Mechanisms of Gene Regulation: Exploring the Activity of Transcription Factors and Mediator Kinases

by

J. D. Rubin

B.S., Cornell University, 2013

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

2020

This thesis entitled:
Mechanisms of Gene Regulation: Exploring the Activity of Transcription Factors and Mediator Kinases
written by J. D. Rubin
has been approved for the Department of Biochemistry

_____
Prof. Dylan J. Taatjes

_____
Prof. Robin D. Dowell

_____
Prof. James A. Goodrich

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Rubin, J. D. (Ph.D., Biochemistry)

Mechanisms of Gene Regulation: Exploring the Activity of Transcription Factors and Mediator Kinases

Thesis directed by Prof. Dylan J. Taatjes

The regulation of gene expression is an essential process required for proper cellular function. This process is especially important within multi-cellular organisms which contain a variety of cell types with specialized roles. Throughout this thesis, I will present work on development of computational methods to infer transcription factor (TF) activity and investigate the role of Mediator kinases in cellular signaling. The first half of this thesis involves work I performed to use signals of functional TF binding to infer TF activity using motifs. This initially involved detection of bidirectional transcripts from nascent sequencing data. I helped develop the first method to perform this quantification called the motif displacement (MD) score. I then improved upon the method developing transcription factor enrichment analysis (TFEA). I show that TFEA outperforms existing techniques and can be widely applied to different types of sequencing data. Next, I present work on how Mediator kinases affect the interferon response. I apply the aforementioned computational methods to show that CDK8 activity is responsible for activation of STAT and Irf family of TFs - the main effectors of the interferon pathway. Finally, I show that Mediator kinase activity is crucial for proper cellular response to serum. Using a variety of -omics techniques, I quantify the effects of Mediator kinase inhibition on immediate transcriptional changes, late gene expression changes, differential phosphorylation of key signaling proteins, changes in cellular metabolism, and defects in cellular proliferation. Overall, this thesis makes significant strides in our understanding of how cells regulate gene transcription/expression and the key players involved in cellular signaling.

Dedication

To my mother Ada Catherine Szeto. I miss you dearly.

Acknowledgements

Core Facilities: Cell Culture: Theresea, Nicole Sequencing Facility: Amber, Catelyn, Kevyn (?) IT: Jon Demasi, Matt Hynes-Grace, Dan Timmons Graduate Mentors: Zachary Poss - Rigor, Joey Azofeifa - Quantitative, Tim Read - Grab the bull by the horns Taatjes Lab, Dowell Lab Dylan, Robin Committee members

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

The regulation of gene expression is an essential cellular process required for life. In individual cells, proper and coordinated control over protein abundance allows a cell to maintain its homeostasis and respond to changes in its environment. In multi-cellular life, the ability to regulate sets of genes gives rise to a diversity of cell types with specialized functions. This can be accomplished while maintaining an identical genomic sequence for ease of propagation and development. Gene regulation is a highly complex process necessary for maintaining correct cellular functioning and when dysregulated can lead to disease. In this introduction, I will discuss our current understanding of how cells control gene expression. Essential to this process is the detection of extracellular signals resulting in their coordinated transmission to transcriptional machinery. The work described here focuses on two aspects of transcription 1) the regulatory activity of signal-specific transcription factors (TFs) and 2) the enzymatic activity of Mediator kinases, a component of the general transcription machinery. What follows is a temporally ordered description of key events involved in cellular detection of external stimuli and the coordinated intracellular response that ultimately leads to changes in gene transcription and protein abundance.

1.1 Cellular Signaling

Generally speaking, a cell receives chemical signals from its environment that it detects through cell surface receptors. In the classic case, these cell surface receptors are transmembrane proteins that have ligand binding domains facing towards the extracellular space and kinase do-

mains contained within the cytosol. Upon binding to a specific ligand, these receptors undergo conformational changes that lead to the activation of their kinase domain. What follows is a kinase signaling cascade in which upstream kinases phosphorylate downstream kinases which then themselves can phoshporylate other targets including other downstream kinases (See Figures 1.1 and 1.2 for examples). This allows for the exponential propagation of signal which is a highly coordinated event that ultimately leads to the expression of specific genes related to responding to the initial ligand. A couple of well-characterized pathways appear throughout this thesis and therefore I will briefly discuss what is known about these below.

1.1.1 Interferon Signaling

Interferons are a class of proteins that are responsible for the anti-viral response in cells. There are currently three types of known interferons (Type I, II, and III). Here I will discuss the signaling pathway associated with Type II interferons, more specifically the only known Type II interferon, IFN-γ. IFN-γ has been shown to play roles in both viral and bacterial infection (ref). It binds specifically to the IFN-γ receptor (IFNGR) which becomes phosphorylated and activated by JAK1/2. Activated IFNGR leads to the phosphoyrlation and activation of STAT1 which dimerizes and translocates into the nucleus to regulate expression of genes related to the inflammatory response. In addition to the inflammatory response, IFN-γ is known to also play roles in cell proliferation through the MAPK pathway and cytokine production through NF-$\kappa\beta$. A summary of the IFN-γ pathway is shown in Figure 1.1

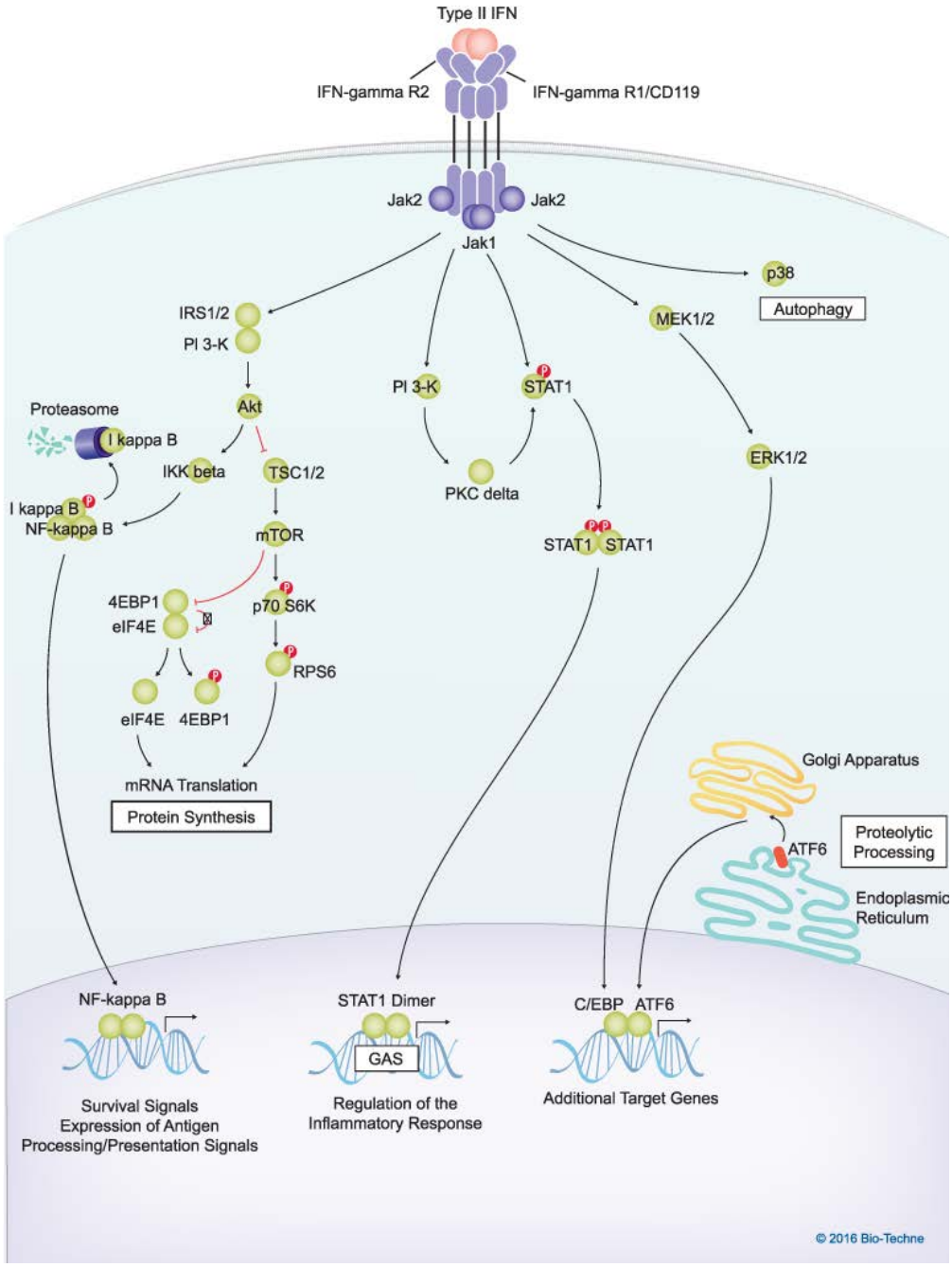


Figure 1.1: **The IFN-γ Pathway.** From https://www.rndsystems.com/pathways/type-ii-interferon-signaling-pathways

1.1.2 Response to Serum

Serum is a heterogeneous mix of nutrients and growth factors typically required for cell proliferation in cell culture models. The serum response is a model system used to study cell proliferation in which cells are first starved of serum followed by induction with serum-containing media. Since serum is a heterogeneous mix, typically derived from bovine fetuses, the serum response pathway is highly complex and not completely understood. Although the main response to serum is cellular proliferation, of particular interest in the studies described here is its interplay with infection pathways (ref). Studies have shown that the main downstream effector of the serum response (Serum Response Factor; SRF), can indirectly modulate Type I interferon response (Xie et al. 2013) and that some mitogens regulate cellular proliferation through the IFN-γ pathway (Ramana et al. 2001).

1.1.3 Cell Proliferative Signaling

Canonically, cell proliferative signaling refers to the mitogen-activated protein kinase (MAPK) pathway. This is a well-studied signaling pathway that ultimately results in cell proliferation. However, despite the simplicity of promoting cell proliferation as a functional outcome, the MAPK pathway is a complex network of protein interactions. This networks includes the activation of other cellular pathways that are necessary to properly regulate this proliferation to ensure an appropriate cellular response to mitogen signaling. The complexitiy of the MAPK pathway is evidenced by the known relationships between dozens of proteins shown in Figure 1.2

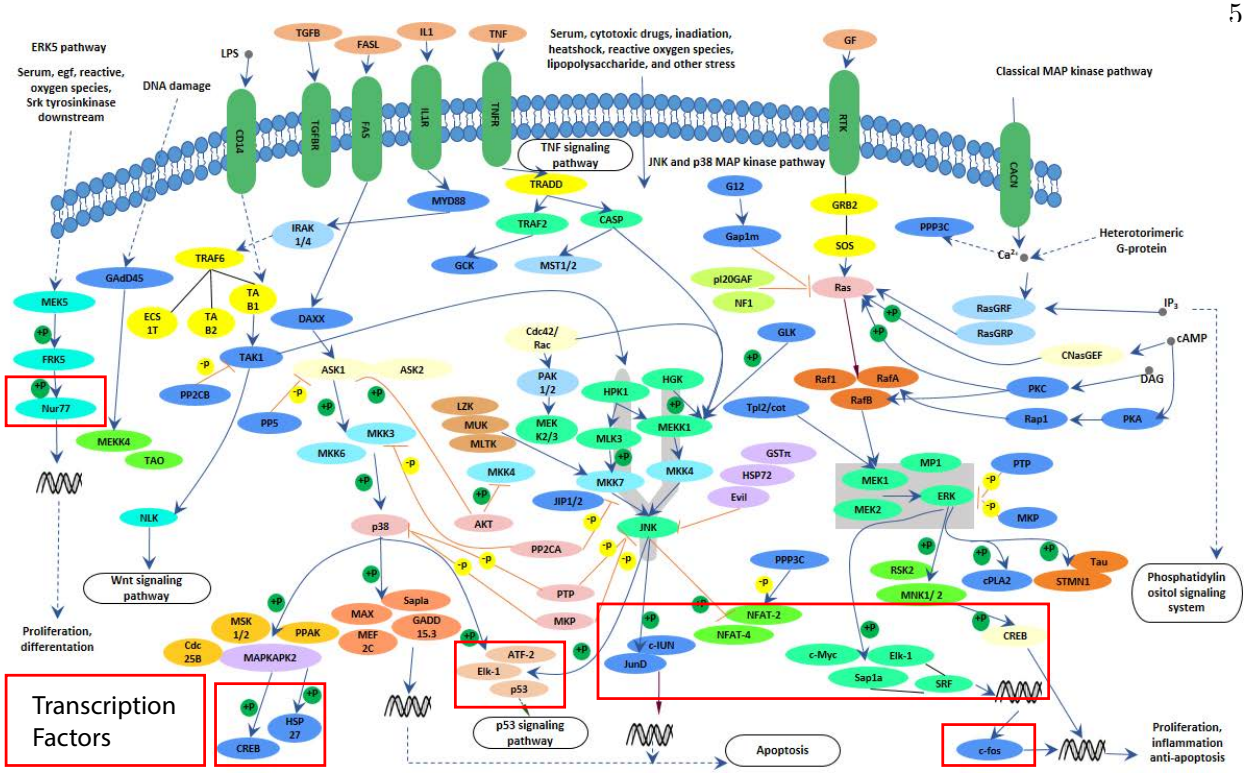


Figure 1.2: **The MAPK Pathway.** From https://www.cusabio.com/pathway/MAPK-signaling-pathway.html

1.2 Transcription Factors

Transcription factors (TFs) are a class of proteins that are responsible for coordinating the transcription of sets of genes. Typically, in its simplest form, TFs contain a DNA binding domain and a transactivation domain (see Fig 1.3a). The DNA binding domain affects what sequences a TF prefers to bind to while the transactivation domain is responsible for effecting changes in transcription through a variety of mechanisms. To effect changes in target gene transcription, TFs can bind to proximally (at promoters) or distally (at enhancers). Whether a TF binds preferentially to promoters or enhancers varies as do their mechanisms of action. In this section I will discuss several important aspects of TFs starting with mechanisms by which a TF can become active. Next, I will briefly describe key aspects that determine where a TF binds and the consequences of

this binding event. Finally, I will touch on how the direct consequences of TF binding can induce or repress target gene transcription.

