (Tukel et al. 2010). *Twist2* knockout mice develop such lesions and lack lipid droplets within the liver and brown fat tissue (Fig. 8G & H) (Šošić et al. 2003; Tukel et al. 2010). Our *in vitro* and *ex vivo* data indicate that TWIST2 acts as a negative regulator of adipogenesis. TWIST2's immediate activation in 3T3-L1 differentiation therefore indicates a negative feedback mechanism to slow differentiation. The TWIST family is a key regulator of the epithelial-mesenchymal transition, further supporting our observation that TWIST2 prevents 3T3-L1 differentiation (Yang et al. 2004). The loss of this negative feedback may result in cell death, leading to the absence of adipose tissue observed *in vivo*. Even with a well-studied system such as adipogenesis, these methods were able to identify *Twist2* as a novel regulator of the differentiation cascade.

Our networks define gene sets that are predominantly regulated by a single TF. We can track changes in RNA polymerse density within the gene sets to identify the target regulatory steps of individual TFs. Stimulated pause release is an established cause of early gene activation in adipogenesis (Wang et al. 2021). We find that GR is largely responsible for the observed increase in pause release. GR is a well-established activator of gene expression (Vockley et al. 2016), often in combination with AP-1 (Biddie et al. 2011). Other activating factors, including AP-1, increase RNA polymerase recruitment to the gene. By acting on separate steps, GR and AP-1 provide non-redundant stimuli to target genes. We find GR and AP-1 are conditionally

dependent upon one another in their potential to activate local genes. A recent study suggests that both AP-1 and CEBP act as pioneer factors that prime the genome for GR-induced transcription activation (Wissink et al. 2019). We find all three of these factor families activate the initial wave of transcription changes, both in combination and in isolation.

We confidently differentiate primary, secondary, and tertiary transcriptional changes by examining multiple, closely spaced time points upon induced adipogenesis. ATAC-seq, ChIP-seq, or chromatin conformation assays alone can only suggest functional relationships between REs and genes (Lieberman-Aiden et al. 2009; Ren et al. 2000). Similarly PRO-seq and RNA-seq return transcription changes with little information regarding upstream regulation. We define individual regulatory relationships more directly by focusing only on the ATAC peaks and genes that are proximal and covary in their dynamics (i.e. significantly change over in the same direction the time course). Our bipartite directed graph networks are unique in the gene regulation field because each edge represents a functional interaction as opposed to an abstract relationship between linked nodes. *Trans* edges represent binding of TF proteins to cognate DNA elements and *cis*-edges describe regulatory interactions between REs and target genes. These networks can define gene sets that are predominantly regulated by a single TF and identify the target regulatory steps of the TF. Highly connected nodes in the network are candidate key regulatory hubs in the differentiation cascade. Moreover, these networks ascribe time attributes to each edge, so subgraphs that respect the flow of time are easily extracted from the larger graph. This integrative genomics approach to network construction can be applied to a multitude of cellular responses and transitions to uncover novel biology and new hypotheses.

METHODS

3T3-L1 culture and differentiation

3T3-L1 cells were provided by Thurl Harris. 3T3-L1 cells were cultured in high glucose DMEM (Gibco) supplemented with 10% newborn calf serum, 1% fetal bovine serum (FBS), 100 U/mL Penicillin G, and 100

µg/mL streptomycin. We induced adipogenesis ~3 days after cells reached confluency by switching cells into high glucose DMEM supplemented with 0.25 µM Dexamethasone, 0.5 mM 3-isobutyl-1-methylxanthine, 2.5 units/mL insulin, 10% FBS, 100 U/mL Penicillin G, and 100 µg/mL streptomycin (Bernlohr et al. 1984; Green and Kehinde, 1974). We collected enough cells at the indicated time points for three replicates of ATAC-seq and PRO-seq.

ATAC-seq library preparation

We prepared ATAC-seq libraries as previously described (Corces et al. 2017). We trypsinized and collected cells in serum-free growth media. We counted ~5 × 10^4 cells per replicate and transferred them to 1.5 mL tubes. We centrifuged cells at 500 × *g* for 5 minutes at 4°C and resuspended the pellet in 50 µL cold lysis buffer (10mM Tris-HCl, 10mM NaCl, 3mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin, adjusted to pH 7.4) and incubated on ice for 3 minutes. We washed the samples with 1 mL cold wash buffer (10mM Tris-HCl, 10mM NaCl, 3mM $MgCl_2$, 0.1% Tween-20). We centrifuged at 500 × *g* for 10 minutes at 4°C and resuspended cells in the transposition reaction mix (25 µL 2X TD buffer (Illumina), 2.5 µL TDE1 Tn5 transposase (Illumina), 16.5 µL PBS, 0.5 µL 1% Digitonin, 0.5 µL 10% Tween-20, 5 µL nuclease-free water) and incubated at 37°C for 30 minutes. We extracted DNA with the MinElute kit (Qiagen). We attached sequencing adapters to the transposed DNA fragments using the Nextera XT Index Kit (Illumina) and amplified libraries with 6 cycles of PCR. We performed PEG-mediated size fractionation (Lis, 1980) on our libraries by mixing SPRIselect beads (Beckman) with our sample at a 0.55:1 ratio, then placing the reaction vessels on a magnetic stand. We transferred the right side selected sample to a new reaction vessel and added more beads for a final ratio of 1.8:1. We eluted the final size-selected sample into nuclease-free water.

ATAC-seq analyses

We aligned reads to the *mm10* mouse genome assembly with Bowtie 2, sorted output *BAM* files with SAMtools, and converted files to *bigWig* format with seqOutBias (Langmead and Salzberg, 2012; Li et al.

2009; Martins et al. 2018). We called accessibility peaks with MACS2 (Gaspar, 2018; Zhang et al. 2008). We sorted reads into peaks using the bigWig R package and identified differentially accessible REs with DESeq2 (Love et al. 2014; Martins, n.d.). We cluster dynamic peaks into response groups using DEGreport (Pantano, n.d.). We performed *de novo* motif extraction on dynamic REs with MEME (e-value cutoff of 0.01) and used TOMTOM (e-value cutoff of 0.05) to match motifs to the HOMER, JASPAR, and UniPROBE TF binding motif databases (Bailey et al. 2015; Heinz et al. 2010; Khan et al. 2018). We use FIMO to identify genome-wide motif occurrences (Cuellar-Partida et al. 2012). We generated DNA sequence logos with ceqLogo (Bailey et al. 2015). We use the bigWig package to assess motif enrichment around ATAC-seq peak summits (Martins, n.d.).

PRO-seq library preparation

We performed cell permeabilization as previously described (Mahat et al. 2016). We trypsinized and collected cells in 10 mL ice cold PBS followed by washing in 5 mL buffer W (10 mM Tris-HCl pH 7.5, 10 mM KCl, 150 mM sucrose, 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 0.5 mM DTT, 0.004 U/mL SUPERaseIN RNase inhibitor (Invitrogen), protease inhibitors (cOmplete, Roche)). We permeabilized cells by incubating with buffer P (10 mM Tris-HCl pH 7.5, KCl 10 mM, 250 mM sucrose, 5 mM $MgCl_2$, 1 mM EGTA, 0.05% Tween-20, 0.1% NP-40, 0.5 mM DTT, 0.004 units/mL SUPERaseIN RNase inhibitor (Invitrogen), protease inhibitors (cOmplete, Roche)) for 3 minutes. We washed cells with 10 mL buffer W before transferring into 1.5 mL tubes using wide bore pipette tips. Finally, we resuspended cells in 500 µL buffer F (50 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 0.1 mM EDTA, 50% Glycerol and 0.5 mM DTT). After counting nuclei, we separated cells into 50 µL aliquots with ~3-5 $\times$ 10^5 cells each. We snap froze our aliquots in liquid nitrogen and stored them at -80°C. We prepared PRO-seq libraries as previously described (Sathyan et al. 2019). We included a random eight base unique molecular identifier (UMI) at the 5′ end of the adapter ligated to the 3′ end of the nascent RNA. We did not perform any size selection in an attempt to not bias our libraries against short nascent RNAs. Raw PRO-seq sequencing files and processed *bigWig* files are available from GEO accession record GSE133147.

PRO-seq analyses

First we used cutadapt to remove adapters from our reads (Martin, 2011). We used fqdedup and the 3′ UMIs to deduplicate our libraries (Martins and Guertin, n.d.). Next we removed UMIs and converted reads to their reverse complement with the FASTX-Toolkit (Gordon, n.d.). As with the ATAC-seq samples, we used Bowtie 2, SAMtools, and seqOutBias to align, sort, and convert reads to *bigWig* files respectively (Langmead and Salzberg, 2012; Li et al. 2009; Martins et al. 2018). We used primaryTranscriptAnnotation to adjust gene annotations based on our PRO-seq data (Anderson et al. 2020). We queried the *bigWig* files within the adjusted genomic coordinates with the bigWig R package and UCSC Genome Browser Utilities (Kent et al. 2010; Martins, n.d.). We identified differentially expressed genes with DESeq2 (Love et al. 2014). We used dREG to define peaks of bidirectional transcription from our *bigWig* files (Wang et al. 2019). As with the ATAC-seq samples, we identified overrepresented motifs in dREG-defined REs with MEME and TOMTOM (Bailey et al. 2015). We evaluate motif enrichment around peak summits and polymerase density in the gene body and pause region with the bigWig package (Martins, n.d.). We define the summit of the pause peak for genes by first identifying the point of maximum density within 1 kb of the TSS. We define the pause region as the 50 bp window around the summit.

Network construction

The bipartite directional networks with gene and RE nodes were inferred using a data-driven rules-based approach. The first rule to infer *trans*-edges from TF families to individual REs is that the RE must contain the cognate motif for the TF family. The second is that the peak must be dynamically accessible over some part of the time course. The third is that at least one gene encoding a member of the TF family must be expressed and activated (or in the case of SP, repressed) over the same time range. We restrict *trans*-edges attributed to GR to the first 40 minutes of the time course for reasons discussed in the text. Similarly, we do not draw edges from SP before 40 minutes. Next, we drew *cis*-edges between REs and proximal genes based on a different rule set. First, REs need to be within 10 kb of gene bodies as defined by primary transcript

annotation of our PRO-seq data. We used BEDTools to find gene-RE pairs that satisfied this rule (Quinlan and Hall, 2010). Next, the peak and the gene need to covary in accessibility and transcription during the same time range. For example, a gene must be activated at the same time as its local RE is increasing in accessibility. We refined the distance requirements by incorporating constraints from our CDF analysis. For each activating factor (AP-1, CEBP, GR, KLF) we find a set of pairwise comparisons within the time course for which factor REs are significantly closer to activated than nondynamic genes. We find a similar set of comparisons for repressive factors (SP, TWIST). For a gene to be linked to a factor RE with a *cis*-edge, we require that the gene must be dynamic in at least one of the comparisons identified by the CDF analysis for that factor. In addition, our CDF analysis also identifies the maximum distance between a factor RE and a regulated gene for each comparison. The RE and the gene's TSS must be within the relevant distance threshold defined by the CDF.

DATA ACCESS

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE150492 (3T3-L1 ATAC-seq), GSE219041 (C7 cells + dexamethasone PRO-seq), and GSE219051 (shTwist2 3T3-L1 RNA-seq). A supplemental code document contains all code used for data analysis and can also be found at https://github.com/guertinlab/adipogenesis.

AUTHOR CONTRIBUTIONS

ABD, BN, and MJG analyzed the data. NMW and FMD performed the 3T3-L1 ATAC-seq and PRO-seq experiments. ABD and DSL performed the 3T3-L1 and primary cell adipogenesis assays. RP, PP, and ABD performed the mouse tissue harvesting and preadipocyte extraction. LW designed and cloned the TWIST2 overexpression construct. ABD and MJG conceptualized and developed the project. ABD and MJG wrote the manuscript.

COMPETING INTEREST STATEMENT

The authors have no competing interests to disclose.

ACKNOWLEDGMENTS

This work was funded by R35-GM128635 to MJG. T32-LM012416 supported ABD. K00-CA253732 supported RKP. R50-CA265089 supported LW. We thank John Lukens for hosting animals. We thank Thurl Harris for providing 3T3-L1 cells and technical guidance. We thank Kevin Janes for providing reagents and suggestions. We thank Sathyan Mattada, Thomas Scott, Jacob Wolpe, Sailasree Rajalekshmi, and Theresa Gibney for critical feedback.

REFERENCES

Ahmad B, Serpell CJ, Fong IL, Wong EH. 2020. Molecular Mechanisms of Adipogenesis: The Anti-adipogenic Role of AMP-Activated Protein Kinase. *Front Mol Biosci* **7**:76.

Anderson WD, Duarte FM, Civelek M, Guertin MJ. 2020. Defining data-driven primary transcript annotations with primaryTranscriptAnnotation in R. *Bioinformatics* **36**:2926–2928.

Ardehali MB, Lis JT. 2009. Tracking rates of transcription and splicing in vivo. *Nat Struct Mol Biol*.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* **43**:W39–49.

Banerjee SS, Feinberg MW, Watanabe M, Gray S, Haspel RL, Denkinger DJ, Kawahara R, Hauner H, Jain MK. 2003. The Krüppel-like factor KLF2 inhibits peroxisome proliferator-activated receptor-γ expression and adipogenesis. *J Biol Chem* **278**:2581–2584.

Benner C, Konovalov S, Mackintosh C, Hutt KR, Stunnenberg R, Garcia-Bassets I. 2013. Decoding a signature-based model of transcription cofactor recruitment dictated by cardinal cis-regulatory elements in proximal promoter regions. *PLoS Genet* **9**:e1003906.

Bernlohr DA, Angus CW, Lane MD, Bolanowski MA, Kelly TJ Jr. 1984. Expression of specific mRNAs during adipose differentiation: identification of an mRNA encoding a homologue of myelin P2 protein. *Proc Natl Acad Sci U S A* **81**:5468–5472.