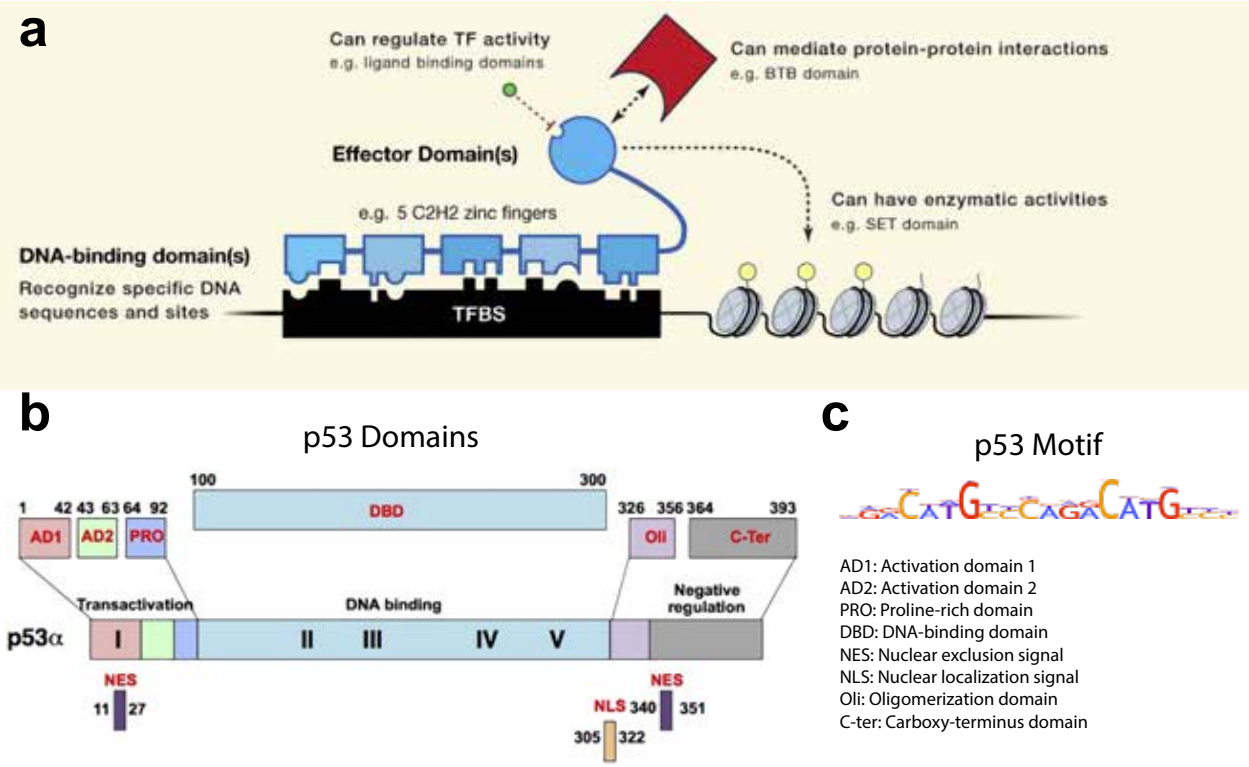


Figure 1.3: **TF Domains.** (Lambert 2018) and (Soussi 2016)

1.2.1 TF Activation

Ultimately, the activation of a cellular signaling cascade results in the activation of one or multiple TFs. These TFs then go on to effect changes in global gene expression. A TF can be activated in several different ways. Classically, these activation mechanisms require that the TF itself or a binding partner of a TF become phosphorylated by some upstream kinase. TFs or co-factors that become activated can then perform designated functions through a variety of mechanisms. Phosphorylation can directly affect TF activity by changing it's binding partners which can be accomplished by modifications to oligomerization domains on TFs, by modifying

binding interfaces between a TF and a co-factor, or by causing conformational changes to the TF or co-factor. These changes to binding partners can then result in several functional outcomes depending on the TF and the co-factor. Additionally, upstream kinases can affect the nuclear localization of a TF (or its co-factor) through phosphorylation of a nuclear localization signal or nuclear exclusion signal thus activating a TF by localizing it (or its co-factor) to the nucleus. Often, TFs have one or multiple of these mechanisms of action that all contribute to affecting its activity.

1.2.2 TF Binding to DNA

Once a TF becomes activated, it then exerts its regulatory influence by binding to DNA. However, a TF cannot simply bind non-specifically to DNA but instead must recognize specific sequences to properly accomplish its regulatory role. Studying the process by which TFs recognize specific sequences of DNA first requires a way to measure where a TF is binding across the genome. This can be accomplished by several experimental methods. The classic approach to measuring TF binding is through a technique called Chromatin Immuno-Precipitation (ChIP) which involves cellular cross-linking followed by pull-down of a protein of interest using an antibody. Sequences bound to the protein of interest are then either determined by microarray (ChIP-chip) or by next generation sequencing (ChIP-Seq). Using this technique, researchers have been able to empirically determine the sequence preference of many TFs, which are referred to as TF motifs (discussed further below). A complementary approach called SELEX accomplishes the same task but using *in vitro* binding of immobilized TFs to a set of random DNA sequences. Each method has its strengths and weaknesses. For instance, SELEX will be blind to chromatin associated effects of TF binding, whereas ChIP may be biased by co-factors, change depending on biological context, and also requires an antibody which has known caveats(Marcon 2015 and Uhlen 2016). Furthermore, sequences obtained via ChIP are always larger than the actual bound sequence due to the details of the technique (although this can be improved using ChIP-exo (Rhee 2012)). Ultimately these techniques result in high quality motifs for hundreds of TFs of interest (of the thousands present in the human genome). A comprehensive list of methods for measuring TF binding specificities is

shown in Figure 1.1.

Table 1.1: **Methods for Measuring TF Specificity.**

		Method	Description	Features			
				Capability of de novo **motif discovery** (approx. length in base pairs with high information content)	Identifies **genomic binding locations** of a TF	Can measure effect of **CpG Methylation**	Can measure **cooperative binding and/or multimers**
High-throughput	*In Vitro Methods*	Protein Binding Microarray (PBM)	A GST-tagged TF is bound to a glass slide that has ~41,000 spots of short immobilized DNA sequences. Fluorescence-based detection of bound spots and k-mer enrichment analysis yields motifs.	✔ (< 12 bp)	✘	✔ Methyl-PBM	✔
		Bacterial one-hybrid	TF binding sites are selected in bacterial cells from a randomized library that is cloned in front of selectable marker genes. Can be reversed to select proteins able to bind a constant DNA sequence using a library of variant protein sequences.	✔ (< 14 bp)	✘	✘	✘
		SELEX-based methods	Systematic evolution of ligands through exponential enrichment (SELEX) involves adding TFs to a DNA pool containing many randomized sequences and selecting for binding in multiple rounds. Related methods include HT-SELEX, SELEX-seq, and Bind-n-Seq. Selection can be performed using affinity tags, or molecular trapping on a microfluidic platform (SMiLE-seq).	✔ (< 25 bp)	✘	✔ Methyl-HT-SELEX	✔ CAP-SELEX ✔ SMiLE-seq
Mid-throughput		DAP-seq	Single step SELEX using a library of fragmented genomic sequences. Sequence diversity is less than HT-SELEX, but genomic sequences that have co-evolved with the TF are included.	✔ Limited by skewed distribution of genomic sequences	✔ Peaks are not necessarily indicative of *in vivo* binding	✔ AmpDAP-seq	✘
		HiTS-FLIP	Uses an Illumina sequencer's flowcell as a PBM chip to measure binding to orders of magnitude more DNA sequences.	✔ (< 17 bp)	✘	✘	✘
		Spec-seq	Single step SELEX using a synthesized library of degenerate sequences of interest. The lower complexity library is useful for quantitatively measuring effects of binding site mutations using sequencing.	✔ Limited by number of sequences assayed	✘	✔ Methyl-Spec-seq	✘
		MITOMI	A microfluidic device is used to isolate DNA-protein complexes from free DNA instantaneously to accurately measure the relative binding affinities of TFs to ~10,000 individual sites.		✘	✘	✘
Low-throughput		EMSA	Tests if a DNA sequence is bound by a protein by observing a shift in the electrophoretic migration of DNA.	✘ Useful for validating known binding sites	✘	✔	✔ EMSA-FRET
		DNA footprinting	DNA is incubated with a TF and then degraded using DNase-I, resulting in cuts in all positions except those that were protected by the bound TF.		✘	✔	✔
		ITC, SPR, MSTP	Isothermal titration calorimetry (ITC), Surface plasmon resonance (SPR) and Microscale thermophoresis (MSTP) measure the binding affinity of TF-DNA interactions.		✘	✔	✔
	In Vivo Methods	ChIP-based assays	Proteins are crosslinked to DNA using formaldehyde and precipitated with an antibody. Bound DNA is detected with qPCR, microarray (ChIP-chip), or sequencing (ChIP-seq). The ChIP-Exo variant incorporates exonuclease treatment to enhance resolution.	✔ Limited by skewed distribution of genomic sequences, and inability to distinguish direct from indirect binding	✔	✔ ChIP + Bisulfite-sequencing	✔ Re-ChIP
		DamID-seq	A TF is expressed in mammalian cells as a fusion to bacterial Dam-methylase. The enzyme methylates a consensus sequence in close proximity to the TF's binding sites, which can be mapped using restriction enzymes and high-throughput sequencing.		✔	✘	✔ Split DamID-seq

1.2.3 TF DNA-binding Motifs

By experimentally measuring TF binding, we can directly either directly measure or infer the sequence specificity of a given TF. Inferences must be made when using techniques that capture flanking sequences as well (*i.e.* ChIP). In reality, a TF does not bind a single sequence exclusively but instead has a range of affinities for different sequences preferring to bind to some over others. These preferences can be described as a position specific scoring matrix (PSSM), also known as a position weight matrix (PWM). Essentially this representation quantifies the probability of finding a given nucleotide at a given position from a set of sequences derived through a TF-binding assay. This is commonly referred to as a TF motif. As sequencing techniques have become more common and cost effective, there has been a drive of the TF community to characterize and curate databases with these TF motifs. Two notable examples are the HOCOMOCO (Kulakovskiy 2017) and JASPAR databases (Fornes 2019). Of an estimated several thousand TFs in the human genome (ref), HOCOMOCO contains 600 and JASPAR contains 1500 motifs. Since these motifs are empirically determined from a variety of sources, their quality can be variable. Additionally, TFs that are related in an evolutionary sense can have very similar motifs. As more experiments and validations are performed, these databases will continue to improve.

An important aspect of these motif databases is that they can be used as a standardized resource for collections of TF motifs. These motifs can then be used to find statistically significant sequence matches across the human genome. Depending on the complexity of the motif and the cutoffs used when scanning, the number of TF motif instances across the genome can be anywhere from tens to millions. However, these hits are typically orders of magnitude more abundant than actual TF binding. This discrepancy can be explained by differences in chromatin accessibility as well as binding of other factors. Despite this caveat, these genome-wide motif hits allow researchers to analyze overlaps and enrichment of these motifs within their own datasets with minimal overhead. Throughout this thesis, these TF motifs will become a critical asset in the development of computational approaches to infer TF activity.

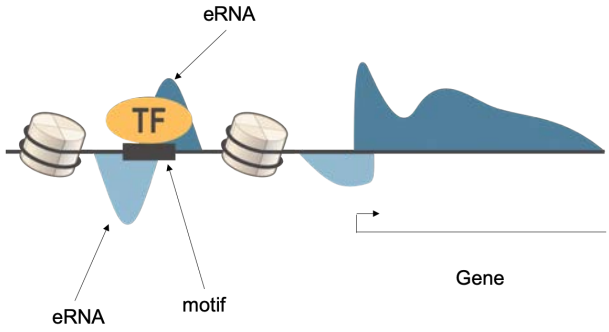


Figure 1.4: **Enhancer Schematic (David Deen (http://www.daviddeen.com/)).**

1.2.4 Enhancers

Classically, TFs bind to regions of the genome called enhancers and promoters. Promoters are well-defined regulatory regions immediately upstream of a gene. Enhancers on the other hand are still being characterized and therefore what is considered an enhancer is still a widely debated topic. The genetic definition of an enhancer is a sequence of DNA that upon insertion into a reporter assay, can drive the expression of a gene regardless of orientation or position relative to the target gene. As sequencing techniques have become more widespread, this definition has incorporated other aspects of observed enhancer characteristics. One such aspect is the deposition of the histone mark H3K27ac which can recruit regulatory protein complexes which can contain bromodomains (Raisner 2018). Other marks such as H3K4me1 and H3K4me2 have also been associated with enhancers (Calo2013), although some argue that these marks are instead simply a readout on transcriptional levels (Core 2014). More recently, nascent sequencing ((*i.e.* Global Run On (GRO) (Core 2008), 4-thiouridine (4sU) (Windhager 2012), Bromouridine (Bru),(Paulsen 2013), and transient transcription (TT-Seq) (Schwalb 2016), Precision Run On (PRO) (Mahat 2016) -Seq, see (Wissink 2019) for a full review) has revealed that some enhancers themselves are transcribed. The resulting enhancer RNA (eRNA) typically (but not always) has the signature of a bidirectional transcript (RNAs transcribed in the forward and reverse strand from a single point of origin Figure 1.4). This bidirectional transcript is presumably due to the lack of strand specificity inherent to the transcriptional machinery.