Bialek P, Kern B, Yang X, Schrock M, Sosic D, Hong N, Wu H, Yu K, Ornitz DM, Olson EN, Justice MJ, Karsenty G. 2004. A twist code determines the onset of osteoblast differentiation. *Dev Cell* **6**:423–435.

Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, Miranda TB, Sung M-H, Trump S, Lightman SL, Vinson C, Stamatoyannopoulos JA, Hager GL. 2011. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* **43**:145–155.

Birsoy K, Chen Z, Friedman J. 2008. Transcriptional regulation of adipogenesis by KLF4. *Cell Metab* **7**:339–347.

Boyle AP, Davis S, Shulha HP, Meltzer 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, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* **109**:21–29.

Chen J, Zhang Z, Li L, Chen B-C, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E, Tjian R, Liu Z. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**:1274–1285.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**:959–962.

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**:1311–1320.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**:1845–1848.

Cuellar-Partida G, Buske FA, McLeay RC, Whitington T, Noble WS, Bailey TL. 2012. Epigenetic priors for identifying active transcription factor binding sites. Bioinformatics 28: 56–62.

Danko CG, Hah N, Luo X, Martins AL, Core L, Lis JT, Siepel A, Kraus WL. 2013. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol Cell* **50**:212–

222.

Distel RJ, Ro HS, Rosen BS, Groves DL, Spiegelman BM. 1987. Nucleoprotein complexes that regulate gene expression in adipocyte differentiation: direct participation of c-fos. *Cell* **49**:835–844.

Dobrian AD. 2012. A tale with a Twist: a developmental gene with potential relevance for metabolic dysfunction and inflammation in adipose tissue. *Front Endocrinol* **3**:108.

Duarte FM, Fuda NJ, Mahat DB, Core LJ, Guertin MJ, Lis JT. 2016. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev* **30**:1731–1746.

Farrar M. 2006. Striped Smith–Waterman speeds database searches six times over other SIMD implementations. *Bioinformatics* **23**:156–161.

Flodby P, Barlow C, Kylefjord H, Ahrlund-Richter L, Xanthopoulos KG. 1996. Increased hepatic cell proliferation and lung abnormalities in mice deficient in CCAAT/enhancer binding protein alpha. *J Biol Chem* **271**:24753–24760.

Freytag SO, Paielli DL, Gilbert JD. 1994. Ectopic expression of the CCAAT/enhancer-binding protein alpha promotes the adipogenic program in a variety of mouse fibroblastic cells. *Genes Dev* **8**:1654–1663.

Fuda NJ, Ardehali MB, Lis JT. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**:186–192.

Fu J, Qin L, He T, Qin J, Hong J, Wong J, Liao L, Xu J. 2011. The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Res* **21**:275–289.

Galmozzi A, Kok BP, Saez E. 2021. Isolation and Differentiation of Primary White and Brown Preadipocytes from Newborn Mice. *J Vis Exp*. doi:10.3791/62005

Gaspar JM. 2018. Improved peak-calling with MACS2. *bioRxiv*. doi:10.1101/496521

Ghaben AL, Scherer PE. 2019. Adipogenesis and metabolic health. *Nat Rev Mol Cell Biol* **20**:242–258.

Gill G, Pascal E, Tseng ZH, Tjian R. 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. *Proceedings of the National Academy of Sciences* **91**:192–196.

Gong XQ, Li L. 2002. Dermo-1, a multifunctional basic helix-loop-helix protein, represses MyoD transactivation via the HLH domain, MEF2 interaction, and chromatin deacetylation. *J Biol Chem* **277**:12310–12317.

Gordon A. n.d. Fastx toolkit. *https://github com/agordon/fastx_toolkit*.

Green H, Kehinde O. 1974. Sublines of mouse 3T3 cells that accumulate lipid. *Cell* **1**:113–116.

Guertin MJ, Lis JT. 2013. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr Opin Genet Dev* **23**:116–123.

Guertin MJ, Lis JT. 2010. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* **6**:e1001114.

Guertin MJ, Martins AL, Siepel A, Lis JT. 2012. Accurate prediction of inducible transcription factor binding intensities in vivo. *PLoS Genet* **8**:e1002610.

Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. 2011. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**:622–634.

Hamamori Y, Sartorelli V, Ogryzko V, Puri PL, Wu HY, Wang JY, Nakatani Y, Kedes L. 1999. Regulation of histone acetyltransferases p300 and PCAF by the bHLH protein twist and adenoviral oncoprotein E1A. *Cell* **96**:405–413.

Hayashi M, Nimura K, Kashiwagi K, Harada T, Takaoka K, Kato H, Tamai K, Kaneda Y. 2007. Comparative roles of Twist-1 and Id1 in transcriptional regulation by BMP signaling. *J Cell Sci* **120**:1350–1357.

Hebrok M, Wertz K, Füchtbauer EM. 1994. M-twist is an inhibitor of muscle differentiation. *Dev Biol* **165**:537–544.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**:576–589.

Hsieh N-H, Resifeld B, Chiu WA. 2020. An R package to apply global sensitivity analysis in physiologically based kinetic modeling. *SoftwareX* **12**. http://dx.doi.org/10.1016/j.softx.2020.100609.

Inuzuka H, Wakao H, Masuho Y, Muramatsu MA, Tojo H, Nanbu-Wakao R. 1999. cDNA cloning and expression analysis of mouse zf9, a Krüppel-like transcription factor gene that is induced by adipogenic hormonal stimulation in 3T3-L1 cells. *Biochim Biophys Acta* **1447**:199–207.

Janes KA. 2015. An analysis of critical factors for quantitative immunoblotting. *Sci Signal* **8**:rs2.

Jonkers I, Lis JT. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol*

Cell Biol **16**:167–177.

Kawamura Y, Tanaka Y, Kawamori R, Maeda S. 2006. Overexpression of Kruppel-like factor 7 regulates adipocytokine gene expressions in human adipocytes and inhibits glucose-induced insulin secretion in pancreatic β-cell line. *Mol Endocrinol* **20**:844–856.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**:996–1006.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**:2204–2207.

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**:D1284.

Kim JY, Park M, Ohn J, Seong RH, Chung JH, Kim KH, Jo SJ, Kwon O. 2022. Twist2-driven chromatin remodeling governs the postnatal maturation of dermal fibroblasts. *Cell Rep* **39**:110821.

Koh HS, Lee C, Lee KS, Park EJ, Seong RH, Hong S, Jeon SH. 2009. Twist2 regulates CD7 expression and galectin-1-induced apoptosis in mature T-cells. *Mol Cells* **28**:553–558.

Kraus NA, Ehebauer F, Zapp B, Rudolphi B, Kraus BJ, Kraus D. 2016. Quantitative assessment of adipocyte differentiation in cell culture. *Adipocyte* **5**:351–358.

Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**:950–953.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359.

Lee YS, Lee HH, Park J, Yoo EJ, Glackin CA, Choi YI, Jeon SH, Seong RH, Park SD, Kim JB. 2003. Twist2, a novel ADD1/SREBP1c interacting protein, represses the transcriptional activity of ADD1/SREBP1c. *Nucleic Acids Res* **31**:7165–7174.

Lefterova MI, Haakonsson AK, Lazar MA, Mandrup S. 2014. PPARγ and the global map of adipogenesis and beyond. *Trends Endocrinol Metab* **25**:293–302.

Lefterova MI, Lazar MA. 2009. New developments in adipogenesis. *Trends Endocrinol Metab* **20**:107–114.

Li D, Yea S, Li S, Chen Z, Narla G, Banck M, Laborda J, Tan S, Friedman JM, Friedman SL, Walsh MJ. 2005. Krüppel-like factor-6 promotes preadipocyte differentiation through histone deacetylase 3-dependent repression of DLK1. *J Biol Chem* **280**:26941–26952.

Li S, Chen K, Zhang Y, Barnes SD, Jaichander P, Zheng Y, Hassan M, Malladi VS, Skapek SX, Xu L, et al. 2019. Twist2 amplification in rhabdomyosarcoma represses myogenesis and promotes oncogenesis by redirecting MyoD DNA binding. *Genes Dev* **33**: 626-640.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**:289–293.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079.

Li H, Quang D, Guan Y. 2019. Anchor: trans-cell type prediction of transcription factor binding sites. *Genome Res* **29**:281–292.

Lis JT. 1980. [42] Fractionation of DNA fragments by polyethylene glycol induced precipitationMethods in Enzymology. Academic Press. pp. 347–353.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550.

Madsen JGS, Madsen MS, Rauch A, Traynor S, Van Hauwaert EL, Haakonsson AK, Javierre BM, Hyldahl M, Fraser P, Mandrup S. 2020. Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. *Nat Genet* **52**:1227–1238.

Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT. 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**:1455–1476.

Marshall NF, Price DH. 1995. Purification of P-TEFb, a Transcription Factor Required for the Transition into Productive Elongation (□). *J Biol Chem* **270**:12335–12338.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10–12.

Martins AL. n.d. R interface to query UCSC bigWig files. *https://github com/andrelmartins/bigWig*.

Martins AL, Guertin MJ. n.d. Remove PCR duplicates from FASTQ files. *https://github com/guertinlab/fqdedup*.

Martins AL, Walavalkar NM, Anderson WD, Zang C, Guertin MJ. 2018. Universal correction of enzymatic sequence bias reveals molecular signatures of protein/DNA interactions. *Nucleic Acids Res* **46**:e9.

McKnight SL, Kingsbury R. 1982. Transcriptional control signals of a eukaryotic protein-coding gene. *Science* **217**:316–324.

McNally JG, Müller WG, Walker D, Wolford R, Hager GL. 2000. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287**:1262–1265.

Moitra J, Mason MM, Olive M, Krylov D, Gavrilova O, Marcus-Samuels B, Feigenbaum L, Lee E, Aoyama T, Eckhaus M, Reitman ML, Vinson C. 1998. Life without white fat: a transgenic mouse. *Genes Dev* **12**:3168–3181.

Mori T, Sakaue H, Iguchi H, Gomi H, Okada Y, Takashima Y, Nakamura K, Nakamura T, Yamauchi T, Kubota N, Kadowaki T, Matsuki Y, Ogawa W, Hiramatsu R, Kasuga M. 2005. Role of Krüppel-like factor 15 (KLF15) in transcriptional regulation of adipogenesis. *J Biol Chem* **280**:12867–12875.

Murray SS, Glackin CA, Winters KA, Gazit D, Kahn AJ, Murray EJ. 1992. Expression of helix-loop-helix regulatory genes during differentiation of mouse osteoblastic cells. *J Bone Miner Res* **7**:1131–1138.

Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* **39**:1507–1511.

Neumayr C, Haberle V, Serebreni L, Karner K, Hendy O, Boija A, Henninger JE, Li CH, Stejskal K, Lin G, Bergauer K, Pagani M, Rath M, Mechtler K, Arnold CD, Stark A. 2022. Differential cofactor dependencies define distinct types of human enhancers. *Nature* **606**:406–413.

Nielsen R, Pedersen TÅ, Hagenbeek D, Moulos P, Siersbæk R, Megens E, Denissov S, Børgesen M, Francoijs K-J, Mandrup S, Others. 2008. Genome-wide profiling of PPARγ: RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev* **22**:2953–2967.

Pan D, Fujimoto M, Lopes A, Wang Y-X. 2009. Twist-1 Is a PPARδ-Inducible, Negative-Feedback Regulator of PGC-1α in Brown Fat Metabolism. *Cell* **137**:73–86.

Pantano L. n.d. DEGreport: Report of DEG analysis. *New Jersey, NJ: R package version*.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**:2444–2448.

Pei H, Yao Y, Yang Y, Liao K, Wu J-R. 2011. Krüppel-like factor KLF9 regulates PPARγ transactivation at the middle stage of adipogenesis. *Cell Death Differ* **18**:315–327.

Ptashne M. 1967. Specific Binding of the λ Phage Repressor to λ DNA. *Nature* **214**:232–234.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.

Ramji DP, Foka P. 2002. CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem J* **365**:561–575.

Rasmussen EB, Lis JT. 1993. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A* **90**:7923–7927.

Rauch A, Haakonsson AK, Madsen JGS, Larsen M, Forss I, Madsen MR, Van Hauwaert EL, Wiwie C, Jespersen NZ, Tencerova M, Nielsen R, Larsen BD, Röttger R, Baumbach J, Scheele C, Kassem M, Mandrup S. 2019. Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. *Nat Genet* **51**:716–727.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290**:2306–2309.

Rosen ED, Hsu C-H, Wang X, Sakai S, Freeman MW, Gonzalez FJ, Spiegelman BM. 2002. C/EBPα induces adipogenesis through PPARγ: a unified pathway. *Genes Dev* **16**:22–26.

Rosen ED, Spiegelman BM. 2006. Adipocytes as regulators of energy balance and glucose homeostasis. *Nature* **444**:847–853.

Rougvie AE, Lis JT. 1988. The RNA polymerase II molecule at the 5’ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell* **54**:795–804.

Rubin CS, Hirsch A, Fung C, Rosen OM. 1978. Development of hormone receptors and hormonal responsiveness in vitro. Insulin receptors and insulin sensitivity in the preadipocyte and adipocyte forms of 3T3-L1 cells. *J Biol Chem* **253**:7570–7578.

Sathyan KM, McKenna BD, Anderson WD, Duarte FM, Core L, Guertin MJ. 2019. An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion. *Genes Dev* **33**:1441–1455.

Scholes C, DePace AH, Sánchez Á. 2017. Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle. *Cell Syst* **4**:97–108.e9.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent transcription from active promoters. *Science* **322**:1849–1851.

Sharabi AB, Aldrich M, Sosic D, Olson EN, Friedman AD, Lee S-H, Chen S-Y. 2008. Twist-2 controls myeloid lineage development and function. *PLoS Biol* **6**:e316.

Siersbæk R, Nielsen R, John S, Sung M-H, Baek S, Loft A, Hager GL, Mandrup S. 2011. Extensive chromatin remodelling and establishment of transcription factor “hotspots” during early adipogenesis. *EMBO J* **30**:1459–1472.

Siersbæk R, Nielsen R, Mandrup S. 2012. Transcriptional networks and chromatin remodeling controlling adipogenesis. *Trends Endocrinol Metab* **23**:56–64.

Siersbæk R, Rabiee A, Nielsen R, Sidoli S, Traynor S, Loft A, Poulsen LLC, Rogowska-Wrzesinska A, Jensen ON, Mandrup S. 2014. Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell Rep* **7**:1443–1455.

Šošić D, Richardson JA, Yu K, Ornitz DM, Olson EN. 2003. Twist regulates cytokine gene expression through a negative feedback loop that represses NF-kappaB activity. *Cell* **112**:169–180.

Spicer DB, Rhee J, Cheung WL, Lassar AB. 1996. Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. *Science* **272**:1476–1480.

Spradling A, Penman S, Pardue ML. 1975. Analysis of drosophila mRNA by in situ hybridization: sequences transcribed in normal and heat shocked cultured cells. *Cell* **4**:395–404.

Steger DJ, Grant GR, Schupp M, Tomaru T, Lefterova MI, Schug J, Manduchi E, Stoeckert CJ Jr, Lazar MA.

2010. Propagation of adipogenic signals through an epigenomic transition state. *Genes Dev* **24**:1035–1044.

Sue N, Jack BHA, Eaton SA, Pearson RCM, Funnell APW, Turner J, Czolij R, Denyer G, Bao S, Molero-Navajas JC, Perkins A, Fujiwara Y, Orkin SH, Bell-Anderson K, Crossley M. 2008. Targeted disruption of the basic Krüppel-like factor gene (Klf3) reveals a role in adipogenesis. *Mol Cell Biol* **28**:3967–3978.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**:663–676.

Tanaka T, Yoshida N, Kishimoto T, Akira S. 1997. Defective adipocyte differentiation in mice lacking the C/EBPβ and/or C/EBPδ gene. *EMBO J* **16**:7432–7443.

Tang QQ, Jiang MS, Lane MD. 1999. Repressive effect of Sp1 on the C/EBPalpha gene promoter: role in adipocyte differentiation. *Mol Cell Biol* **19**:4855–4865.

Thompson B, Varticovski L, Baek S, Hager GL. 2016. Genome-Wide Chromatin Landscape Transitions Identify Novel Pathways in Early Commitment to Osteoblast Differentiation. *PLoS One* **11**:e0148619.

Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, Gnirke A, Meissner A. 2015. Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**:344–349.

Tukel T, Šošić D, Al-Gazali LI, Erazo M, Casasnovas J, Franco HL, Richardson JA, Olson EN, Cadilla CL, Desnick RJ. 2010. Homozygous nonsense mutations in TWIST2 cause Setleis syndrome. *Am J Hum Genet* **87**:289–296.

Unamuno X, Gómez-Ambrosi J, Rodríguez A, Becerril S, Frühbeck G, Catalán V. 2018. Adipokine dysregulation and adipose tissue inflammation in human obesity. *Eur J Clin Invest* **48**:e12997.

van Kruijsdijk R, Van Der Wall E, Visseren FLJ. 2009. Obesity and Cancer: The Role of Dysfunctional Adipose TissueObesity and Cancer. *Cancer Epidemiol Biomarkers Prev* **18**:2569–2578.

Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, Diegel M, Dunn D, Neri F, Haugen E, Rynes E, Reynolds A, Nelson J, Johnson A, Frerker M, Buckley M, Kaul R, Meuleman W, Stamatoyannopoulos JA. 2020. Global reference mapping of human transcription factor footprints. *Nature* **583**:729–736.

Vockley CM, D'Ippolito AM, McDowell IC, Majoros WH, Safi A, Song L, Crawford GE, Reddy TE. 2016. Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. *Cell* **166**:1269–1281.e19.

Wang N-D, Finegold MJ, Bradley A, Ou CN, Abdelsayed SV, Wilde MD, Taylor LR, Wilson DR, Darlington GJ. 1995. Impaired Energy Homeostasis in C/EBPα Knockout Mice. *Science* **269**:1108–1112.

Wang X, Wang H-Y, Hu G-S, Tang W-S, Weng L, Zhang Y, Guo H, Yao S-S, Liu S-Y, Zhang G-L, Han Y, Liu M, Zhang X-D, Cen X, Shen H-F, Xiao N, Liu C-Q, Wang H-R, Huang J, Liu W, Li P, Zhao T-J. 2021. DDB1 binds histone reader BRWD3 to activate the transcriptional cascade in adipogenesis and promote onset of obesity. *Cell Rep* **35**:109281.

Wang Z, Chu T, Choate LA, Danko CG. 2019. Identification of regulatory elements from nascent transcription using dREG. *Genome Res* **29**:293–303.

Westwood JT, Clos J, Wu C. 1991. Stress-induced oligomerization and chromosomal relocalization of heat-shock factor. *Nature* **353**:822–827.

Wingender E, Schoeps T, Haubrock M, Krull M, Dönitz J. 2018. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res* **46**:D343–D347.

Wissink EM, Vihervaara A, Tippens ND, Lis JT. 2019. Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet* **20**:705–723.

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.

Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA. 2004. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**:927–939.

Yeh WC, Cao Z, Classon M, McKnight SL. 1995. Cascade regulation of terminal adipocyte differentiation by three members of the C/EBP family of leucine zipper proteins. *Genes Dev* **9**:168–181.

Yousfi M, Lasmoles F, Lomri A, Delannoy P, Marie PJ. 2001. Increased bone formation and decreased osteocalcin expression induced by reduced Twist dosage in Saethre-Chotzen syndrome. *J Clin Invest* **107**:1153–1161.

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**:2227–2241.

Zeitlinger J, Stark A, Kellis M, Hong J-W, Nechaev S, Adelman K, Levine M, Young RA. 2007. RNA

polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* **39**:1512–1516.

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

Fig. 1 - CEBP, TWIST, SP, KLF, GR, and AP-1 TF families drive either increased or decreased chromatin accessibility in adipogenesis: A) Preadipocyte fibroblast 3T3-L1 cells were treated with an adipogenesis cocktail and harvested at the indicated time points for ATAC-seq and PRO-seq experiments. B) Temporal classification of ATAC-seq peaks revealed five major dynamic classes. Each dynamic ATAC peak is a red or blue trace with the number of peaks in the class indicated in the lower right; the x-axis represents time and the y-axis indicates normalized accessibility. C) *De novo* motif analysis identified the top six DNA motifs enriched within dynamic peaks. The individual TFs listed in the wedge below the DNA sequence logo recognize the respective DNA motifs. The heatmap quantifies the local protein sequence alignment of the DNA binding-domains for the genes, as determined by the Smith-Waterman algorithm (Farrar, 2006). D) Dynamic ATAC-seq peaks are classified by the presence of each DNA motif. The red bars represent the number of dynamic ATAC-seq peaks within the Immediate Increase, Transient Increase, and Gradual Increase categories; the blue bars correspond to the Transient Decrease and Gradual Decrease classes. E) Red, blue, and grey traces are composite motif densities relative to ATAC peak summits for the increased, decreased, and nondynamic peak classes. The y-axis quantifies the density of the indicated position-specific weight matrix and each motif instance is weighted by its conformity to a composite motif. F) Dynamic traces of peaks that exclusively contain the specified motif indicate that CEBP, GR, and AP-1 associate with increasing accessibility; SP and TWIST associate with decreasing accessibility. Peak traces are colored as in panel (B). These conclusions are consistent with the reciprocal analysis from panel (E).

Fig. 2 - SP, NRF, and E2F6 TF families drive bidirectional transcription dynamics at regulatory regions within gene bodies and promoters: A) The heatmap illustrates over 200,000 putative REs with a bidirectional transcription signature. B) Both dREG and ATAC-seq identify a RE within the promoter of *Cops8*. The intragenic RE is only identified by its bidirectional PRO-seq signature while the upstream intergenic RE is only identified by ATAC-seq. C) Dynamic ATAC-seq and dREG-defined REs largely overlap in promoter regions. Intragenic regions are defined based on primary transcript annotation of PRO-seq data, promoters are between 150 bp upstream and 50 bp downstrean of TSSs, and intergenic regions are the remainder of the genome. D) Dynamic ATAC-seq peaks are enriched for a more diverse set of TF motifs than dynamic dREG peaks. E) Motif density distinguishes TFs associated with dynamic bidirectional transcription from those associated with dynamic accessibility. For example, TWIST and GR motifs are enriched within dynamic ATAC-seq peaks but are rarely found within dynamic dREG peaks. F) SP is only associated with bidirectional transcription at promoters and not distal REs. The top plot shows the average normalized PRO-seq signal for plus and minus strands around all 1,135,731 SP motif instances while the bottom plot displays all SP motifs excluding those in promoters (1,118,185). The distinct dual peak profile of bidirectional transcription collapses when only considering SP motifs outside promoters. G) Dynamic bidirectional transcription peaks found in promoters are stratified by the presence or absence of TF motifs. The left plot quantifies the total number of peaks and the right plot scales to the proportion of peaks in each category. The x-axis factor motif categories are defined by the presence or absence of ATAC-associated factors (AP-1, CEBP, GR, KLF, and TWIST) and dREG-associated factors (SP, E2F6, and NRF). dREG-associated factor motifs are enriched in peaks that decrease bidirectional transcription, suggesting a link between SP, NRF, and E2F6 factors and an early and pervasive decrease in promoter initiation at their target genes.

Fig. 3 - Chromatin accessibility, transcription dynamics, and proximity guide inference of *cis*-edges between REs and target genes: A) Change in gene expression correlates with local accessibility change over the first hour. Each data point represents a gene within 10 kb of either only increased (red) or decreased (blue) peaks. The y-axis indicates the number of increased or decreased accessibility peaks near the gene and the x-axis represents the normalized change in gene transcription over the first hour. B) *Klf5* (left) is part of a cluster of 1717 immediately and transiently activated genes (gray traces on the right). C) Cumulative distribution plots showing distance between GR-bound REs and genes either activated (red), repressed (blue), or unchanged (grey) over the first hour of the time course. The left-shift of the red curve suggests that ATAC-seq peaks with GR motifs are closer to the 20 vs. 0 min activated gene class. The inset plot reports the difference in cumulative distribution between the activated and unchanged gene classes as distance from the TSS increases. The leveling off of the traces at 114 kb from the TSSs suggests that GR-mediated transcription activation requires GR to bind within 100 kb of the TSS. D) *Stat1* (left) is part of a cluster of repressed genes (gray traces on the right). E) ATAC-seq peaks with SP motifs are closer to the 60 vs. 120 min repressed gene class. Traces converge 1,900 bases from the start sites, suggesting that the functional distance of SP-mediated gene repression is within 2 kb of TSSs.

Fig. 4 - Constrained networks downstream of AP-1, GR, and SP identify genes regulated by individual factors: Simplified networks highlight the number of REs and genes that are combinatorially or individually regulated by (A) AP-1, (B) GR, and (C) SP. Factors bind / dissociate from REs (purple circles) and regulate genes (blue squares). Colored arrows and numbers indicate the contribution of non-lead factors to RE activity. Combinatorially regulated REs are bound by the lead TF and either one or more of the other TFs. Composite PRO-seq signal is plotted relative to the promoter-proximal pause peak of (D) 1224 genes solely regulated by AP-1, (E) 174 genes regulated by GR, and (F) 1127 genes regulated by SP. Inset violin plots illustrate the change in pause index for the gene set for the indicated time points. Each data point is a gene and all genes were input from the composite.

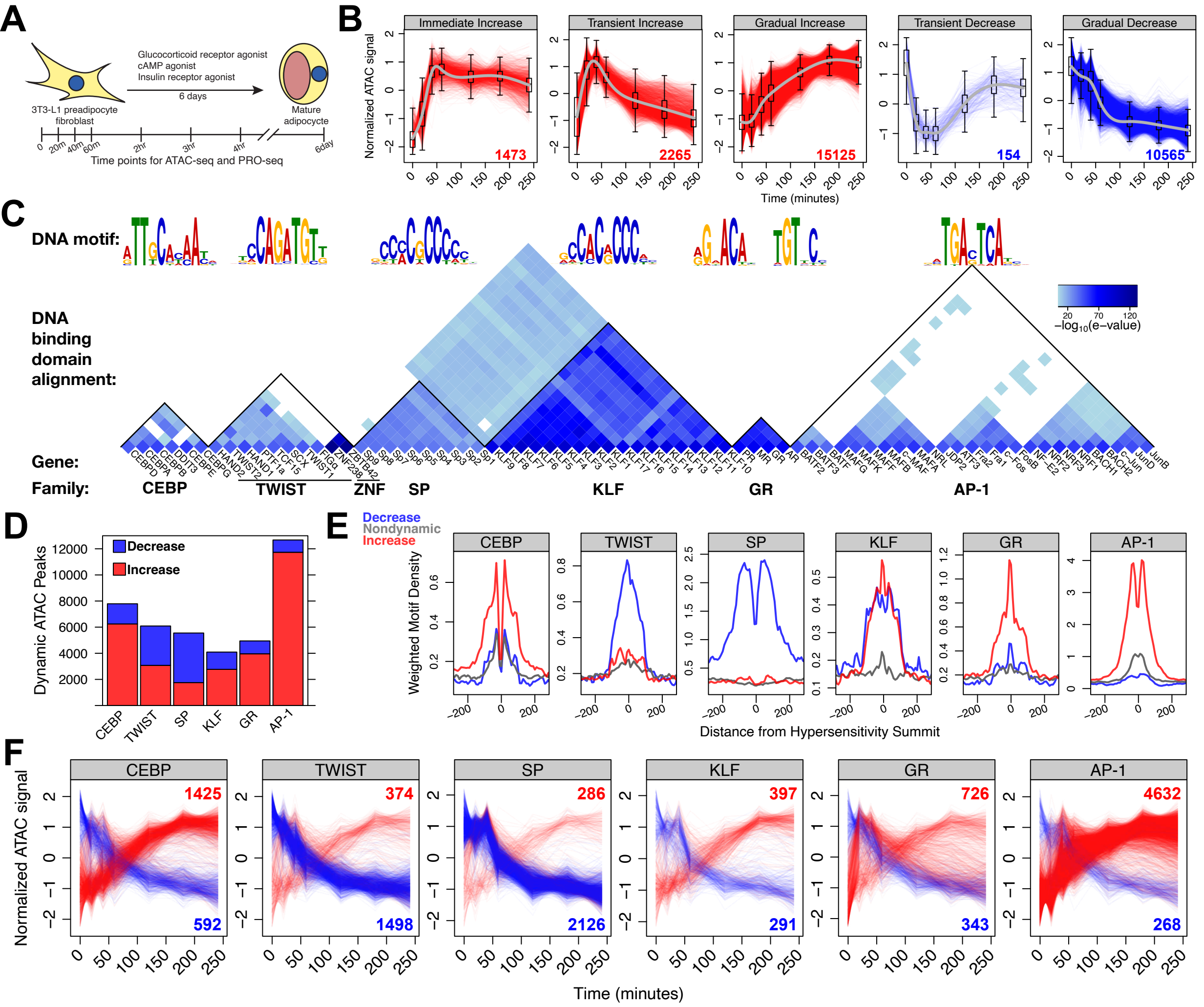
Fig. 5 - TFs induce changes to transcriptional rate constants: A) The compartment model of transcription contains rate constants for RNA Polymerase II (PolII) initiation / recruitment, premature termination, pause release, and elongation. Polymerase occupancy in the pause or gene body compartment is measured directly from PRO-seq. Differential equations relate rate constants to the rate of change of polymerase density in the compartments. B) We use the compartment model to estimate how rate constants change between 0 and 20 minutes for the 174 genes regulated solely by GR (left), 60 and 120 minutes for the 1127 genes regulated by SP (middle), and 40 and 60 minutes for the 1224 genes regulated by AP-1 (right). For each estimation, we hold k_{pre} and k_{elong} constant and calculate values of k_{init} and k_{rel} that fit the observed occupancy in the pause and body regions. C-D) We simulated composite profiles for a set of parameters from (B left) and (B middle).