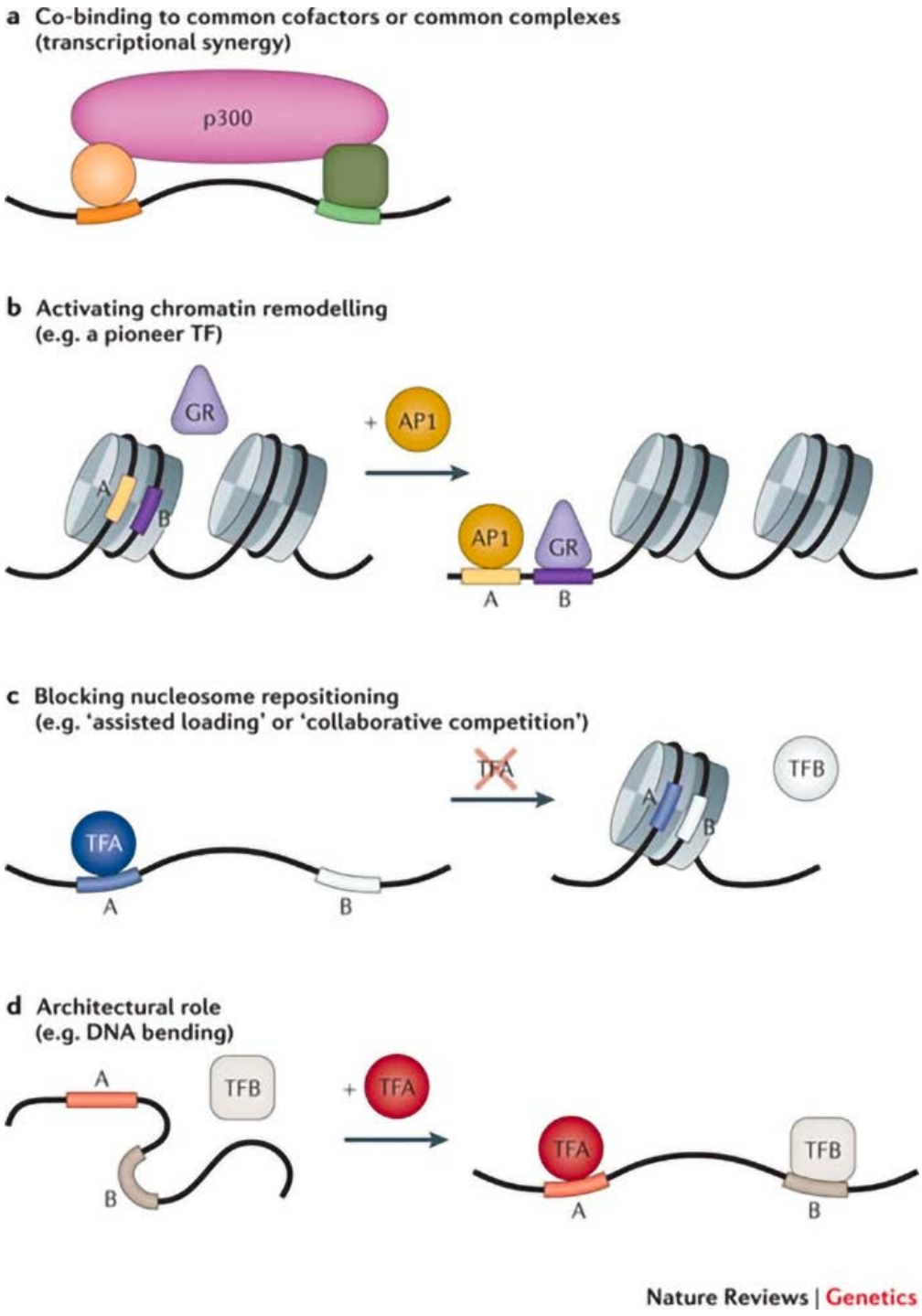


Figure 1.5: **Mechanisms of TF-dependent Enhancer Activation (Francois 2012).**

1.2.4.1 Enhancer Function

How enhancers and eRNAs function to regulate gene transcription is still a widely debated topic. This is complicated by the fact that not all TF binding results in eRNA transcription and assigning an enhancer to its target gene is a difficult task to perform in a high-throughput manner using current experimental approaches. While the effects of eRNA production are still being explored, several mechanisms have been proposed on how, after TF binding, a bound enhancer region can recruit co-factors or change the chromatin landscape (Figure 1.5). While these mechanisms describe molecularly what happens at enhancers, it does not explain why there are many more enhancers than genes and many TF binding sites per enhancer. These aspects of enhancers are to presumably have tighter control over gene expression. This explanation makes sense if instead of viewing gene expression as a binary on/off switch, we posit that the actual levels of gene expression are important for correct cellular function (Lee 2013). In this case, having many enhancers that you can tightly control using TFs allows a cell to fine tune the expression of genes (see Figure 1.6). Another prevailing theory (and these are not mutually exclusive) is that enhancers function to create phase-separated domains.

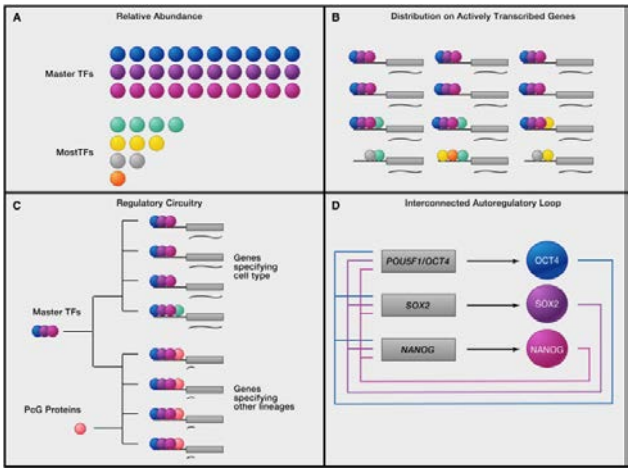


Figure 1.6: **TF and enhancer circuitry**

Phase separation is a way to compartmentalize biological molecules within a cell without

the use of a lipid membrane. Researchers have proposed that due the weak interactions between eRNAs and the intrinsically disordered domains found on TFs and many of the general transcription machinery could create phase separated domains. These liquid-like droplets could be a mechanism by which cells can maintain high concentrations of biological molecules required for transcription thus mitigating the need to search the large genomic space for genes to transcribe. This mechanism also allows for re-engaging of polymerase after it finishes transcribing at highly expressed genes. A summary of potential interactors in the phase separation model is shown in Figure 1.7.

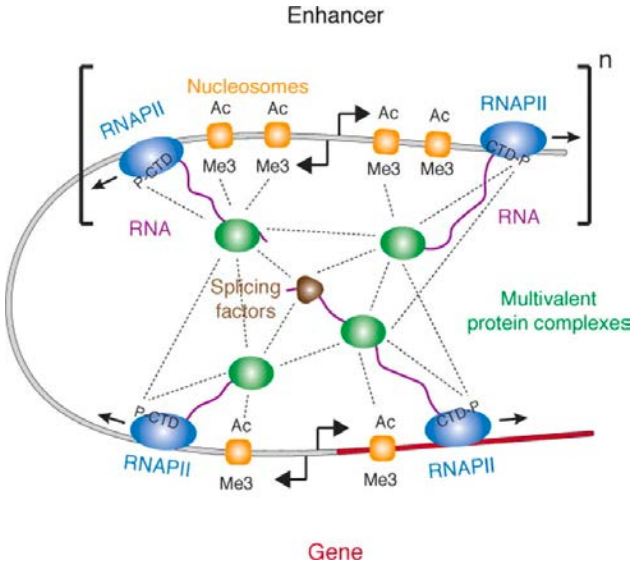


Figure 1.7: **Enhancer Phase Separation Model. (Hnisz 2017)**

1.3 Transcription

Transcription is the process by which DNA is converted to RNA. In eukaryotes, three RNA Polymerase (RNAP) enzymes are responsible for transcription of different RNAs (RNAPI - ribosomal RNA, RNAPII - protein coding genes, RNAPIII - 5S ribosomal subunit, tRNAs, and other small RNAs). Transcription occurs generally in three stages: loading/initiation, elongation, and cleavage/termination. Each stage has critical regulatory function and events important for proper RNA processing occurs at each stage. Below I will discuss these stages of transcription as they relate

to the behaviour of the main transcription enzyme of protein coding genes RNAPII. A summary of the transcriptional stages discussed in these sections is show in Figure 1.8.

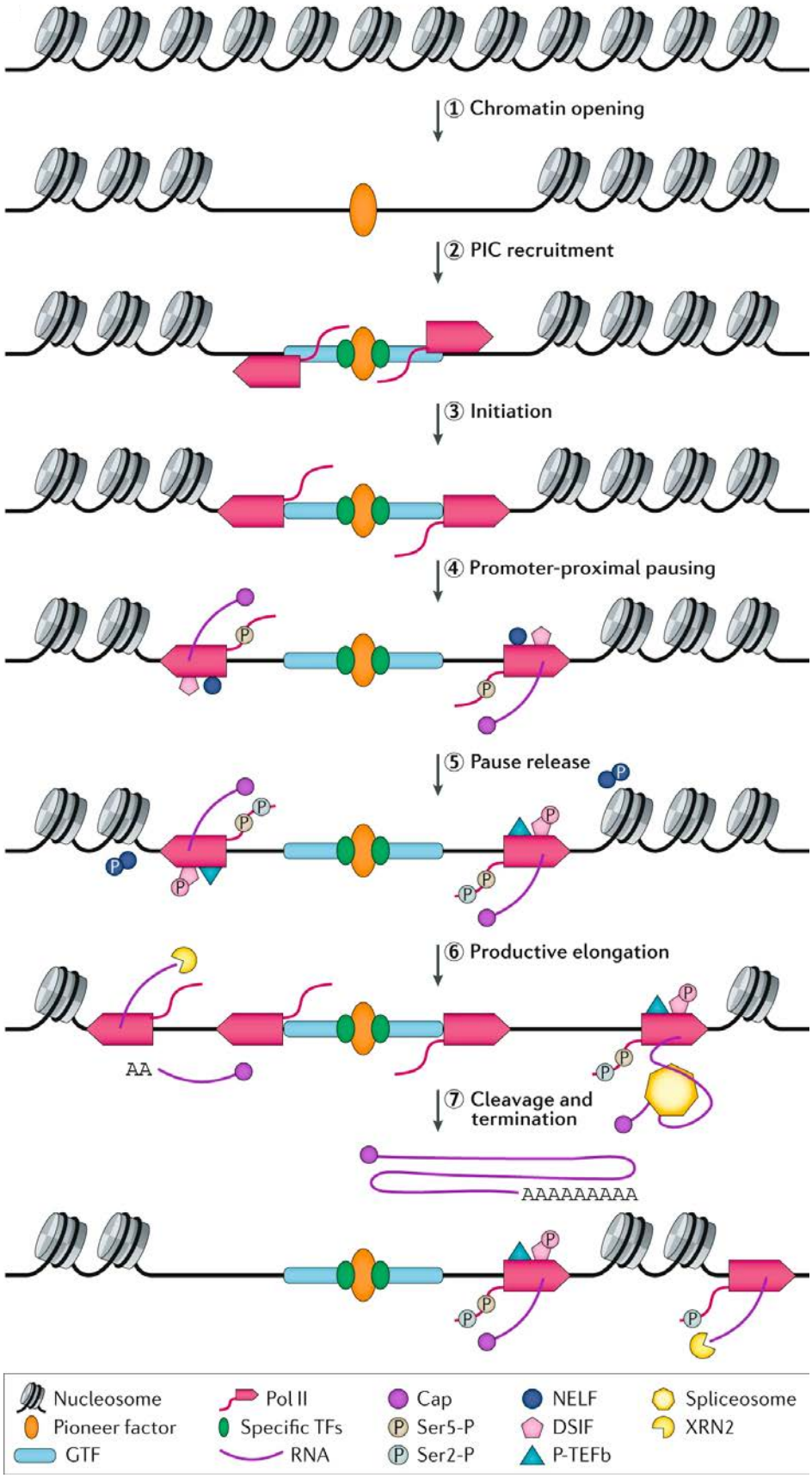


Figure 1.8: **Stages of Transcription.**

1.3.1 Chromatin Opening

Chromatin refers to the organizational structures in a cell containing DNA, RNA, and protein that make up chromosomes. The main structural unit of chromatin are large protein complexes called histones that wrap around DNA forming nucleosomes. Organization of the DNA into chromatin allows for regulation of gene transcription and proper packaging of genetic material for cell division. To begin transcription of a gene, the region encompassing that gene must first be accessible for binding of other factors and enzymes (step 1 in Figure 1.8b). Chromatin accessibility can range from completely inaccessible to open usually determine by specific marks deposited on histones (see Klemm 2019 for review). As described previously, TFs are largely responsible for opening chromatin by directly recruiting specific enzymes that can displace nucleosomes or by indirectly depositing histone marks that then recruit said enzymes. Once a region of chromatin becomes accessible, protein factors can be recruited to begin the process of transcription.

1.3.2 The Pre-Initiation Complex

Transcription is a highly regulated and complex process that involves the interplay of many protein factors. Generally, to transcribe a gene, a coordinated assembly of a pre-initiation complex (PIC) is required (Figure 1.9, step 2 in Figure 1.8b). Assembly of the PIC involves recruitment of several general transcription factors (GTFs) including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. These GTFs play specific roles in recruiting RNA Polymerase II (RNAPII) to designated regions along the genome and stabilizing different complexes (see Gupta 2016 for review). Another key component of the PIC is the Mediator complex that I will discuss in more detail in section 1.4.

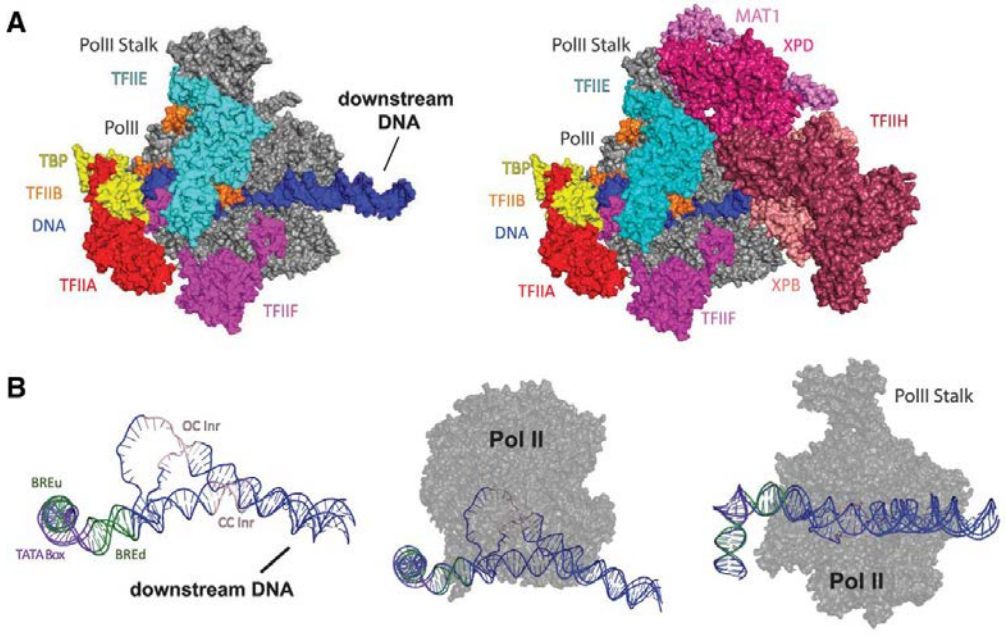


Figure 1.9: **Assembly of the Pre-Initiation Complex.**

1.3.3 Loading and Initiation

Loading and initiation are thought to be the main point of regulation within the transcription process. To load, a RNAPII enzyme is first recruited onto DNA and then initiates by transcribing a small amount (8-14nt) (Sims III 2004) (step 3 of Figure 1.8b). Since there are two strands of DNA, RNAPII must choose the correct "sense" strand to transcribe the coding region of a specified gene. Inherently, the RNAPII enzyme itself is not strand-aware and must rely on components of the PIC to guide its decision of which strand to transcribe. This decision is not always a completely binary one leading to the observation of bidirectional transcripts or RNA molecules produced from both strands originating from a common point of RNAPII loading (see Figure 1.4 and Figure 1.8a). This phenomenon occurs both in regions that do not contain a sense direction (*i.e.* enhancers) but also in regions that have a coding gene and therefore a sense direction. This bidirectionality is likely due to the inherent stochasticity of RNAPII loading and strand decision. While these bidirectional transcripts have not been well studied, they can be used as a signature for sites of RNAPII loading which will be discussed later in this thesis.

1.3.3.1 Pausing

Once a RNAPII enzyme loads onto DNA, picks a strand, and initiates, it releases from the PIC and transcribes another small amount (60bp) before pausing (Adelman 2012) (step 4 in Figure 1.8b). This pause is largely seen in nascent sequencing data and can theoretically be due to a physical resident pause of RNA polymerase or abortive transcription. Either way, we will refer to this phenomenon as pausing as that is the prevailing model and referenced most in the literature (Adelman 2012). RNAPII pausing is an elusive process that is hard to study due to the sparsity of methods that can be used to quantify this phenomenon. It is estimated that around 30% of genes exhibit pausing (Adelman 2012). The molecular mechanisms of pausing involve association of several factors with the transcribing complex containing RNAPII. NELF and DSIF are two canonical pausing factors that associate with transcribing RNAPII causing it to pause (Muse 2007, Wada 1998, Yamaguchi 1999). More recently, the GTF TFIID was shown to be sufficient for RNAPII pausing (Fant 2020). Pausing has been proposed to serve a variety of functions for gene transcription. Broadly speaking, these proposed mechanisms involve maintaining open chromatin (Wu 1980, Costlow 1984, Gilchrist 2010), synchronizing gene transcription (Gressel 2017, Shao 2017), and facilitating RNA processing (Rasmussen 1993, Tome 2018).

1.3.4 Elongation

Paused RNAPII complexes become released into productive elongation following the phosphorylation of NELF and DSIF by the super elongation complex (SEC) containing PTEF-b (step 5 in 1.8b). Elongation is the process by which RNAPII transcribes along a gene body (*i.e.* the region of a gene that is between the pause site and transcription end site, step 6 in Figure 1.8b). Depending on the gene and cellular context, elongation proceeds at a rate of 2-4kb/min (refs).

As a nascent RNA is being transcribed, protein factors can associate with the RNAPII complex or the RNA itself to facilitate splicing. Splicing is the process by which non-coding regions of a gene (introns) are excluded in favor of coding regions (exons).This process happens

co-transcriptionally and is also proposed to involve phase-separation (ref). Splicing allows for different functional isoforms of a single gene by varying inclusion/exclusion of different exons, modifying transcription start sites, and differential inclusion of introns. Splicing is a key player in gene regulation and has gained attention as many diseases are associated with defects in the splicing process (ref). Additionally, changes to elongation rates can have effects on splicing (refs).

1.3.5 Termination

Termination of transcription describes the process by which the polymerase enzyme releases from DNA. This is temporally preceded by cleavage of the nascent RNA. Termination is an essential process that is required for proper processing of RNAs following transcription and therefore essential for the proper expression of genes into protein products. Two models have been proposed for how termination is accomplished molecularly. The first model states that the RNAPII enzyme detects poly-adenylation sequences (PAS) and undergoes conformational changes leading to proper termination via the recruitment of the cleavage and poly-adenylation complex (CAP) (Zhang 2015). The second model, called the torpedo model, states that after cleavage of the nascent RNA, the exonuclease Xrn2 begins degrading the RNA being produced by the still-elongating RNAPII enzyme. Once Xrn2 catches up with RNAPII, this triggers its release from DNA (Connelly 1988, Proudfoot 1989). Recently, termination has been shown to be affected by cellular stress with the observation that after heat shock, global polymerase occupancy extends far beyond the observed PAS in non-treated cells (Cardiello 2018).

1.4 Mediator

Among the GTFs that make up the PIC is a large protein complex called Mediator, an essential factor for regulated transcription by RNAPII. Mediator functions as a molecular bridge, communicating the signals from distal TF binding events to the PIC. This is accomplished by undergoing conformational changes in response to co-factor binding such as RNAPII or different TFs (Figure 1.10). These conformational changes then act as signals that can be relayed to the

PIC.

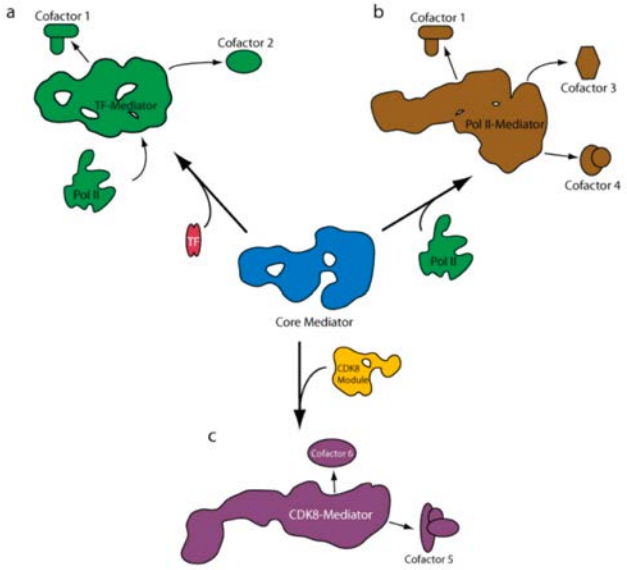


Figure 1.10: **Mediator Conformational Changes.** (Allen 2015)

Mediator is considered a regulatory hub and required for the downstream activation of genes in response to TFs following a cellular signaling cascade. Specific TFs can bind to regions across the Mediator complex resulting in different functional outcomes (Figure 1.11). In addition to the core Mediator complex, there are two kinases (in humans) that form their own submodules and can reversibly associate with Mediator - CDK8 and CDK19.

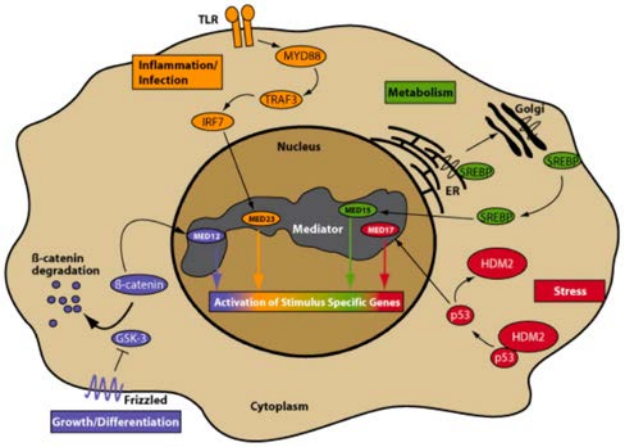


Figure 1.11: **Mediator Functions Downstream of Signaling.** (Allen 2015

1.4.1 Mediator Kinases

As briefly described above, Mediator kinases are cyclin dependent kinases (CDKs). In contrast to the classic CDKs that associate mostly with the cell cycle, Mediator kinases are part of so-called transcriptional CDKs. Transcriptional CDKs are known to have effects on transcription as part of larger complexes such as CDK7 which associates with TFIIH and CDK9 which associates with PTEF-b. While CDK7 and CDK9 are relatively well-studied, Mediator kinases remain enigmatic. Below I discuss several known aspects of Mediator kinase function including their role in disease.

1.4.2 CDK8 Transcriptional Effects

Because of their association with Mediator, Mediator kinases were first studied as a part of the general transcription machinery. Their association with Mediator and the general PIC was confirmed using pulldown techniques (ref). Initially, CDK8 was found to repress transcription independent of its kinase activity in *in vitro* transcription assays (Knuesel 2009). Supplementation with additional RNAPII enzyme was found to relieve that repression suggesting these might compete for binding within the PIC. This was further confirmed by the observation that stably associated RNAPII-containing PICs were not affected by addition of CDK8. This has been described as a

mechanism by which CDK8 could facilitate promoter escape (ref matt galbraith?) by binding to Mediator. Upon promoter escape of RNAPII, CDK8 would then dissociate allowing for a new RNAPII molecule to be recruited to the PIC (Figure 1.12). This molecular switch mechanism described a way in which cells could consistently transcribe genes without the need to re-assemble a new PIC for each transcription event.

Several studies have supported a role of CDK8 in transcription. Within the serum response, researchers discovered that the CDK8 protein was required for elongation of serum response genes. This study attributed this elongation defect to a lack of PTEF-b recruitment, a component of the super elongation complex (SEC). Further support for the role of CDK8 in regulating the SEC came from studies characterizing the direct targets of Mediator kinase activity (poss). [Other studies?]