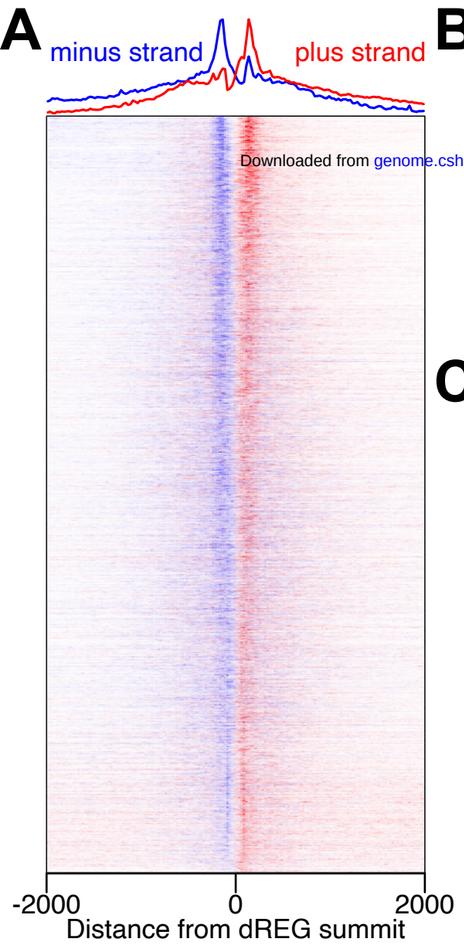
Fig. 6 - TFs bind REs and activate genes individually and combinatorially: A) Schematics and modular networks illustrate 4 variations in AP-1, CEBP, GR, and KLF binding patterns that can lead to activation of gene targets: all 4 factors bind to a single RE (orange), 2 REs each bound by 2 factors (purple), 3 factors bind one RE and 1 factor binds another (blue), and a more complex combination with redundant factor contributions at multiple REs (green). B) A wide network depicts cooperation between TFs. Top squares indicate GR, AP-1, CEBP, and KLF TF families. Second row circles represent 15 classes of REs each bound by a different combinations of factors. Third row squares represent 15 classes of genes each regulated by a different combination of factors. There are multiple potential combinations of REs that can produce the same gene class, as illustrated in (A). The middle square representing the 82 genes activated by all 4 factors is an example of variable regulatory combinations. Colored arrows and numbers correspond to RE combinations as depicted in (A). C) Genes were sorted into categories based on proximity to AP-1 motifs, proximity to GR motifs, and activation status. We calculated 95% confidence intervals for odds ratios based on contingency tables consisting of genes stratified by their proximity to inferred AP-1 and GR regulatory elements and by whether the genes were dynamic. The confidence interval for genes that are not proximal to inferred AP-1 regulatory elements is 0.97 to 1.49. The confidence interval for genes that are proximal to inferred AP-1 regulatory elements is 2.3 to 2.89. The increase in odds ratio prediction suggests that activated genes proximal to AP-1 are significantly more likely to be proximal to GR than non activated genes. D) The confidence interval for genes not proximal to GR is 0.96 to 1.27. The confidence interval for genes proximal to GR is 1.94 to 2.87. Again, the increase indicates that AP-1 and GR factors coordinate to activate transcription. E) The fraction of repressed genes proximal to TWIST increases regardless of the presence of SP with odds ratio confidence intervals of 2.2-3.07 and 2.21-2.79 when not proximal and proximal to SP peaks respectively. F) We find a lower proportion of repressed genes proximal to SP motifs both in the presence and absence of TWIST. The odds ratio confidence intervals do not change with the presence of TWIST, going from 0.71-0.89 to 0.64-0.9 when in proximity to predicted TWIST REs.

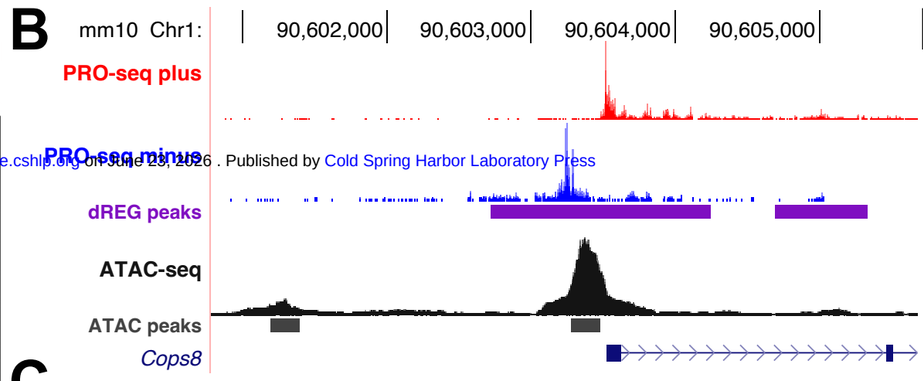
Fig. 7 - Variations in TF gene expression lead to downstream changes in accessibility and transcription: A) A deep network highlights temporal components of the regulatory cascade. Node and edge characteristics are illustrated as in Fig. 4. In addition, we added time interval attributes to respective edges to indicate when REs and genes are changing. B-D) UCSC Genome Browser shots indicate accessibility and peaks (top three tracks) as well as nascent transcription (next two tracks) dynamics between the indicated time points. E) Wedged bar plots quantify the regulatory kinetics across the time course for indicated factors. The x-axis intervals represent the time range in which the indicated number (y-axis) of genes are regulated (connected in the network) by the specified factor. Wedges between bars indicate carryover elements from previous time interval and the outer "wings" represent elements that are not included in the previous time interval. The top shaded blue wedges represent genes regulated by multiple factors; bottom wedges represent genes that are solely regulated by the indicated factor. F) A *Twist2*-centric network illustrates the high connectivity and influence of *Twist2*.

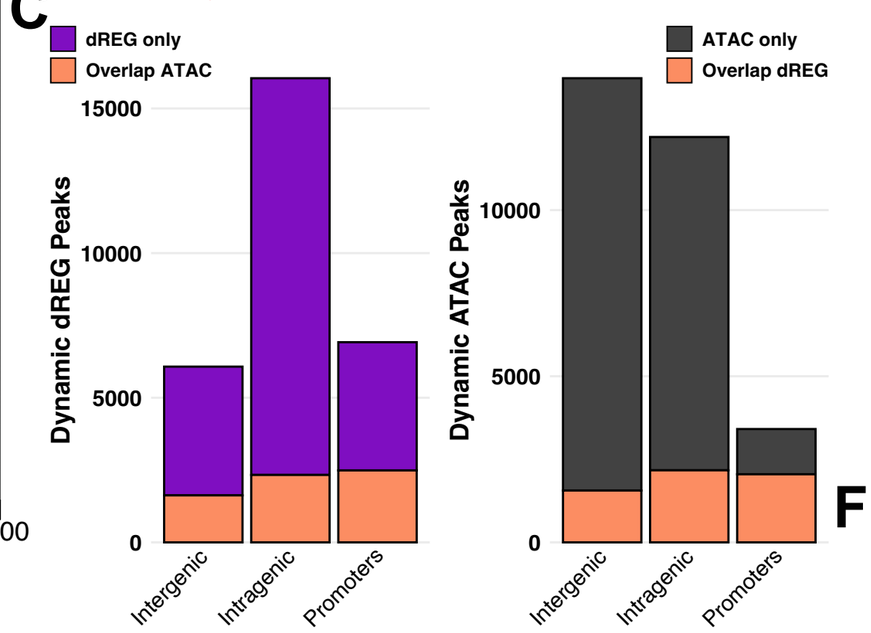
Fig. 8 - TWIST2 represses target genes and inhibits adipogenesis: A) *Twist2* expression as measured by RNA-seq at indicated time points following differentiation initiation. shTwist2.v1 is a single shRNA construct while shTwist2.v2 is a combination of shRNAs targeting *Twist2*. B) Expression of predicted TWIST2 target genes that are repressed in the RNA-seq time course before addition of the adipogenic cocktail. Y-values indicate fold change of gene expression in *Twist2* knockdown versus control. X-values indicate basal expression of each gene. The majority of genes are expressed to a higher degree in the knockdown samples, suggesting that loss of basal TWIST2 leads to derepression of target genes. C) Each data point represents one of the 520 predicted TWIST2 target genes that are repressed in the RNA-seq time course. We identified the time point comparison for which each gene exhibits the greatest degree of repression. X-values represent fold change for the relevant comparison in the shTwist2 samples and Y-values represent fold change for the same comparison in the shControl samples. Predicted TWIST2 target genes largely exhibit a greater degree of repression in the shTwist2 samples due to a relatively greater magnitude of *Twist2* activation. D) shControl or shTwist2 3T3-L1s were stained with Oil Red O after 6 days of differentiation. Absorbance at 540nm represents total lipid uptake of the cell population. Lines connect control and experimental samples from the same experimental replicate. Inset plot displays the difference in absorbance between the indicated shRNA treatment and the control knockdown for each replicate. TWIST2 depletion causes an increase in fat uptake, implicating the TF as a negative regulator of differentiation. E) 3T3-L1s overexpressing either LacZ or TWIST2 were stained with Oil Red O after 6 days of differentiation. TWIST2 overexpression causes a decrease in fat uptake, supporting the conclusion that TWIST2 is a negative regulator of adipogenesis. Lines and inset plot are used as in (D). F) Primary WAT preadipocytes harvested from P3 mice were induced to differentiate for 6 days and stained with Oil Red O. We pooled preadipocytes from animals with the same genotype. Wild type and heterozygous samples consist of preadipocytes pooled from 2 and 6 animals respectively. The pooled samples were plated in triplicate for the experiment. Preadipocytes harvested from *Twist2*$^{+/-}$ pups exhibit increased Oil Red O staining compared to those harvested from wild type mice. G) Hematoxylin and eosin staining of skin shows collapse of dermal (purple bars) and subcutaneous WAT (orange bar) in P3 *Twist2*$^{+/-}$ mice compared to wild type. Scale bars indicate 200 µm. H) Hematoxylin and eosin staining of interscapular brown fat shows

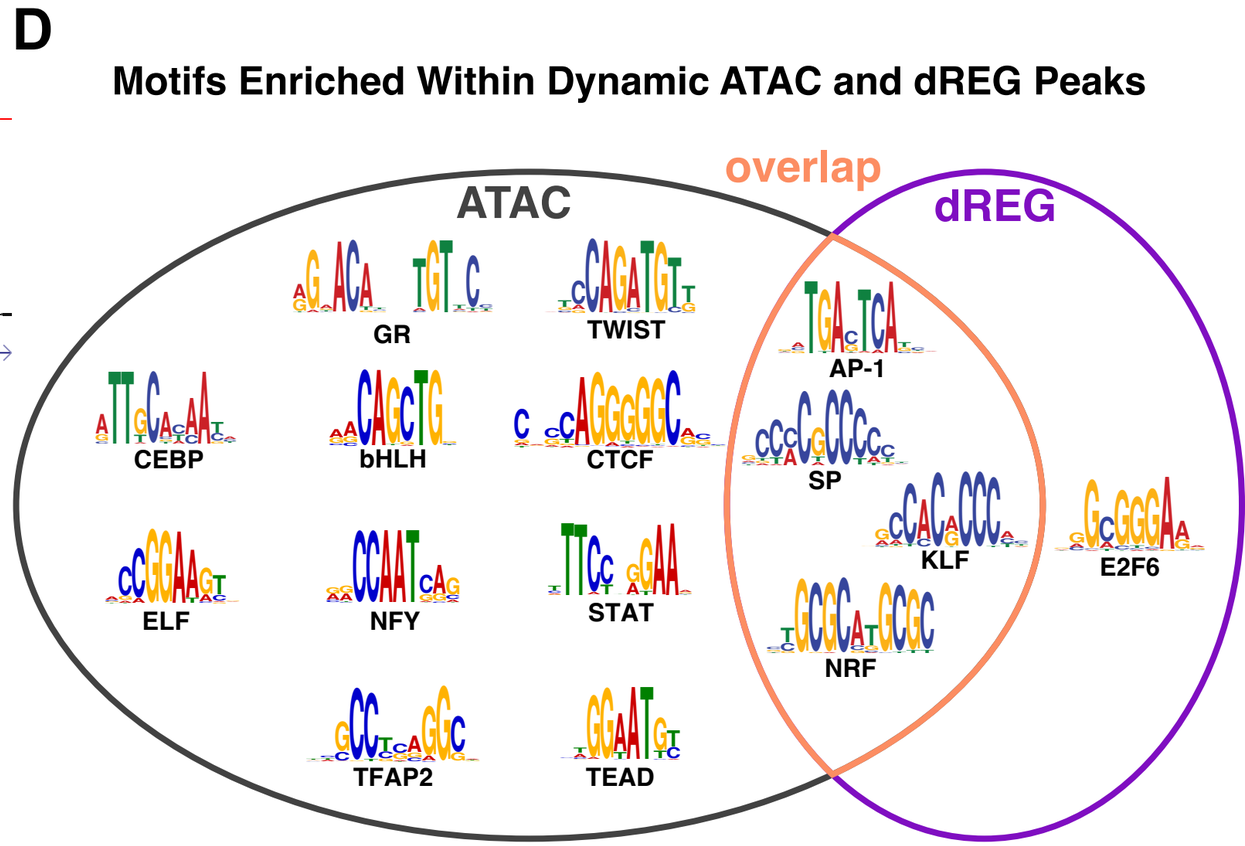
reduced fat droplets (large white/light colored circles) in P14 *Twist2*$^{-/-}$ mice compared to wild type. Images taken at either 40× (top) or 63× (bottom) magnification.

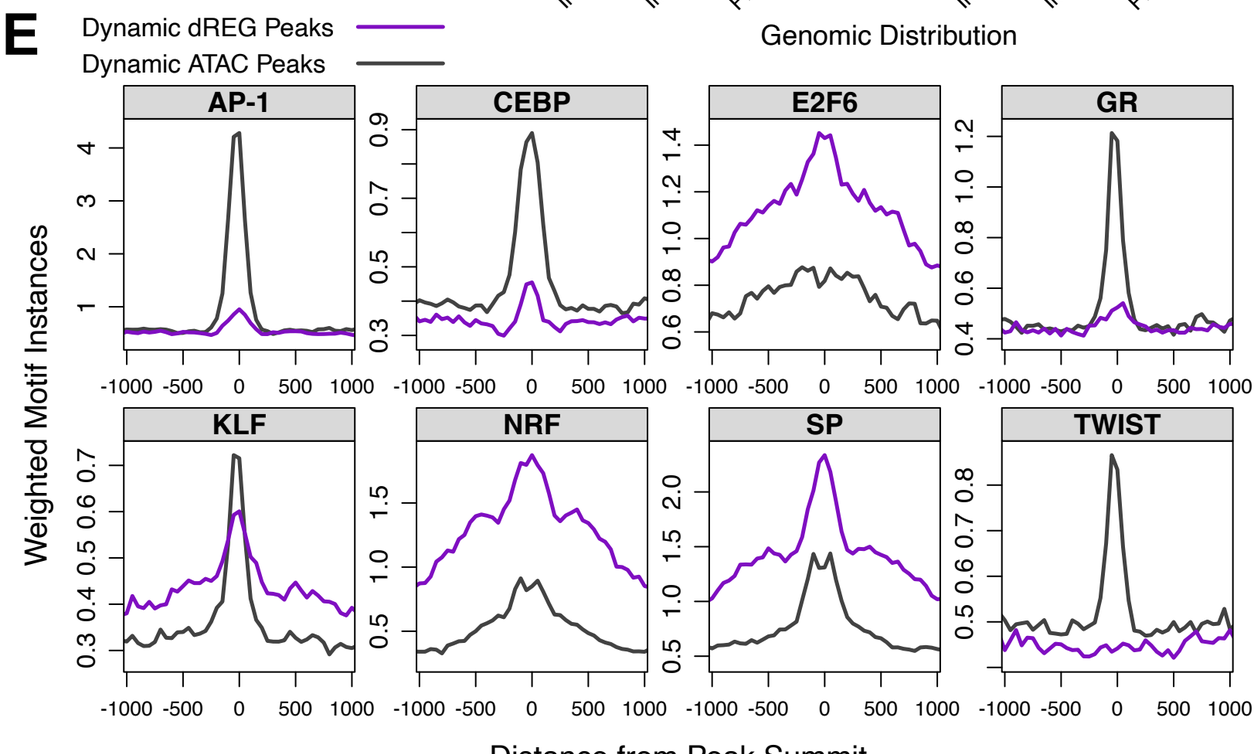

A
Glucocorticoid receptor agonist
cAMP agonist
Insulin receptor agonist
6 days
3T3-L1 preadipocyte fibroblast
Mature adipocyte
0
20m
40m
60m
2hr
3hr
4hr
6day
Time points for ATAC-seq and PRO-seq
B
Immediate Increase
Transient Increase
Gradual Increase
Transient Decrease
Gradual Decrease
Normalized ATAC signal
1473
2265
15125
154
10565
Time (minutes)
C
DNA motif:
DNA binding domain alignment:
Gene:
Family:
CEBP
TWIST
ZNF
SP
KLF
GR
AP-1
20
70
120
−log10(e−value)
D
Decrease
Increase
Dynamic ATAC Peaks
CEBP
TWIST
SP
KLF
GR
AP-1
E
Decrease
Nondynamic
Increase
Weighted Motif Density
Distance from Hypersensitivity Summit
F
Normalized ATAC signal
1425
592
374
1498
286
2126
397
291
726
343
4632
268
Time (minutes)