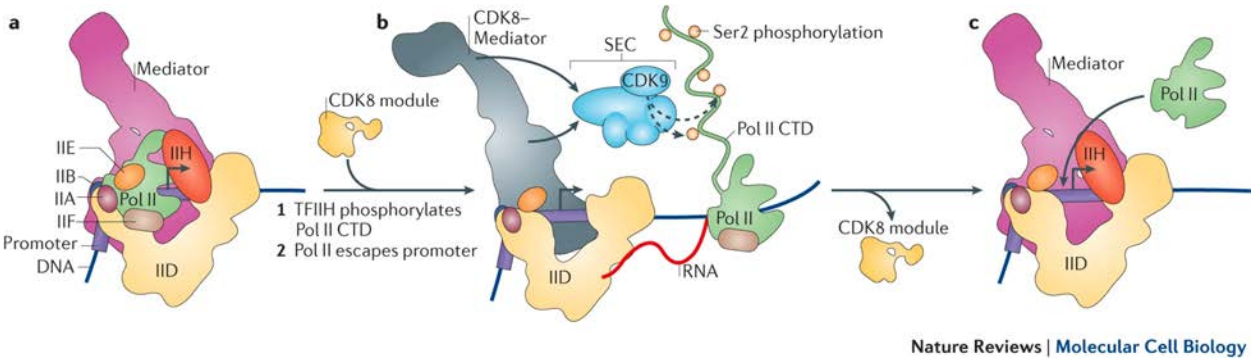


Figure 1.12: **CDK8 Function within the PIC.**

1.4.3 CDK8 Signaling Effects

In addition to its roles as a regulator of general transcription, CDK8 has also been shown to play roles in cellular signaling. Among its known targets, transcription factors (TFs) make up a significant portion (25% see Figure 1.13) (Poss 2016). In initial studies, CDK8 was found to be a positive co-regulator of p53 target genes (Donner et al. 2007). This study found an association of CDK8 module binding at p53 target genes using different cellular stimuli to activate the p53 pathway. In a cancer study, CDK8 expression was shown to correlate with the expression

of β-catenin, a key player in Wnt signaling, in colorectal cancer patients with poor prognosis. However, CDK8 knockdown did not affect the expression of β-catenin in this study (ref Seo et al. 2010). Additionally, CDK8 was shown to be involved in lipogenesis during starvation through the regulation of the TF SREBP-1 in *Drosophila*, mice, and humans (Zhao et al. 2012). These effects were shown to be mediated by SREBP-1 phosphorylation and inhibited by the overexpression of Cyclin C, the kinase activating cycling associated with CDK8. Furthermore, a recent study implicated CDK8 as a regulator of gene expression within the TGF-β/SMAD signaling pathway (ref Liang et al. 2018). CDK8 was shown to regulate the expression of matrix metalloproteinases through regulation of the microRNA miR-181b. This study found CDK8 inhibition to be an effective means to limit growth of metastatic colorectal tumors in the liver. Finally, several studies have implicated Mediator kinases within the IFN-γ pathway (ref Steinparzer 2018, Guo 20190 in response to either IFN-γ treatment or cellular differentiation. These studies point a role of CDK8 in cellular signaling, especially in response to changes in the cellular environment due to perturbations or cellular context. However, most research in this field lacks true mechanistic insight into exactly how Mediator kinases drive these transcriptional reprogramming events.

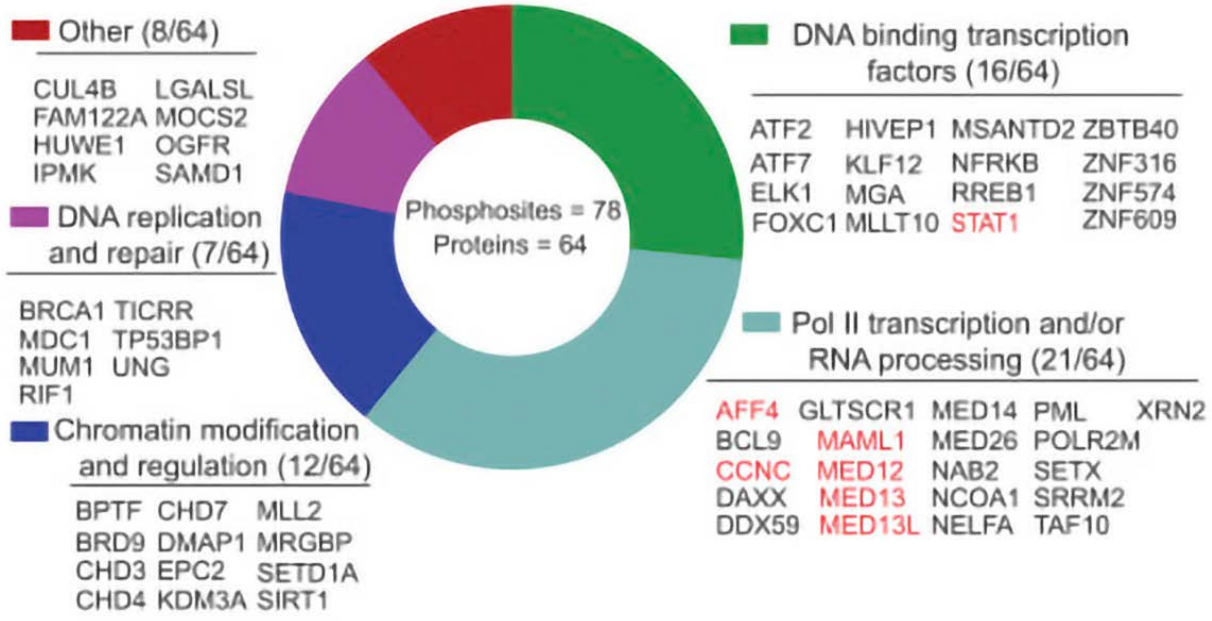


Figure 1.13: **CDK8 Targets.**

1.4.4 Mediator Kinases in Disease

Mediator kinases have gained attention largely due to their association with disease. In particular, CDK8 has been shown to be overexpressed in colorectal cancers with overexpression correlating with negative outcomes in patients (ref). Additionally, CDK8 and CDK19 play crucial roles in development - as evidenced by the embryonic lethality of CDK8 (Westerling 2007) and the association of CDK19 with developmental disorders (Mukhopadhyay 2010, Chung2020). Because of its disease association, time and effort have been spent in investigating the mechanisms by which Mediator kinases can cause disease. However, these mechanisms are not completely understood, in part because of its various roles in different cellular contexts.

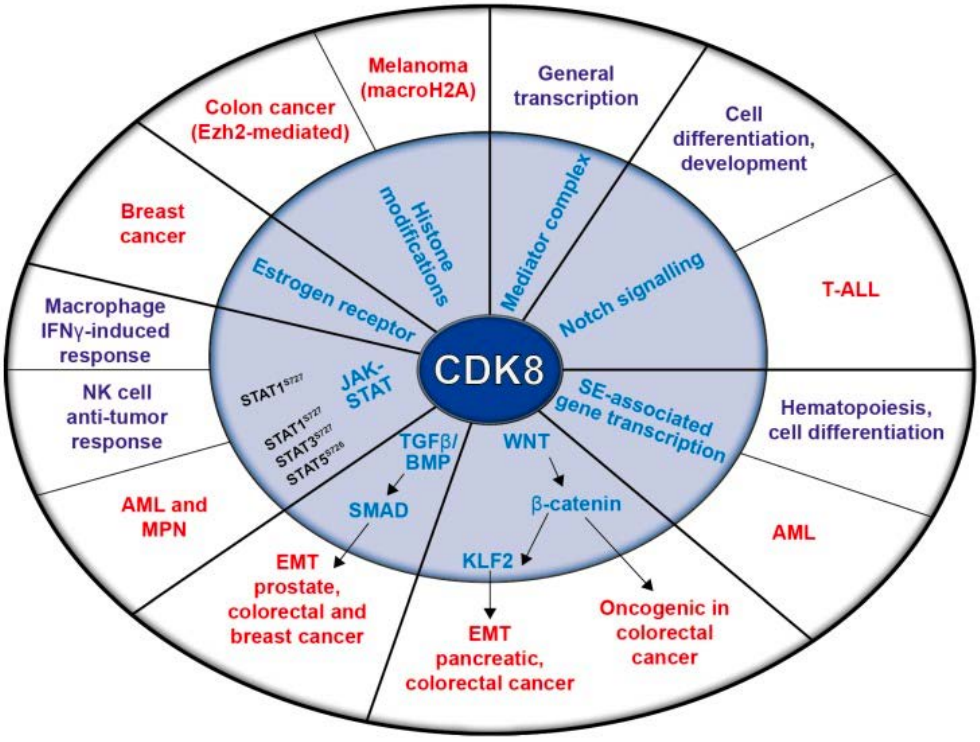


Figure 1.14: **CDK8 in Cellular Signaling and Disease.**

1.4.5 Mediator Kinase Inhibition

One major question left to be answered is whether Mediator kinase activity plays a major role in these phenotypes. In several cancers, including AML, the kinase activity itself has been shown to be necessary for cell proliferation. This has led researchers to screen small molecules to be used as therapeutics. One of these molecules, called Cortistatin A (CA) was found in sea sponges in a search for antiproliferative compounds against AML specifically (Ref). CA is an especially potent and specific inhibitor of Mediator kinase activity, found to be selective in kinome-wide screens for CDK8 and CDK19 over all other kinases. CA is the main inhibitor used in the second half of this thesis and has been shown to decrease tumor burden in animal models of AML (Pelish). The cell proliferative effects of Mediator kinase inhibition has already been shown to be inconsistent across cell types. Despite its overexpression in colorectal cancer, similar cell proliferation defects were not found in cell culture models following treatment with Mediator kinase inhibitors. However, one

study found defects in liver metastases of colon cancer tumors in mice following Mediator kinase inhibition. Why Mediator kinase inhibition is cell type specific and the mechanistic roles of these kinases in cell proliferation are still open questions.

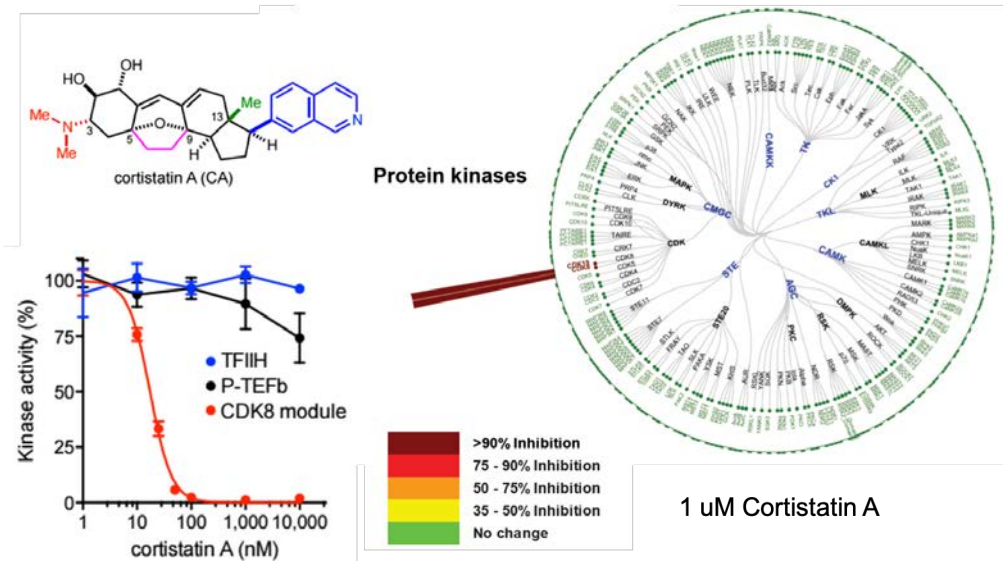


Figure 1.15: **Cortistatin A Selectivity for Mediator Kinases.**

1.5 Thesis Summary

In this thesis, I will present work that describes the development of a computational method for detecting TF activity followed by the application of that method to study the role of Mediator kinases within signaling networks. In the first chapter, I present early work on the motif-displacement (MD) score approach, a technique that takes advantage of eRNA detection in combination with *in silico* derived TF motif locations to quantify TF activity. This technique was shown to accurately capture known TF activation following perturbation. In the second chapter, I expand on this technique with another computation technique called Transcription Factor Enrichment Analysis (TFEA). TFEA adds several improvements to the MD-Score method including the ability to detect non-binary changes to eRNA transcription as well as a principled way to handle sample replicates. I show that TFEA outperforms the MD-Score in single timepoint perturbation experiments.

Using data simulation, I show that TFEA also outperforms similar software in its minimization of false positives. Finally, using publicly available time series datasets, I show that TFEA can deconvolute complex regulatory networks and arrive at mechanistic inferences across different data types. In the third chapter, I transition to studying Mediator kinase activity, specifically how it relates to the IFN-γ response in mouse and human cells. Employing the MD-Score approach, I show that the activity of IFN-γ related TFs decreases following Mediator kinase inhibition. This study also showed the relative contributions of CDK8 and CDK19 to this response. Finally, in the fourth chapter of this thesis, I present work on how Mediator kinase inhibition affects cellular proliferation during serum response in a colon cancer cell culture model. Using a variety of -omics techniques, I probe the effects of Mediator kinase inhibition on transcription, gene expression, metabolism, protein phosphorylation, and cell proliferation. The results of this study ultimately lead to key insights into Mediator kinase activity and suggest that this could be a therapeutic strategy for treatment of colon cancer.

Chapter 2

Motif Displacement Scores

2.1 Preamble

An observation from nascent sequencing is that sites of polymerase loading and initiation often exhibit bidirectional transcription (Core 2008). Furthermore, distal sites of bidirectional transcription were found with no annotated gene nearby. Later, these sites were found to be associated with enhancers and these distal (relative to genes) bidirectional transcripts became known as enhancer RNAs (eRNAs). Several groups then found that upon activating a transcription factor, the levels of eRNAs increased and these eRNAs overlapped with the activated TFs motif (refs). We therefore sought to determine whether this was a general phenomenon of TF activation and whether it could be used as a marker for TF activity.

2.1.1 Significance

A cell undergoes functional reprogramming to determine cell fate through the activation of transcription factors (TFs). Therefore, measuring TF activity is a global readout of cellular processes. Measuring this activity however is a difficult task, especially since there are estimated to be 1800 TFs in the human genome. Historically, measuring TF activity has relied on detecting TF binding via ChIP. However, recent evidence shows that not all TF binding events are functional (ref). In addition, ChIP relies on antibodies which have known caveats. ChIP is also a low-throughput technique only capable of measuring binding of one TF at a time. Gene expression has been combined with ChIP to obtain more reliable measures of activity however, the poor temporal

dynamics of RNA-Seq coupled with the non-trivial need to assign TF binding sites to target genes has limited the effectiveness of this approach.

In this work, we devised a method to computationally measure TF activity based on the locations of eRNAs relative to TF motifs. This allowed us to measure the activity of any TF whose motif was known. Because there exist repositories with such motifs, this method allowed us to simultaneously measure the activity of hundreds of TFs from a single experiment. We further show evidence here that eRNA transcription correlates with target gene transcription, a crucial point that supports a functional role for eRNAs.

This work was published in Genome Research Volume 28, Pages 334-344 on February 15, 2018 (doi: 10.1101/gr.225755.117).

2.1.2 Contributions

This was a highly collaborative project with many researchers contributing significantly to its completion. The majority of this work involved metric development and testing which was completed by Dr. Azofeifa. Data husbandry was completed by Dr. Allen and Josephina Hendrix. Finally, Dr. Read made significant contributions to method evaluation. What follows are my specific contributions to this project.

I joined this project shortly after a then graduate student in the lab, (now) Dr. Azofeifa, had developed an algorithm to model RNA polymerase II behaviour to detect the presence of bidirectional transcripts which were putative enhancers and gene promoters[**?**]. A then post-doc in the lab, Dr. Allen, had observed that in response to the small molecule activator of p53 Nutlin, eRNAs became transcribed over p53 motifs[**?**]. We therefore sought to devise a way to quantify this effect more broadly.

I began under the guidance of Dr. Azofeifa to first determine whether the p53 motif could be recovered from bidirectional transcripts. I then used this motif to determine the co-localization of the p53 motif relative to the locations of identified bidirectional transcripts. While performing this task, I realized there was no need to perform *de novo* motif discovery as there were publicly

available databases containing hand-curated TF motifs (HOCOMOCO, JASPAR). I then wrote scripts to perform the motif displacement analysis on all motifs within the HOCOMOCO database. The first measure of TF activity was performed by fitting the histogram of motif hits to a normal distribution. We began by using the p-value associated with that fit as our measure of TF activation. The first evidence to support the validity of this approach involved measuring these p-values for the p53 motif in HCT116 cells treated with DMSO, Nutlin, or a p53 -/- cell line. We observed that the normality of the distribution of p53 motifs over bidirectionals increased as we moved from the knockout cell line, to DMSO treated cells, to Nutlin treated cells. Eventually, we switched to a simpler metric that measured the number of motif hits within a small window compared to the motif hits within a larger motif window called the motif displacement score (MD-Score). The data and experiments associated with this analysis would eventually become Figure 3a of this paper.

In addition to these preliminary analyses, I performed supporting work associated with this paper. I wrote scripts to perform an intersection of identified bidirectional transcripts with all available histone marks in HCT116 cells from the ENCODE repository used in Figure 1b. I also developed a support vector machine to determine which of these marks was most predictive of the presence of bidirectional transcripts - unsurprisingly, the result was H3K27ac.

What follows is the culmination of several years of weekly meetings and many hours of analysis and re-analysis, mostly by Dr. Azofeifa. The results of this study were published in Genome Research and are as follows.

Research

Enhancer RNA profiling predicts transcription factor activity

Joseph G. Azofeifa,[1,2] Mary A. Allen,[2] Josephina R. Hendrix,[1,3] Timothy Read,[2,4] Jonathan D. Rubin,[4] and Robin D. Dowell[1,2,3]

[1]*Department of Computer Science, University of Colorado, Boulder, Colorado 80309, USA;* [2]*BioFrontiers Institute, University of Colorado, Boulder, Colorado 80309, USA;* [3]*Department of Molecular, Cellular and Developmental Biology,* [4]*Department of Biochemistry, University of Colorado, Boulder, Colorado 80309, USA*

Transcription factors (TFs) exert their regulatory influence through the binding of enhancers, resulting in coordination of gene expression programs. Active enhancers are often characterized by the presence of short, unstable transcripts termed enhancer RNAs (eRNAs). While their function remains unclear, we demonstrate that eRNAs are a powerful readout of TF activity. We infer sites of eRNA origination across hundreds of publicly available nascent transcription data sets and show that eRNAs initiate from sites of TF binding. By quantifying the colocalization of TF binding motif instances and eRNA origins, we derive a simple statistic capable of inferring TF activity. In doing so, we uncover dozens of previously unexplored links between diverse stimuli and the TFs they affect.

[Supplemental material is available for this article.]

Transcription is orchestrated by the sequence-specific binding of transcription factors (TFs) to DNA, resulting in regulation of gene expression programs (Spitz and Furlong 2012). Hence, TFs function as major determinants of cell state (Takahashi and Yamanaka 2006; Rackham et al. 2016). Chromatin immunoprecipitation (ChIP) studies have identified binding sites for many of the approximately 1400 TFs encoded within the human genome (Vaquerizas et al. 2009), allowing estimation of a DNA-binding motif model for more than 600 factors (Kulakovskiy et al. 2013). However, studies comparing TF binding events to RNA expression levels have revealed that many TF binding sites have no apparent effect on nearby transcription (Li et al. 2008; Fisher et al. 2012; Read et al. 2016). Distinguishing such "silent" TF binding events from those with regulatory capacity is a fundamental challenge. Despite their critical importance for controlling cellular phenotypes, it is difficult to ascertain when a TF is active, e.g., contributes to nearby transcription.

One notable attempt to infer TF activity leveraged patterns of TF motif instances at annotated protein coding genes to explain changes in expression (The FANTOM Consortium and Riken Omics Science Center 2009; Balwierz et al. 2014). Yet, most TF binding occurs within regions of the genome distal to protein coding genes (Spitz and Furlong 2012). These binding events often correspond to enhancer regions known to be important for regulation of gene expression and cellular identity (Heintzman et al. 2009). Active enhancers are often characterized by the presence of short, unstable, bidirectional transcripts termed enhancer RNAs (eRNAs). When a specific TF is activated, eRNA transcription generally increases at the location of the TF binding event (Danko et al. 2013; Hah et al. 2013; Allen et al. 2014; Puc et al. 2015). While the functions of eRNAs are only beginning to be understood (Hah et al. 2013; Li et al. 2013; Sigova et al. 2015), their presence is nonetheless an indicator of enhancer activity (Andersson et al. 2014; Danko et al. 2015).

eRNA detection requires extremely sensitive methods, both in the laboratory as well as computationally. Because they are unstable, eRNAs are rarely observed via steady-state RNA assays such as RNA-seq. Nascent transcription assays capture transcription throughout the genome, including eRNA transcription (Core and Lis 2008; Core et al. 2014; Nojima et al. 2015). We recently described a model capable of estimating sites of bidirectional transcript initiation at single-base-pair resolution (Azofeifa and Dowell 2017). Transcription fit (Tfit) leverages the known behavior of RNA polymerase II (RNAP) to identify individual transcripts within nascent transcription data (Azofeifa and Dowell 2017). Although Tfit does not implicitly assume polymerase initiation will be bidirectional, we observed bidirectional transcription at both promoters and enhancers (Azofeifa and Dowell 2017). Whether bidirectional (two transcripts) or unidirectional (one transcript), our model precisely infers the point of RNA polymerase loading, i.e., the origin point of transcription.

Here, we leverage the Tfit model to ascertain TF activity. We show that, by calculating the frequency of TF binding motif instances relative to the location of eRNA initiation, the activity of the TF itself can be inferred from nascent transcription data alone. We apply our model to hundreds of publicly available human and mouse nascent transcription data sets to discover previously unknown links between TF activity and diverse biological phenomena.

Results

eRNA origins mark sites of regulatory TF binding

To utilize Tfit across a broad set of nascent transcription data sets, we modified the algorithm both to rapidly identify all sites of transcript initiation genome-wide and to account for the variable

Corresponding author: robin.dowell@colorado.edu
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.225755.117. Freely available online through the *Genome Research* Open Access option.

distances between forward and reverse strand transcripts observed across distinct nascent transcription data sets (see Methods). As a first application and validation of this revised algorithm, we identified 39,633 putative sites of bidirectional transcription in a K562 GRO-cap data set (Core et al. 2014), of which 30,324 were not associated with an annotated promoter (Supplemental Figs. S1, S2). As previously observed (Danko et al. 2015; Azofeifa and Dowell 2017), marks of active chromatin as well as TF binding events strongly associate with Tfit-predicted sites of bidirectional transcription (Supplemental Figs. S3–S5; Supplemental Table S1). Given their distal location relative to promoters, their overwhelming co-association with marks of active chromatin, and their association with TF binding complexes (Supplemental Fig. S6), we refer to non-promoter-associated Tfit polymerase loading positions as eRNA origins.

Although the vast majority of eRNA origins localize with TF binding, only a fraction of TF binding sites overlap eRNA origins (Supplemental Fig. S3A). Previous efforts to predict sites of TF binding using joint eRNA and TF-DNA motifs focused on only a small set of TFs (Danko et al. 2015). We extended this analysis to include 139 TF ChIP-seq experiments and observed a wide spectrum of association between TF binding sites and eRNA presence, suggesting that eRNA presence alone is not sufficient to fully explain TF binding (Fig. 1A). These data are consistent with the observation that only a fraction of TF binding sites result in a concomitant change in nearby gene expression (Cusanovich et al. 2014; Savic et al. 2015).

Given the strong relationship between active chromatin and eRNA transcription, we asked whether eRNAs discriminate "silent" from "active" TF binding. In support of this hypothesis, TF binding sites occurring at sites of eRNA origination display a significantly increased overlap with canonical marks of active chromatin relative to non-eRNA-associated TF binding (Fig. 1B). Moreover, no statistical difference is detected between these categories for repressive chromatin marks.

Although regulatory TF binding is often enriched for open and active chromatin, functional TF binding must ultimately lead to a change in gene expression. To this end, we considered TF binding events within enhancers conserved between two cell types but differing in terms of eRNA presence with the hypothesis that neighboring gene expression would be elevated in the eRNA-harboring cell type (Fig. 1C). There are 95 TFs profiled in at least two cell types for which cell-type–matched nascent transcription is available (Supplemental Table S2). For example, binding of the TF NR2F2 was profiled in both K562 and MCF-7 cell lines, yielding 30,618 and 16,678 binding peaks, respectively, with 3491 peaks shared between the two cell types (Fig. 1D). Of these cell-type–invariant peaks, 25% harbor an eRNA origin in both cell types, 7% only in K562, and 12% only in MCF-7, and 56% do not harbor an eRNA origin in either cell type. Measuring the transcription level of nearby target genes (TF binding site <10 kb of gene promoter) revealed that eRNA presence is significantly correlated with elevated local gene expression (*P*-value $<10^{-6}$). After making a total of 262 possible pairwise cell type comparisons (95 TFs, four cell types), we noted that 73% of these comparisons display such dynamics (Fig. 1E; Supplemental Table S2). In the same vein, TF binding sites that overlap a region with strong enhancer activity—as measured by a CapStarr-seq enhancer assay (Vanhille et al. 2015)—are five times more likely to associate with eRNAs than regions

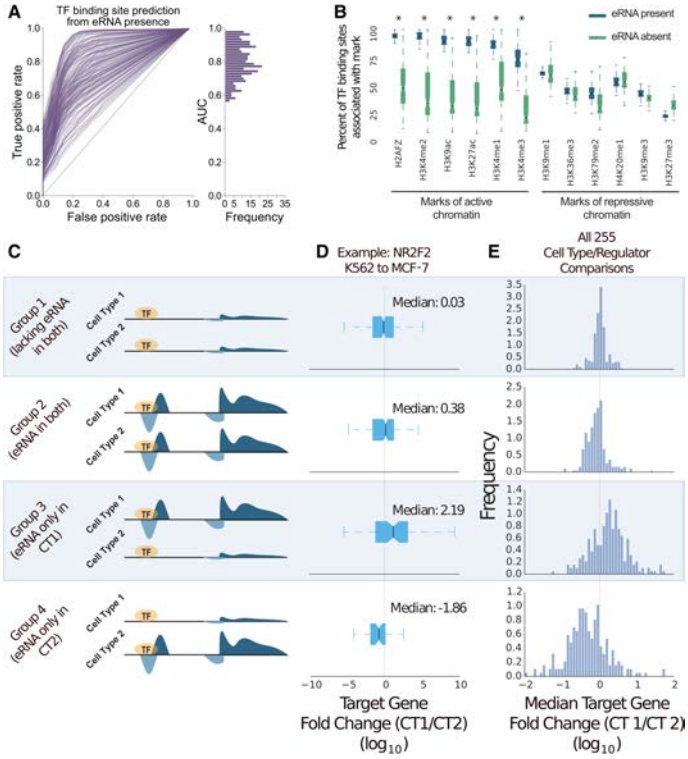


Figure 1. Enhancer RNA (eRNA) presence marks the active subset of TF binding. (*A*) ROC analysis of TF binding site prediction via eRNA presence. False-positive and true-positive rates are varied by thresholding the penalized likelihood ratio statistic generated from Tfit. (*B*) TF binding peaks (Supplemental Table S1) were grouped according to eRNA association. A box-and-whiskers displays the median/variability in proportion of histone mark association between the groups across all TFs (Supplemental Table S1). Asterisks indicate a *P*-value $<10^{-10}$ by *z*-test. All data in *A* and *B* are K562 cells. (*C*) Pairwise cell type–associated TF binding peaks were grouped according to eRNA presence from matched cell types (Supplemental Table S2). A gene was considered "neighboring" by a distance <10 kb. (*D*) Log base 10 FPKM fold change of "neighboring" genes related to eRNA-grouped NR2F2 binding peaks. (*E*) Histogram of Log base 10 FPKM fold change of "neighboring" genes for all possible eRNA-grouped TF ChIP-seq data sets ($n = 255$).