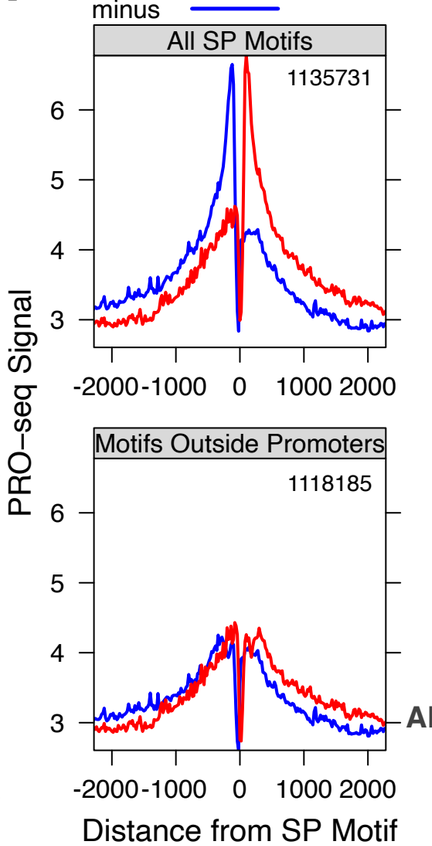


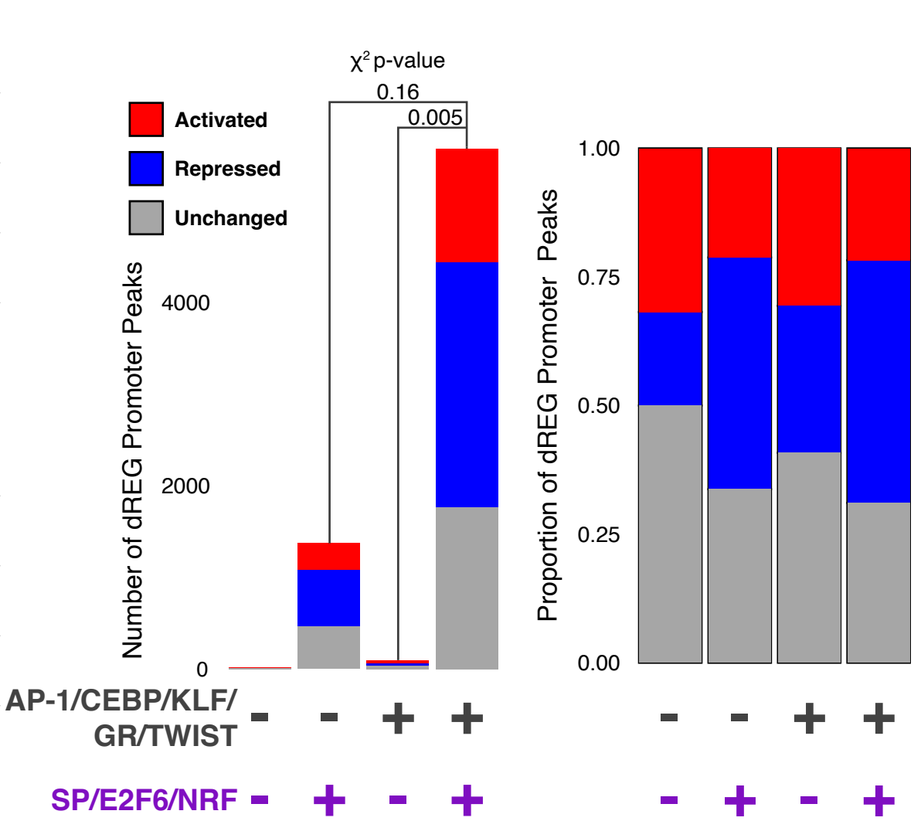


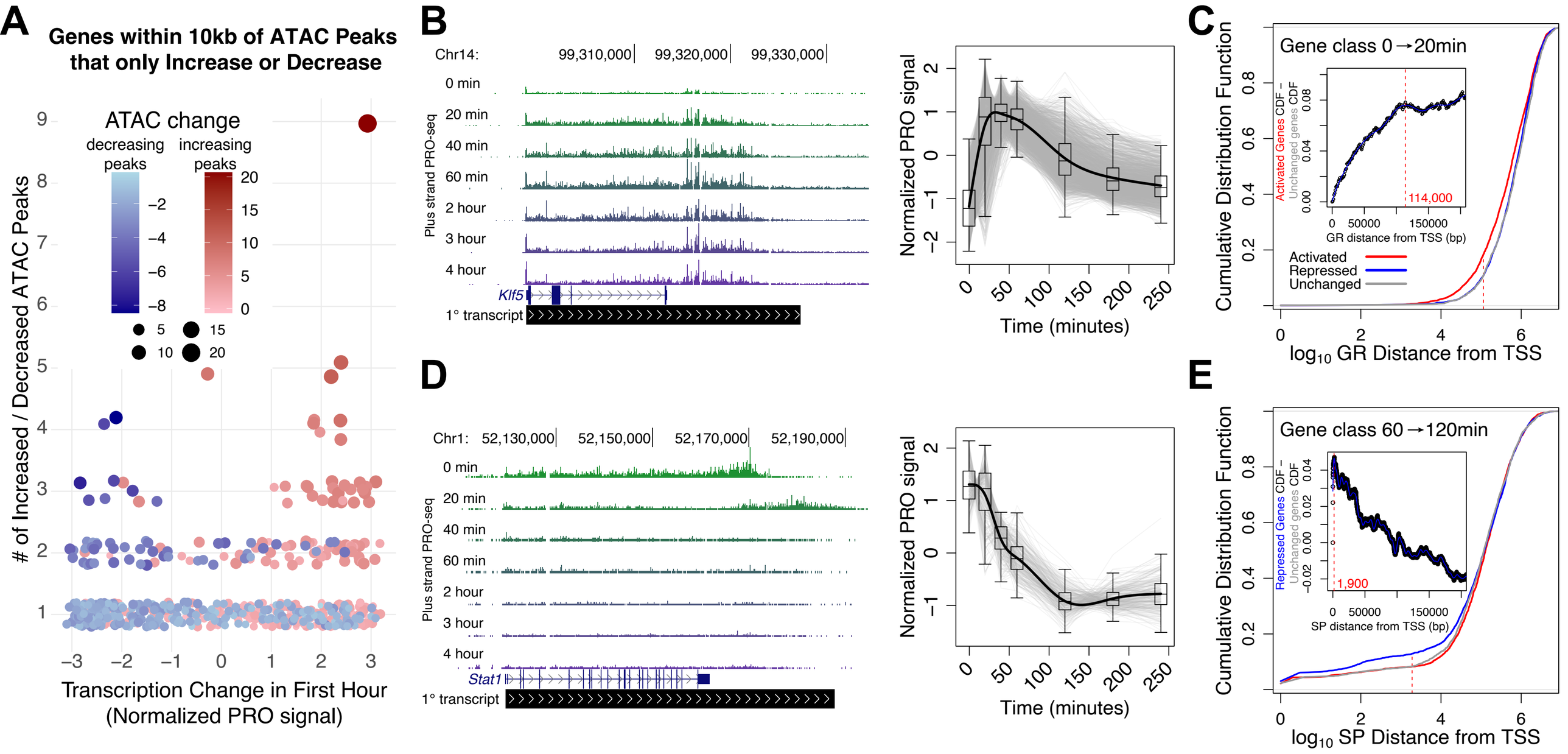




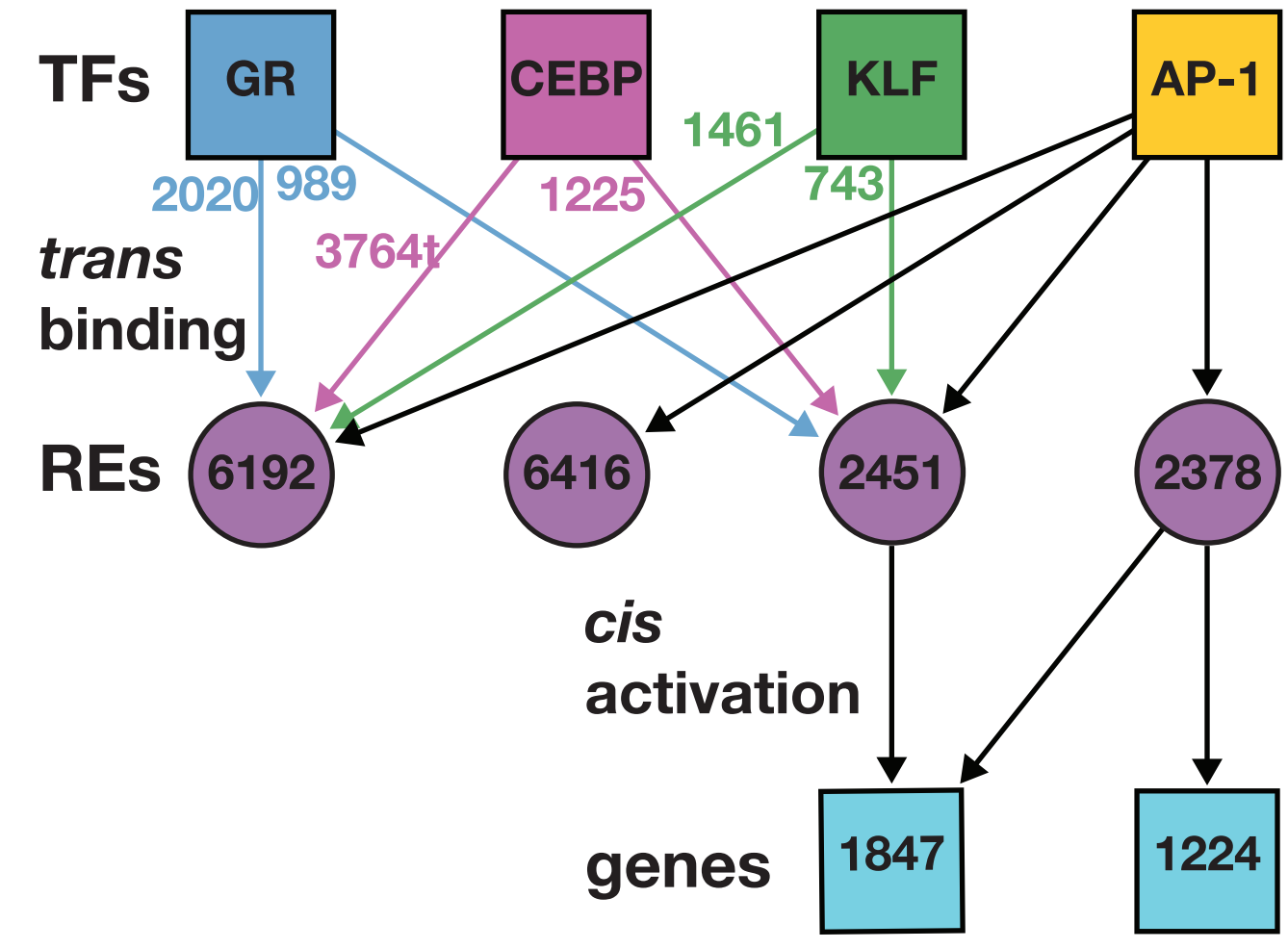




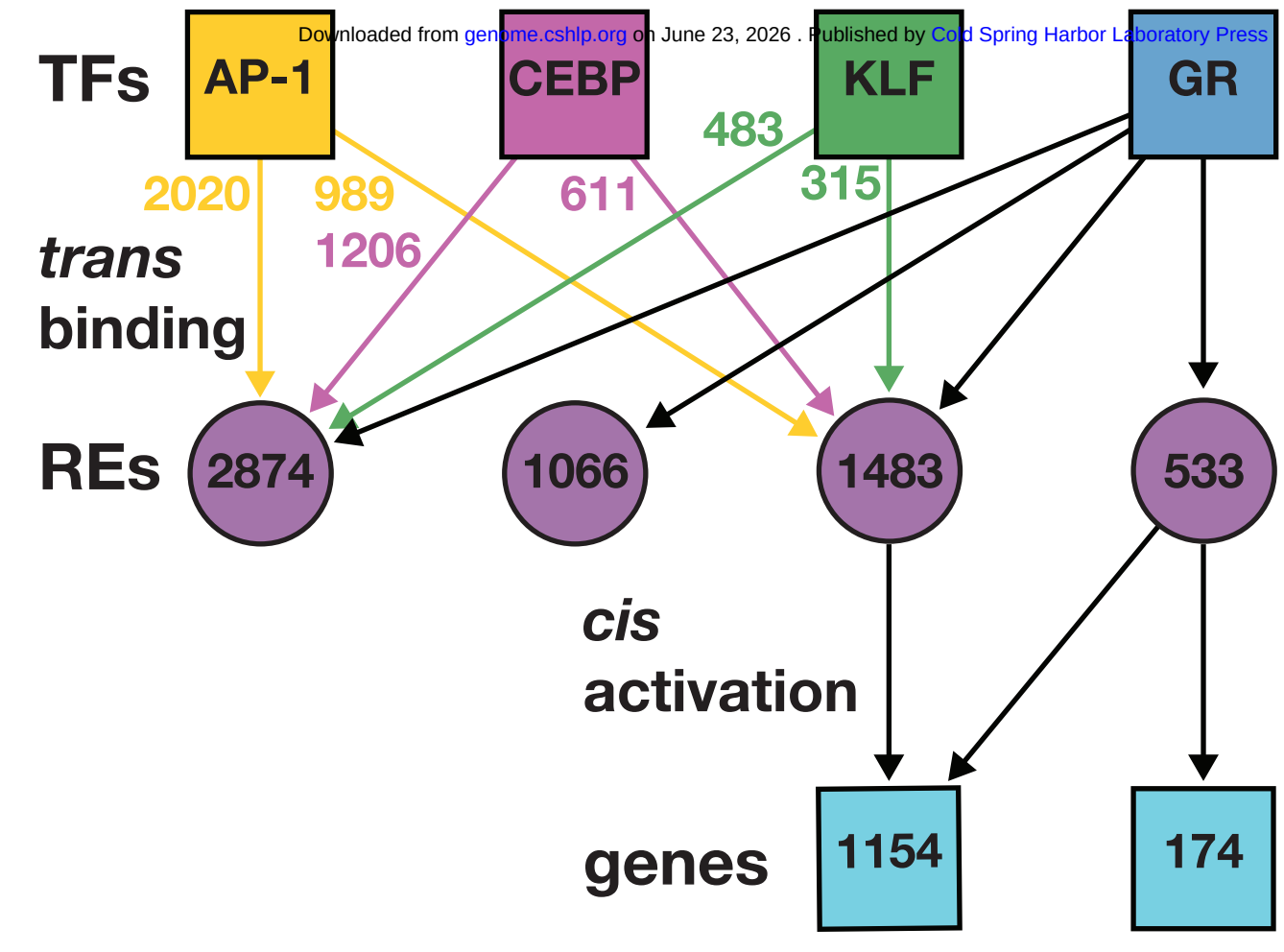



A
Genes within 10kb of ATAC Peaks that only Increase or Decrease
ATAC change
decreasing peaks
increasing peaks
# of Increased / Decreased ATAC Peaks
Transcription Change in First Hour (Normalized PRO signal)
B
Chr14:
99,310,000
99,320,000
99,330,000
0 min
20 min
40 min
60 min
2 hour
3 hour
4 hour
Plus strand PRO-seq
Klf5
1° transcript
Normalized PRO signal
Time (minutes)
C
Gene class 0→20min
Cumulative Distribution Function
Activated Genes CDF – Unchanged genes CDF
114,000
GR distance from TSS (bp)
Activated
Repressed
Unchanged
log10 GR Distance from TSS
D
Chr1:
52,130,000
52,150,000
52,170,000
52,190,000
Stat1
E
Gene class 60→120min
Repressed Genes CDF – Unchanged genes CDF
1,900
SP distance from TSS (bp)
log10 SP Distance from TSS