considered inactive by the enhancer assay (P-value $<10^{-19}$, hypergeometric). These results are consistent with a model where eRNA presence discriminates silent from functional TF binding.

eRNA origins colocalize with TF binding motif instances

Given that many TFs bind DNA in a sequence-specific manner, we next sought to determine the precise spatial relationship between instances of the TF-DNA motif model and eRNA transcription. To this end, we measured the distance between genomic instances of the TF motif model and eRNA origins in a K562 GRO-cap data set (Core et al. 2014). We observed a stark colocalization of the motif instance with the eRNA origin specifically in the TF-bound fraction of eRNAs (Supplemental Fig. S7A), suggesting that the motif sequence is present at the precise point of eRNA origination. This led to the speculation that the genome-wide patterns of motif sequence to eRNA co-occurrence could identify the set of active TFs directly regulating eRNA transcription, even when ChIP data are not available.

To investigate this hypothesis systematically requires a measurement of the colocalization of motif instances with eRNA origins. With this in mind, we devised a simple statistic—the motif displacement score (MD-score)—which computes the proportion of TF sequence motif instances within an h-radius of eRNA origins relative to a larger local H-radius (Fig. 2A). Similar to the average length of a nucleosome free region (Yadon et al. 2010), we set the h-radius based on the average estimated distance between the forward and reverse strand transcript peaks at eRNA origins ($h = 150$ bp; Supplemental Fig. S7B) and the H-radius as the average length of chromatin marks associated with active regulatory loci ($H = 1500$ bp; Supplemental Fig. S8). Consistent with the patterns observed in ChIP data, the MD-score is elevated in the bound set of eRNAs relative to the not bound set (Supplemental Fig. S7C).

In order to expand our approach to include TFs for which no ChIP-seq is available, we leveraged a hand-curated database of TF binding motif models (HOCOMOCO, 641 motif models) (Kulakovskiy et al. 2013) and measured the distribution of motif instances proximal to K562 eRNA origins (Fig. 2B). Under a uniform nucleotide background model, 32% of the motif models colocalized significantly with eRNAs (P-value $<10^{-6}$). However, similar to gene promoters and TF binding motifs, enhancers exhibit heightened GC content (Fenouil et al. 2012; The ENCODE

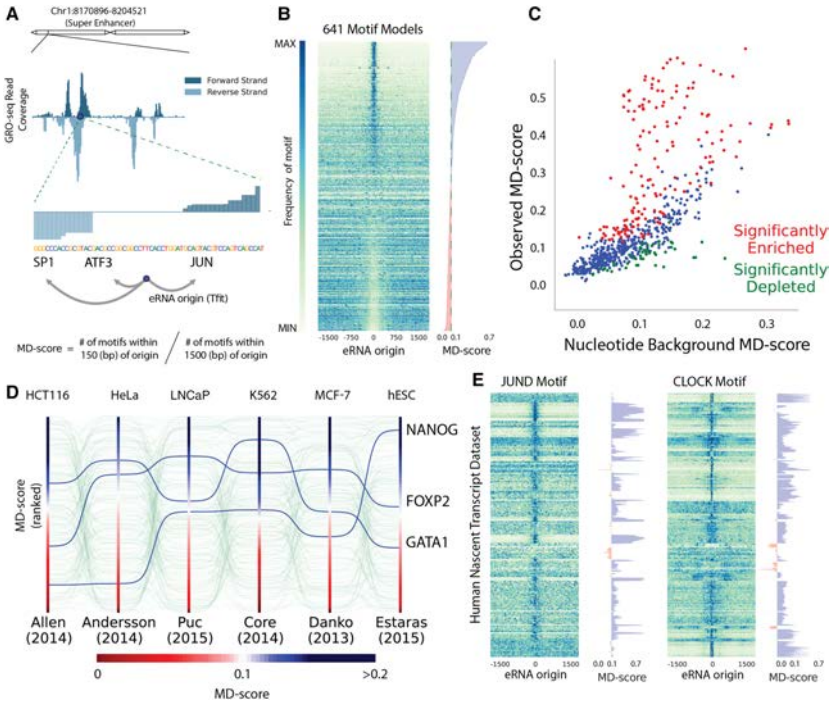


Figure 2. Motif colocalization with eRNA origins varies by cell type. (*A*) An example locus of GRO-seq, the inferred eRNA origin, and computation of "motif displacement" (MD) and the associated MD-score. (*B*) Each row is a TF motif model, and each column is a bin of a histogram (100) where heat is proportional to the frequency of a motif instance at that distance from an eRNA origin. (*C*) A comparison between the expected MD-score for a motif model (*x*-axis) and the observed MD-score in a K562 GRO-cap experiment (Core et al. 2014). Red and green dots indicate a P-value $<10^{-6}$ above or below expectation hypothesis tests, respectively. (*D*) MD-scores were computed and ranked under six nascent transcription data sets. (*E*) Each row corresponds to a nascent data set, and each column relates to motif frequency. These MD distributions are shown for two demonstrative examples (JUND and CLOCK) and the associated MD-scores, sorted by publication.

Project Consortium 2012), which may artificially induce GC-rich motif presence at eRNA origins (Supplemental Fig. S9A). To control for local sequence bias in our colocalization metric, we developed a simulation-based method to perform empirical hypothesis testing of the MD-score (Supplemental Fig. S9B). We observed that—even in light of a significant nucleotide bias—27% of motif models remain significantly colocalized with eRNA origins in the K562 GRO-cap data set (Fig. 2C).

Interestingly, a subset of TFs display significantly lowered MD-scores relative to expectation (green dots in Fig. 2C), suggesting that in these cases, the instances of the motif model are significantly depleted at eRNA origins. Consistent with this observation, a previously published knockout of the Rev-Erb family of transcriptional repressors (*Nr1d1* and *Nr1d2*) resulted in the gain of eRNAs (Lam et al. 2013). Taken together, these results suggest that repressors suppress eRNA activity proximal to their DNA response element.

Significant enrichment or depletion of a motif model near eRNA origins likely indicates that the TF protein is present and functionally active, as either an activator or repressor, respectively. To validate that MD-scores reflect TF activity, we first examined the MD-scores of all motif models across a set of nascent transcription data sets from six distinct cell types. Our analysis revealed wide fluctuations in MD-scores of several motif models across experiments (Fig. 2D). Importantly, we observed that the MD-score associated with cell-type–specific TFs are elevated in their known lineage of activity. For example, NANOG is elevated in embryonic stem cells, consistent with its role in maintaining pluripotency (Mitsui et al. 2003). Additionally, GATA1 is elevated in K562 cells, consistent with its role in leukemia (Shimamoto et al. 1995).

To further evaluate the MD-score, we predicted eRNA origins in a large collection of publicly available nascent transcription data sets (67 publications, 34 cell types and 205 treatments; Supplemental Table S3). Our compendia include a diverse collection of nascent transcription protocols, cell types, sequencing depths, and laboratory of origin. Across the compendium, the spatial relationship between eRNA transcription and motif sequence is exceedingly dynamic (Supplemental Fig. S10), as exemplified by the JUND and CLOCK motif models (Fig. 2E). Given that we observed a modest correlation between sequencing depth and eRNA-identification (Supplemental Fig. S11), we next sought to determine the extent to which the inferred MD-score simply reflected batch effects. To this end, we leveraged the fact that many TFs play a pivotal role in cell fate and identity (Mitsui et al. 2003). Indeed, dimensionality reduction of our MD-score compendium (491 human nascent transcription experiments) revealed statistical influences based predominantly on underlying cell type (Supplemental Figs. S12, S13). Notably, 78% of motif models in HOCOMOCO are significantly colocalized with eRNA origins in at least one data set. While the experimental details clearly influence the ability to infer specific eRNAs, the aggregation of genome-wide signal makes MD-scores relatively robust to experimental variability. Importantly, key cell-type–specific TFs show elevated MD-scores only in the relevant cell type (Fig. 2D), suggesting that MD-scores quantify activity for broad classes of TFs across cell types, despite differences in protocol, sequencing depth, and/or laboratory of origin. Overall, these results indicate that MD-scores fluctuate across cell types and conditions in a manner that suggests changes in TF activity.

As an alternative validation, we examined the transcription patterns of the gene encoding the TF. For many TFs, we observed higher transcription of the TF when the MD-score significantly differed from expectation (Supplemental Fig. S14A). Overall, 45% of TFs show a correlation across all samples between the eRNA inferred MD-score and the transcription level (FPKM) of the gene encoding the TF (Supplemental Fig. S14B), suggesting that some TFs are themselves regulated at transcription. However, the observed correlations were often weak and complex—typically neither linear or monotonic—consistent with the observation that expression levels of a gene are poorly correlated with protein levels (Vogel and Marcotte 2012). Many TFs, including TP53 (Supplemental Fig. S14C), are post-transcriptionally or post-translationally modified to regulate their activity, and therefore, FPKM and MD-scores are not expected to correlate (Oren 1999; Everett et al. 2010).

MD-scores quantify TF activity

To better investigate whether MD-scores reflect TF activity, we turned to experiments where the activity of individual TFs is perturbed (Supplemental Table S4). We reasoned that alterations in TF activity should be detected as significant changes in the MD-score. In previous work, we utilized the drug Nutlin-3a to activate TP53 in HCT116 cells (Allen et al. 2014). Here we observe a significant increase in the colocalization of the TP53 motif sequence and eRNA origins following 1 h of Nutlin-3a exposure (ΔMD-score 0.17, P-value $<10^{-33}$). In fact, of the 641 available TF-motif models, only TP53 and TP63, which have nearly identical motif models, displayed elevated MD-scores following Nutlin-3a treatment (P-value $<10^{-6}$) (Fig. 3A). A number of other studies have specifically activated TFs, including tumor necrosis factor (TNF, also known as TNF-alpha) activation of the NF-κB complex (NFKB1/NFKB2/REL/RELA/RELB) (Luo et al. 2014) and estradiol activation of ESR1 (Hah et al. 2013). In both cases, we observed dramatic shifts in the MD-score for the TF(s) known to be activated by each stimulus (Fig. 3B, C). Despite the fact that treatments involving Nutlin-3a, TNF, and estradiol are known to modulate gene expression (Hah et al. 2013; Allen et al. 2014; Luo et al. 2014), we observed no detectable differences in MD-scores when considering only promoter-associated bidirectional transcript sites (Supplemental Fig. S15). In all three cases (Fig. 3A–C), TF activation resulted in the production of new eRNAs that are uniquely enriched for the relevant motif model, effectively elevating the TF's MD-score (Supplemental Fig. S16).

We next sought to evaluate the robustness of the ΔMD-score approach for inferring altered TF activity. First, differential MD-score analysis between biological replicates revealed no significant shifts in motif sequence to eRNA colocalization, indicating that our false-discovery rate is low (Supplemental Fig. S17). Second, we randomly subsampled reads from the Nutlin-3a experiment to generate data sets with considerably lower depth. With increasingly less depth, fewer eRNAs are detected and the inferred MD-score drops. However, the magnitude of the ΔMD-score remains relatively consistent, indicating that the metric is largely robust to sequencing depth (Supplemental Fig. S18). Finally, we varied the h-radius from 0 to 1500 (the full H-radius) to assess the impact of the h-radius on differential MD-score analysis. We found detectable differences in the MD-score across a broad range of h-radius values, indicating that detection of significant ΔMD-score is robust to the choice of h-radius (Supplemental Fig. S19). Collectively, these results indicate that differential MD-score analysis is a robust method of detecting changes in TF activity.