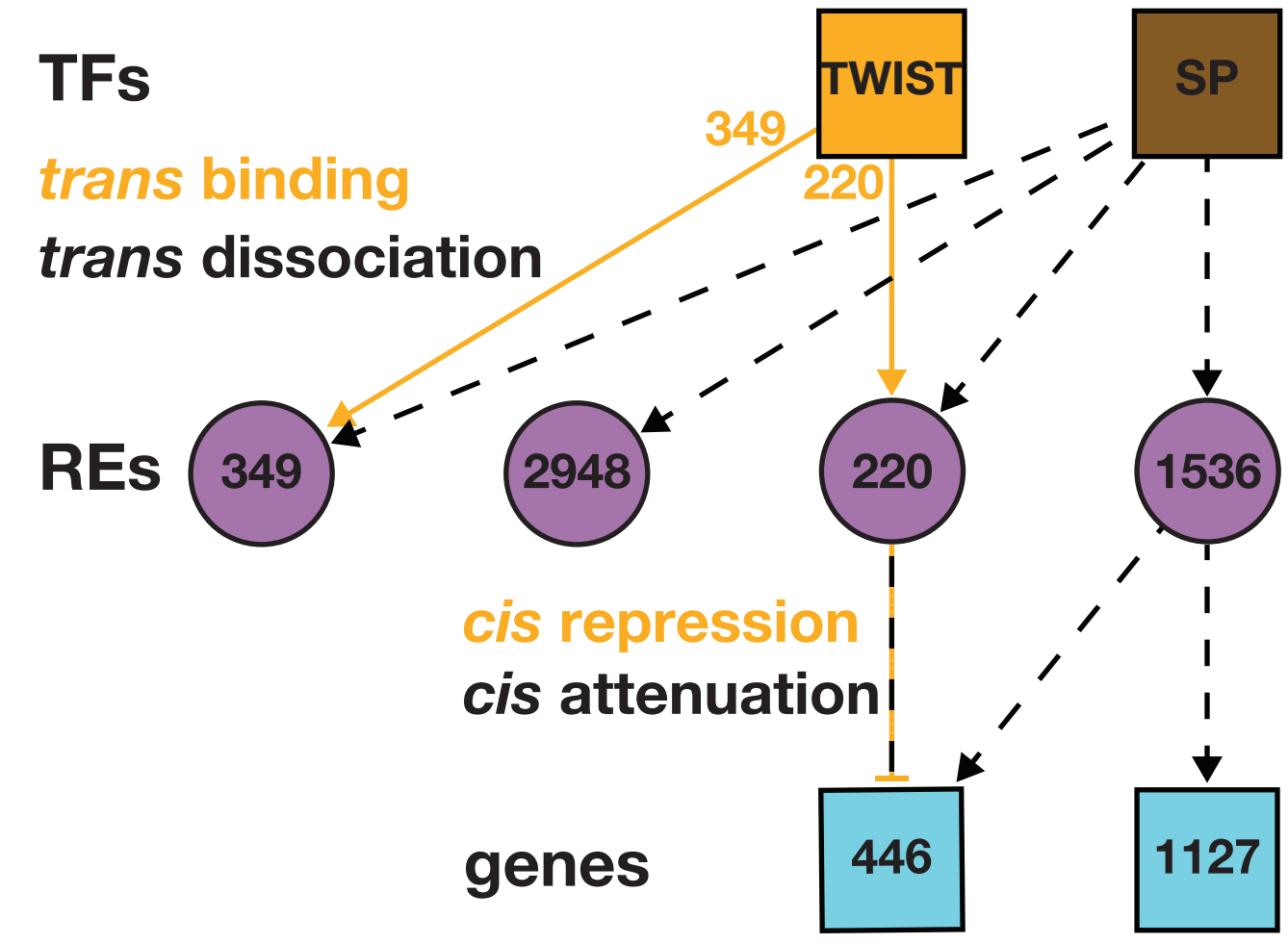
A



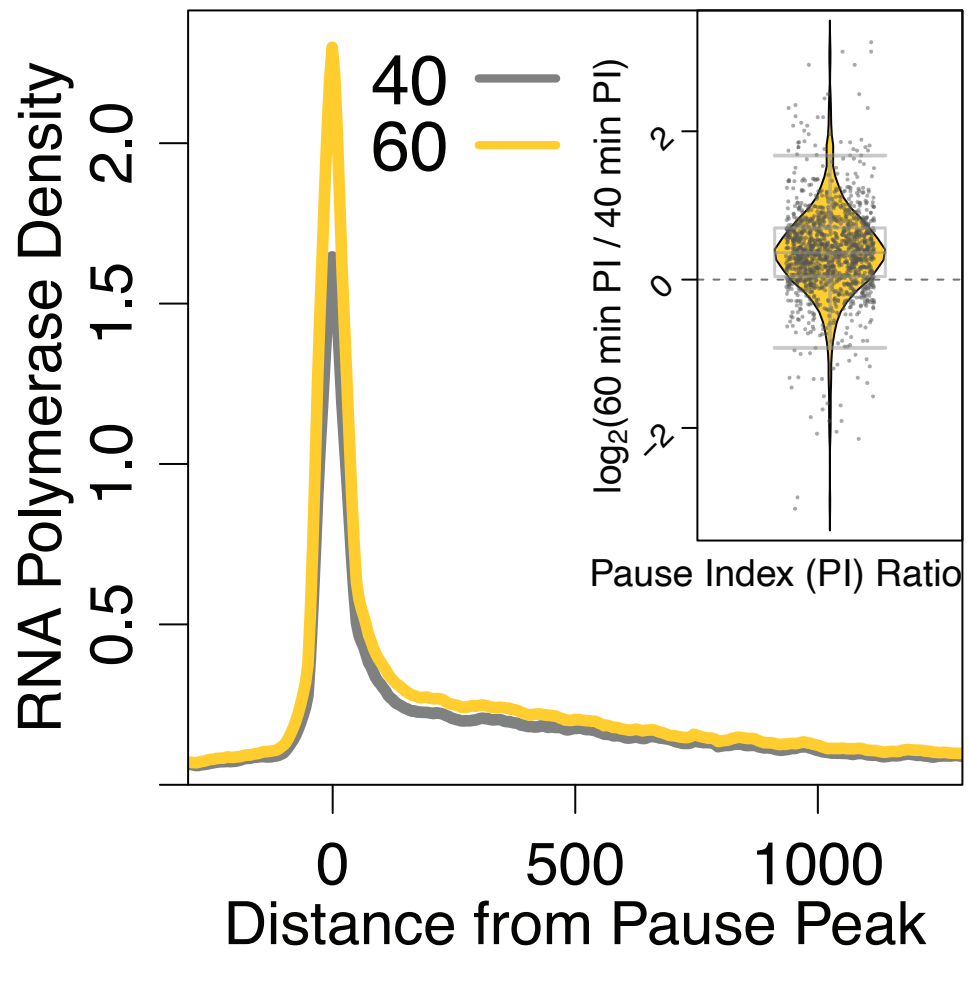
B



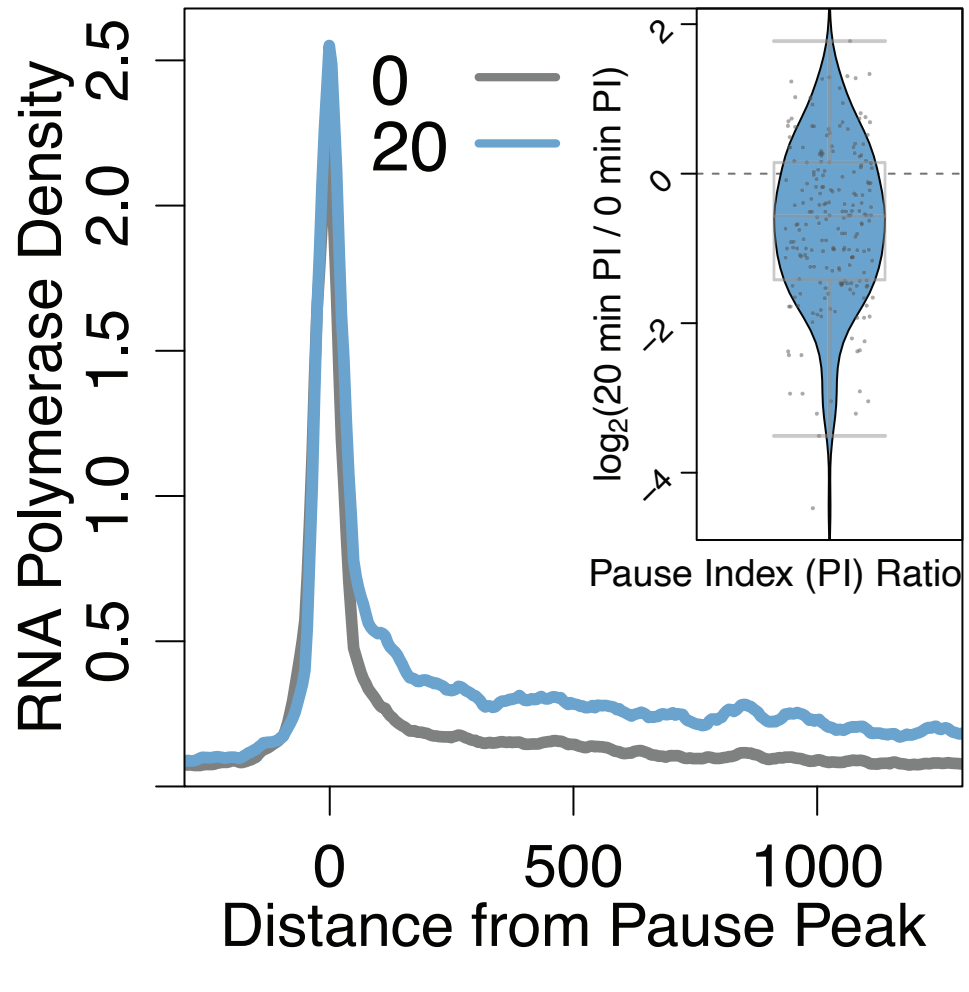
C



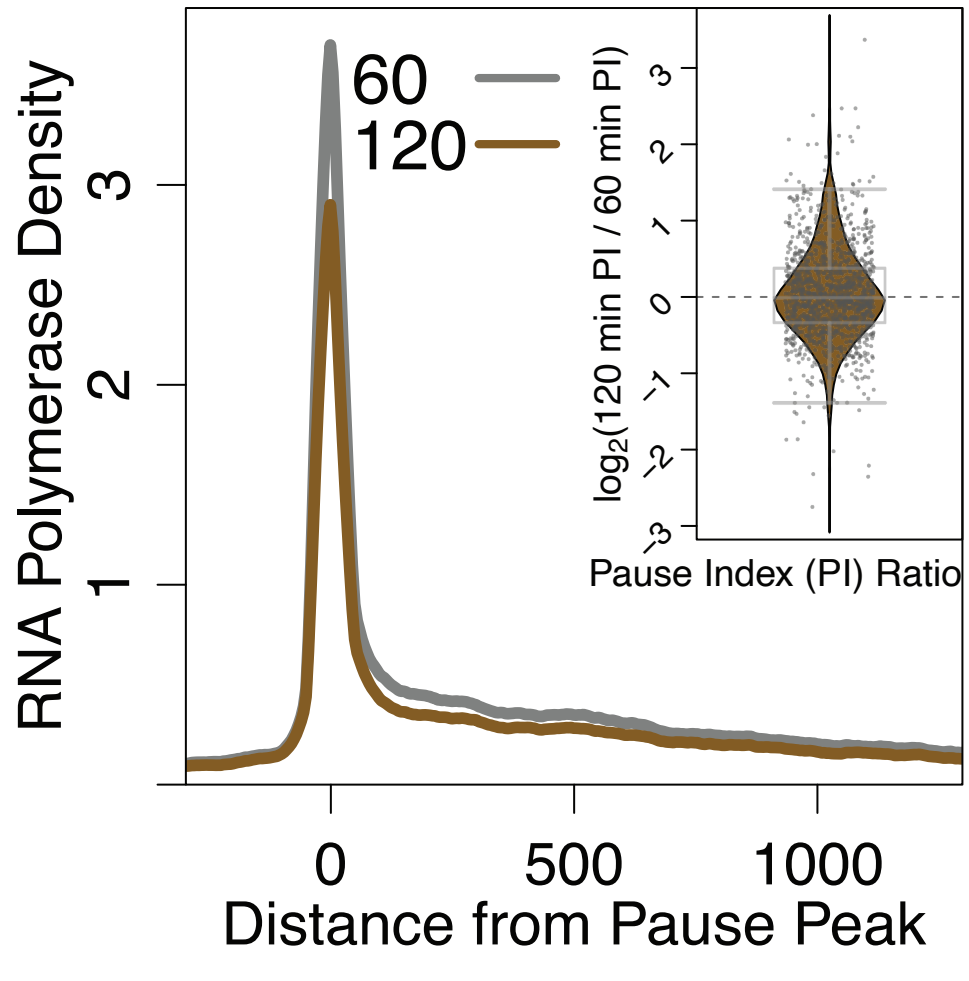
D

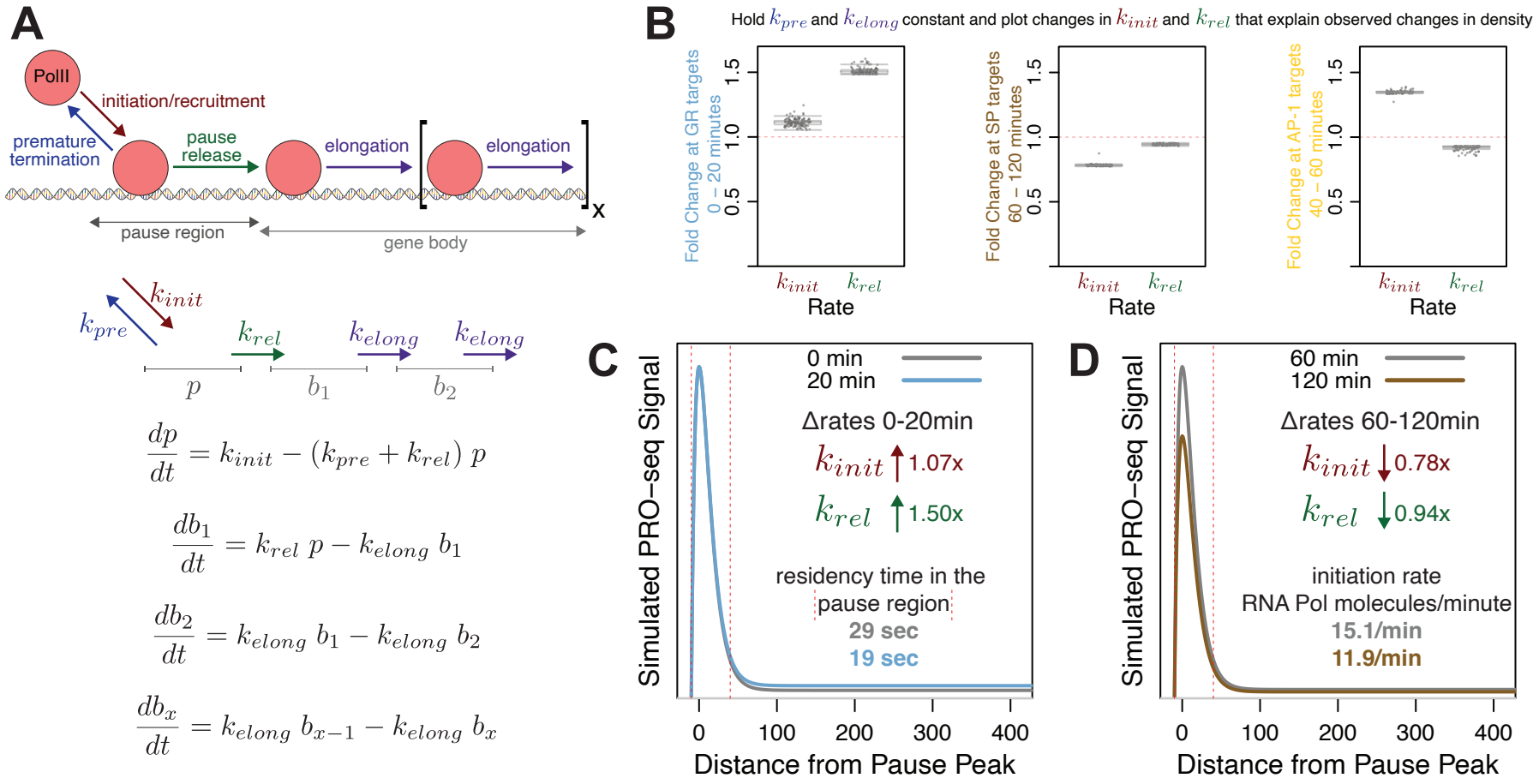


E

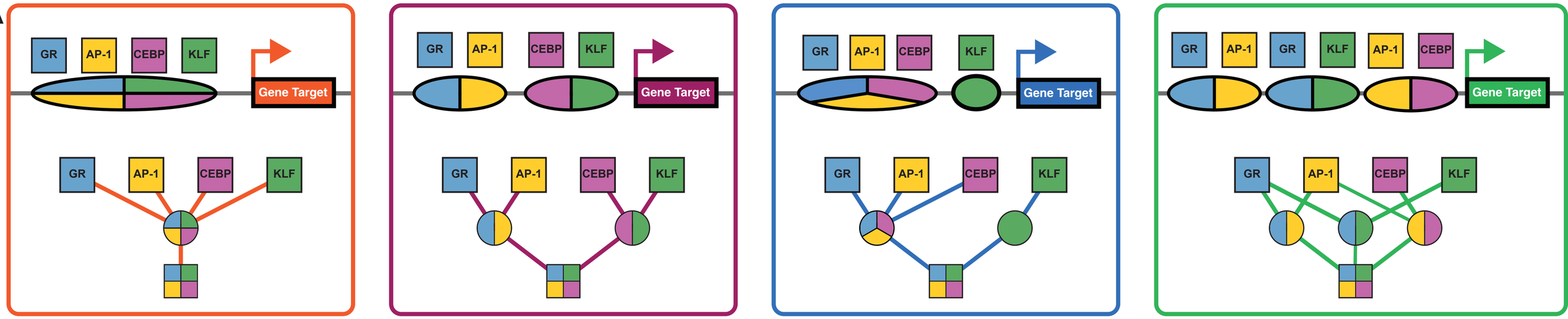


F

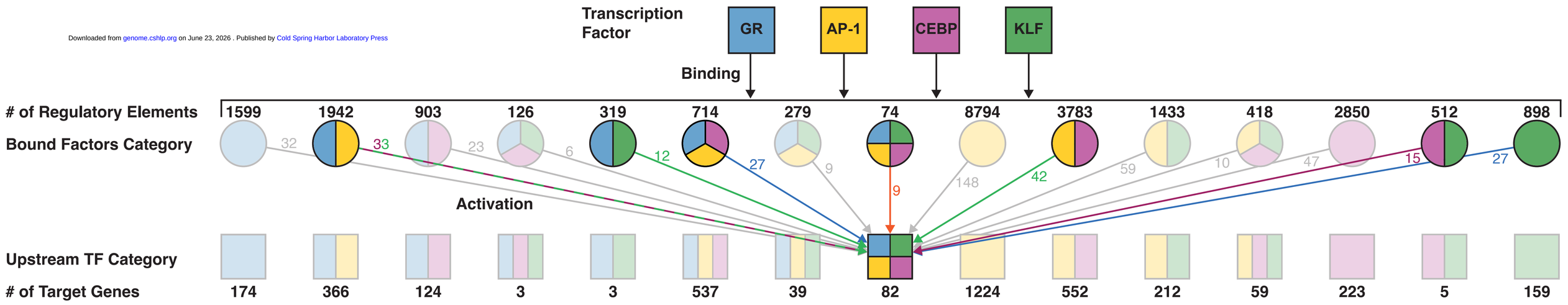



A
PolII
initiation/recruitment
premature termination
pause release
elongation
elongation
x
pause region
gene body
k_{init}
k_{pre}
k_{rel}
k_{elong}
k_{elong}
p
b_1
b_2
$\frac{dp}{dt} = k_{init} - (k_{pre} + k_{rel})\, p$
$\frac{db_1}{dt} = k_{rel}\, p - k_{elong}\, b_1$
$\frac{db_2}{dt} = k_{elong}\, b_1 - k_{elong}\, b_2$
$\frac{db_x}{dt} = k_{elong}\, b_{x-1} - k_{elong}\, b_x$
B
Hold k_{pre} and k_{elong} constant and plot changes in k_{init} and k_{rel} that explain observed changes in density
Fold Change at GR targets 0 – 20 minutes
Fold Change at SP targets 60 – 120 minutes
Fold Change at AP-1 targets 40 – 60 minutes
0.5
1.0
1.5
k_{init}
k_{rel}
Rate
C
Simulated PRO–seq Signal
0 min
20 min
Δrates 0-20min
k_{init} ↑1.07x
k_{rel} ↑1.50x
residency time in the pause region
29 sec
19 sec
0
100
200
300
400
Distance from Pause Peak
D
Simulated PRO–seq Signal
60 min
120 min
Δrates 60-120min
k_{init} ↓0.78x
k_{rel} ↓0.94x
initiation rate
RNA Pol molecules/minute
15.1/min
11.9/min
0
100
200
300
400
Distance from Pause Peak

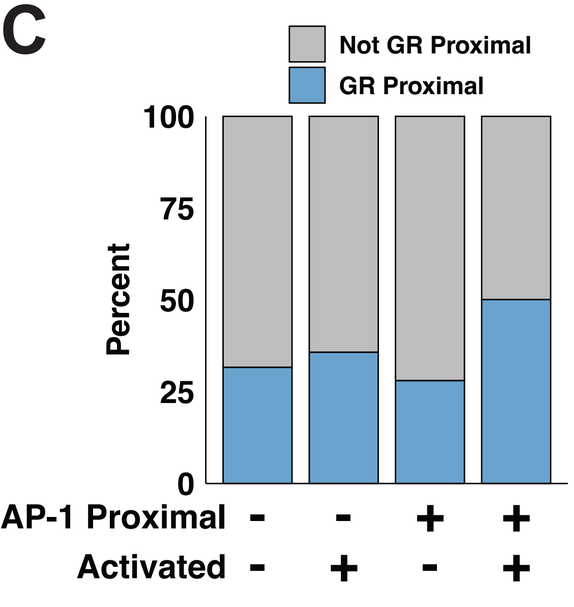
A



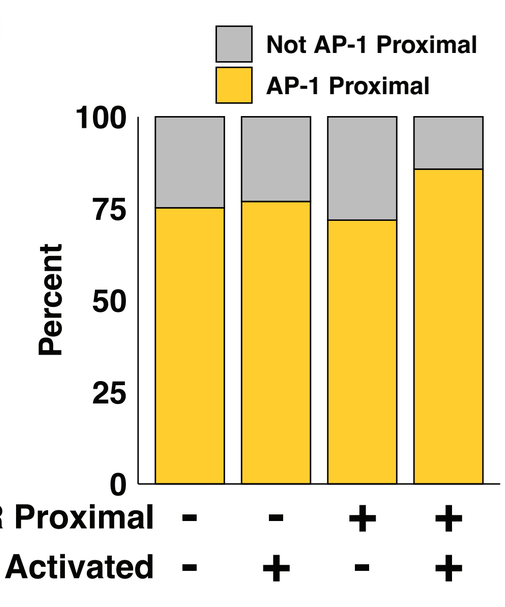
B



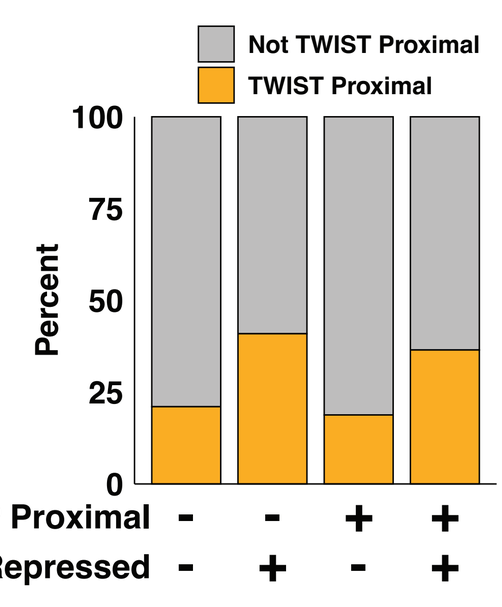
C



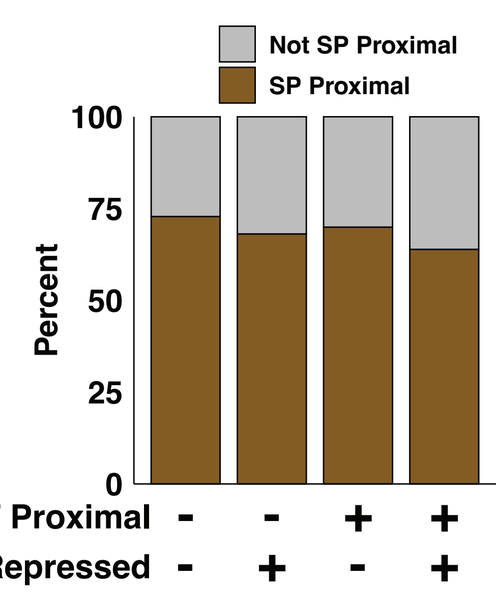
D

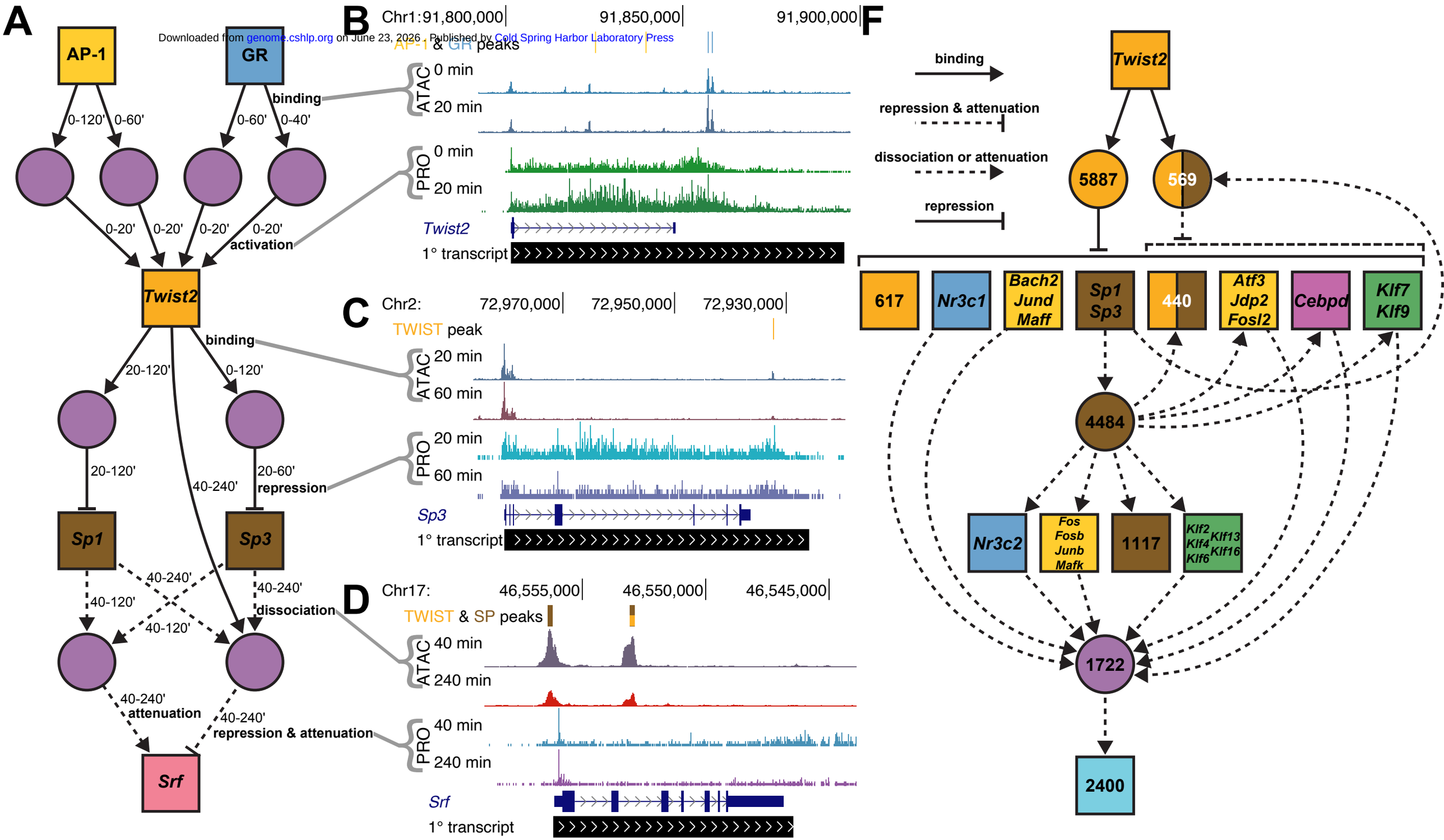


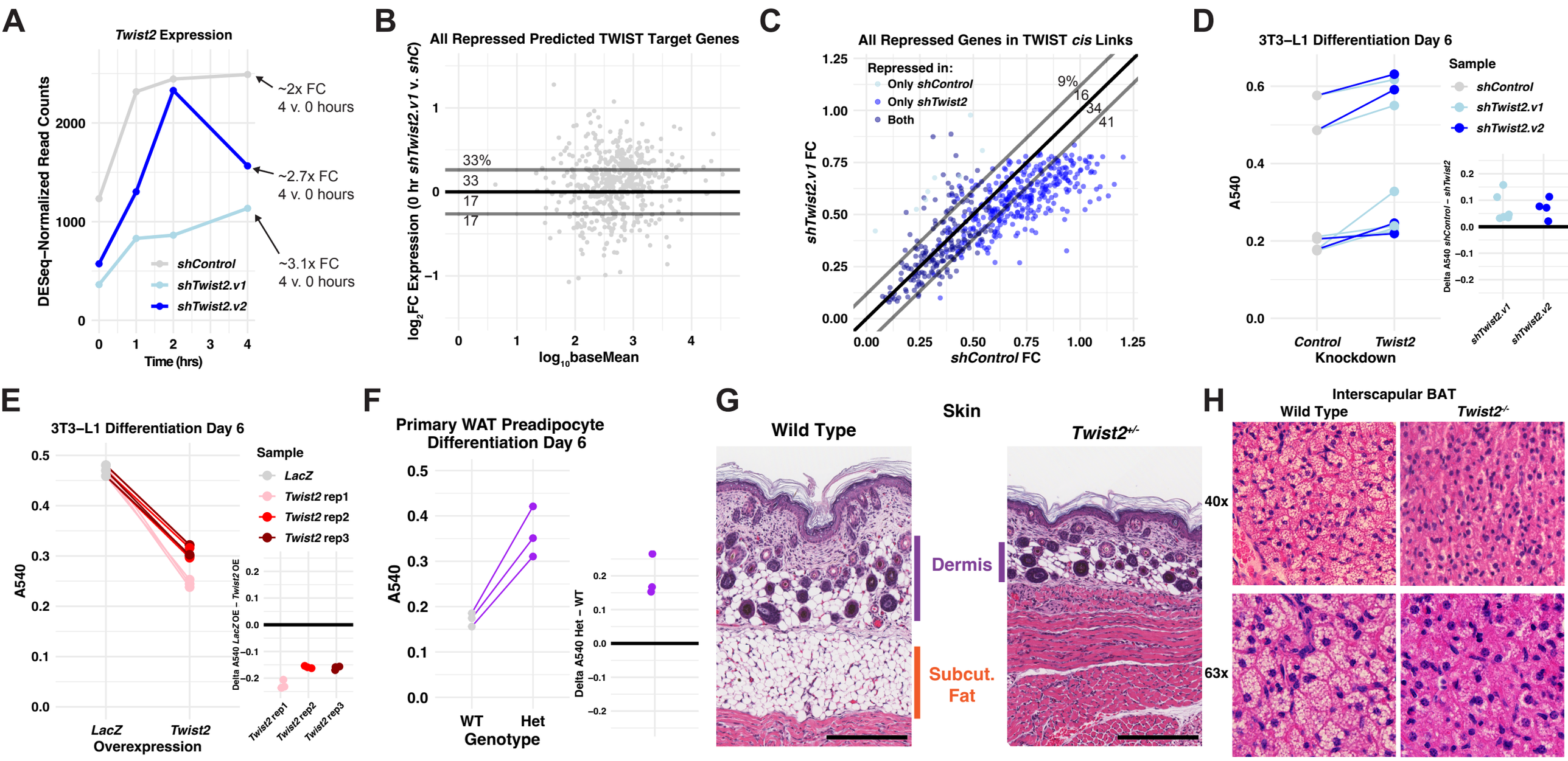
E



F






A
Twist2 Expression
DESeq–Normalized Read Counts
Time (hrs)
shControl
shTwist2.v1
shTwist2.v2
~2x FC
4 v. 0 hours
~2.7x FC
4 v. 0 hours
~3.1x FC
4 v. 0 hours
B
All Repressed Predicted TWIST Target Genes
log2FC Expression (0 hr shTwist2.v1 v. shC)
log10baseMean
33%
33
17
17
C
All Repressed Genes in TWIST cis Links
Repressed in:
Only shControl
Only shTwist2
Both
9%
16
34
41
shTwist2.v1 FC
shControl FC
D
3T3–L1 Differentiation Day 6
Sample
shControl
shTwist2.v1
shTwist2.v2
A540
Control
Twist2
Knockdown
Delta A540 shControl – shTwist2
E
3T3–L1 Differentiation Day 6
Sample
LacZ
Twist2 rep1
Twist2 rep2
Twist2 rep3
A540
LacZ
Twist2
Overexpression
Delta A540 LacZ OE – Twist2 OE
F
Primary WAT Preadipocyte Differentiation Day 6
A540
WT
Het
Genotype
Delta A540 Het – WT
G
Skin
Wild Type
Twist2+/-
Dermis
Subcut. Fat
H
Interscapular BAT
Wild Type
Twist2-/-
40x
63x