In each of the aforementioned perturbations, nascent transcription was assessed at a ≤1-h time point. Therefore, we next

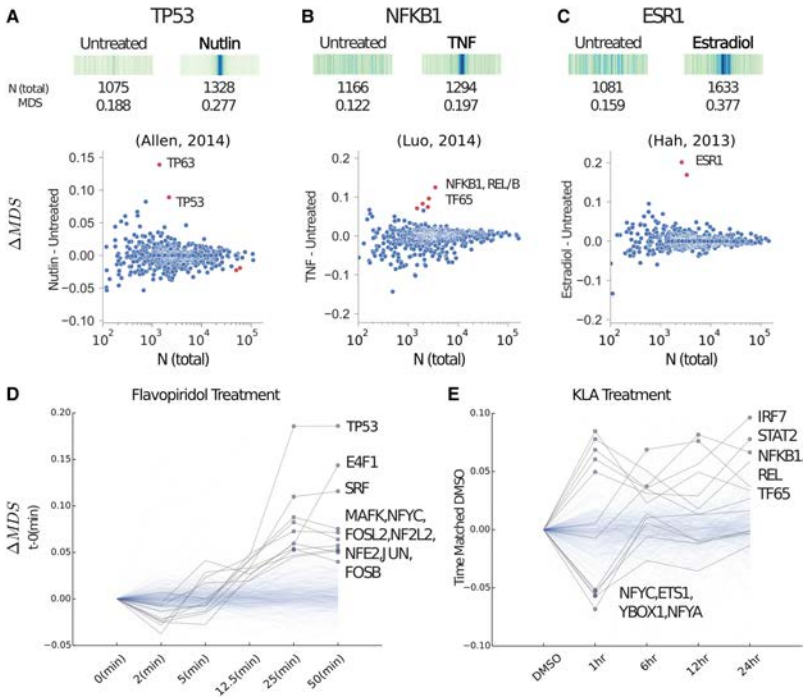


Figure 3. MD-scores predict TF activity. (*A*, *top*) The MD distribution, MD-score, and the number of motifs within 1.5 kb of any eRNA origin before and after stimulation with Nutlin-3a (e.g., Nutlin) on TP53 (Allen et al. 2014), the TF known to be activated. (*Bottom*) For all motif models (each dot), the change in MD-score (Δ*M DS*) following perturbation (*y*-axis) relative to the number of motifs within 1.5 kb of any eRNA origin (*x*-axis). Red points indicate significantly increased and/or decreased MD-scores, respectively (*P*-value $<10^{-6}$). Similar analysis for TNF activation of the NF-κB complex (*B*) (Luo et al. 2014) and estradiol activation of estrogen receptor (ESR1; *C*) (Hah et al. 2013). (*D*) A time series data set following treatment with flavopiridol (Jonkers et al. 2014). The *y*-axis indicates the MD-score change relative to time point zero. Blue dots indicate a MD-score difference $<10^{-6}$. A darker shaded line indicates a time trajectory with at least one significant MD-score. (*E*) Time series data set following treatment with Kdo2-lipid A (KLA) where each time point is normalized to time-matched DMSO (Kaikkonen et al. 2014). Therefore, the *y*-axis indicates MD-score difference relative to the time point–matched DMSO sample. NCBI Sequence Read Archive (SRA) SRR numbers of these comparisons are outlined in Supplemental Table S4.

sought to determine whether MD-scores could capture TF activity across broader time frames. First, we observed that detectable changes in TF activity are exceedingly rapid, as exemplified by flavopiridol (a CDK9 inhibitor)-treated mouse embryonic cells (Laitem et al. 2015), which display a dramatic and monotonic increase in the MD-scores of TP53 and E4F1 (Fig. 3D). For a number of TFs, MD-scores trend upward at 12.5 min and show significant changes within 25 min of exposure. Interestingly, this result indicates that eRNA activity proximal to key TFs increases at short time points, even though flavopiridol is a general repressor of transcription. Mouse T cells treated for a longer time course with Kdo2-lipid A (a highly specific TLR4 agonist) (Kaikkonen et al. 2013) showed dynamic and time-ordered shifts in MD-scores for a number of key TFs (Fig. 3E), including interferon (IRF7) and STAT2. Furthermore, YBOX1 decreases in colocalization (reduced MD-score), consistent with its known role as a transcriptional repressor that increases in expression after KLA exposure (Liu et al. 2009). Collectively, these results indicate that profiles of eRNA transcription—when combined with motif models—identify shifts in TF activity in response to perturbation.

Discussion

We leveraged the observation that eRNAs mark the functional activity of TFs to develop a simple statistic that reflects a TF's functional activity. Importantly, we do not assign TFs to individual enhancers, because most eRNAs have numerous motif instances proximal to their origin. Our approach does not determine which of these possibilities is critical to the regulation of the eRNA. Instead, our statistic, the MD-score, measures the global colocalization of eRNAs with a TF motif model in order to capture changes in TF activity after diverse stimuli.

While the biological functions of eRNAs remain largely unknown, eRNAs clearly represent a powerful readout for TF functional activity. Previous work demonstrated that the presence of eRNAs correlates with active regulatory regions and, consequently, a subset of TF binding sites (Danko et al. 2015). Separately, it has been noted that some binding sites are apparently "silent" with respect to transcription (Cusanovich et al. 2014) or reflect artifacts of ChIP (Teytelman et al. 2013; Worsley Hunt and Wasserman 2014). Therefore, to determine whether eRNAs mark sites of TF

activity, we leveraged binding events across cell lines that differed only in their eRNA activity. Our results indicate that TF binding sites that correspond to eRNA synthesis are more likely to positively affect nearby gene expression than those lacking eRNA transcription. Undoubtedly, assigning enhancers to the nearest gene is not optimal, as many enhancers are known to regulate target genes at great distances (Yao et al. 2015). However, incorrect enhancer to gene assignments would only increase noise within our comparison. Thus, given the instability and short half-lives of eRNAs (Li et al. 2016), their presence within a cell reflects ongoing TF activity.

Consequently, we directly assess TF activity from motif models and nascent transcription. We observe that many motif models show significantly enriched colocalization with eRNA origins beyond expectation, suggesting that these TFs are both present and functionally active in regulation. As the detection of eRNAs is dependent on sequencing depth, future TF-activity inference methods should consider both eRNA-motif colocalization as well as read depth. Even still, we show that TF activity is a strong predictor of cell type, even across distinct protocols, sequencing depths, and laboratory of origin. Hence, our approach has utility in identifying potentially diagnostic signatures of TF activity.

Most importantly, MD-scores can be used to identify when the activity of a TF differs between two data sets, due to either an experimental stimulus or differences in cell type. Our metric utilizes the genome-wide patterns of TF motif sequence colocalization with eRNA origins to identify changes in TF activity, regardless of whether the TF functions as an activator or repressor. Implicitly, changes in MD-score must thus reflect the gain and loss of eRNAs between two conditions, suggesting a direct relationship between functional TF binding and eRNA transcription initiation. However, we and others have observed changes in eRNA transcription levels after stimulus (Hah et al. 2013; Allen et al. 2014), suggesting that our metric could be improved by including changes in the transcription levels of pre-existing eRNAs.

Notably, our differential MD-score approach has some limitations. First, as described, our model considers the influence of each TF on transcription activity independently, yet TFs are often known to work cooperative or in combination (Spitz and Furlong 2012). If two (or more) TFs collaborate to induce eRNA activity and each motif model is enriched over expectation, both would be detected. However, if only the combination is enriched, we would not detect it in our current framework. Second, some families of TFs have similar recognition motifs, making distinguishing between them difficult. In a few cases, one or more family members is not transcribed. For example, upon stimulation with Nutlin-3a, both TP53 and TP63 show significant increases in MD-score (Fig. 3A), but in this cell type (HCT116), only TP53 is transcribed. Thus in this case, we can confidently assert that Nutlin-3a activates TP53. However, in most cases, we will not be able to distinguish family members apart. Finally, we focus here on colocalization of TF motif instances with eRNAs. However, a small set of TFs preferentially bind to promoters (The ENCODE Project Consortium 2012). For these factors, stronger signals may be obtained by computing MD-scores from all sites of polymerase initiation (promoters and enhancers).

In conclusion, we showed that addition of diverse chemical stimuli to cells resulted in activation or deactivation of specific TFs. It is compelling to think that had we not known the nature of each stimulus, we could have inferred their effects from the unique eRNA profile obtained immediately after addition of the compound. As methods for measuring eRNA production become simpler and cheaper, our approach could eventually serve as a screen capable of discriminating between the direct mechanistic impact of closely related compounds and, hence, serve as another layer of information about the effects of a drug. Such data could help to define previously poorly understood molecular mechanisms underlying a drug's activity.

Methods

Public data sets

We examine the relationship (association and/or overlap) between genomic features such as TF binding peaks, chromatin modifications, DNA sequence, TF binding motif models, and eRNA presence. Data for all features were obtained from publicly available sources and compared relative to a human and mouse genome versions hg19 and mm10, respectively. Human and mouse nascent transcription data were obtained from the NCBI Gene Expression Ombnibus (Supplemental Table S3). ENCODE peak data were obtained from https://www.encodeproject.org/matrix/?type=Experiment. Most data were provided relative to hg19, but when necessary, ENCODE files were converted to hg19 via the Python LiftOver package. Accession numbers for all ENCODE data utilized are provided in Supplemental Table S1. Motif models were obtained from the HOCOMOCO v. 10 (Kulakovskiy et al. 2013, 2016) database and scanned against the genome. For complete details on the processing and remapping of these data sets, refer to the Supplemental Methods.

Tfit modification and parameters

In prior work (Azofeifa and Dowell 2017), we leveraged the known behavior of RNAP to identify individual transcripts within nascent transcription data. Our model (Azofeifa and Dowell 2017), known as transcription fit (Tfit), infers the precise point of RNA polymerase loading, e.g., the origin point of transcription. Formally, this origin point (μ) represents the expected value of a Gaussian (normal) random variable, discussed in great detail in our previous publication (Azofeifa and Dowell 2017).

For analysis of numerous nascent data sets, here we modify our previous approach in two ways. First, to rapidly identify all sites of transcription initiation genome-wide, we compute a likelihood ratio statistic between a fully specified exponentially modified Gaussian (Equation 1, the loading/initiation/pausing phase of our earlier Tfit model) (Azofeifa and Dowell 2017) against a uniform distribution background model (Equation 2) at some genome interval $[a, b]$. We hereafter refer to this approach as template matching. Second, we amend our earlier estimate of the loading step of polymerase activity to permit variable distances between the forward and reverse strand transcripts, hereafter referred to as a polymerase footprint. For completeness, we now describe both modifications in full detail below. We then validated the modified Tfit by comparison of predictions to histone marks and TF binding data (for full description of validation, see Supplemental Methods).

Template matching

The loading/initiation/pausing portion of our earlier model, fully specified in Azofeifa (Azofeifa and Dowell 2017), describes the initial activity of RNAP and captures initiating transcription, which is often bidirectional, genome-wide. Briefly, our model assumes RNAP is first recruited and binds to some genomic coordinate X as a Gaussian-distributed random variable with parameters μ, σ^2, where μ might represent the typical loading position (e.g., origin of any resulting transcript either TSS or enhancer locus) and σ^2 the amount of error in recruitment to μ. Upon recruitment,

RNAP selects and binds to either the forward or reverse strand, which we characterize as a Bernoulli random variable S with parameter π. Following loading and preinitiation, RNAP immediately escapes the promoter and transcribes a short distance, *Y*. We assume that the initiation distance is distributed as an exponential random variable with rate parameter λ. In this way, the final genomic position *Z* of RNAP is a sum of two independent random variables ($X + SY$), where the density function (resulting from the convolution/cross-correlation) is given in Equation 1. Note that, in keeping with traditional notation, we let uppercase, non-Greek alphabet letters represent random variables and the associated lowercase letters refer to instances or observations of the stochastic process.

$$h(z, s; \mu, \sigma, \lambda, \pi) = \lambda\phi\left(\frac{z-\mu}{\sigma}\right)R\left(\lambda\sigma - s\frac{z-\mu}{\sigma}\right)1(s)$$
$$1(s) = \begin{cases} \pi & : s = +1 \\ 1-\pi & : s = -1 \end{cases}. \quad (1)$$

Above, $\phi(.)$ refers to the standard normal density function and $R(.)$ refers to the Mill's ratio.

In contrast, reads obtained outside of initiation regions are captured by a uniform distribution (Equation 2).

$$u(z; a, b) = \frac{\hat{\pi}}{b-a}, \quad (2)$$

where $\hat{\pi}$ refers to the maximum likelihood estimator for the strand bias (Equation 3).

$$\hat{\pi} = \sum_{i=1}^{N} I(s_i > 0)/N, \quad (3)$$

where $I(.)$ is an indicator function. Finally, the (log-)likelihood of the exponentially modified Gaussian (LL_{emg}) and uniform (L_u) distribution computed at a genomic interval $[a,b]$ using aligned read counts is given in Equation 4.

$$LL_{emg} = \sum_{i=a}^{b} \log\ h(z_i, s_i; \hat{\mu}, \hat{\sigma}, \widehat{1/\lambda}, \hat{\pi}),$$
$$LL_u = \sum_{i=a}^{b} I(s_i > 0)\log\ \frac{\hat{\pi}}{b-a} + I(s_i < 0)\log\ \frac{1-\hat{\pi}}{b-a}, \quad (4)$$
$$LLR = LL_{emg} - LL_u.$$

Here, $\hat{\mu}$ refers to the center of the window. Based on our previous study (Azofeifa and Dowell 2017), we set $\{\hat{\sigma}, \widehat{1/\lambda}, \hat{w}, \hat{\pi}\} = \{34.2, 391.7, 0.358, 0.501\}$.

The algorithm is a simple sliding window of LLR computations. Overlapping (1-bp) regions of interest (LLR > τ) are merged. In every study profiled for bidirectional transcription by Tfit, $\tau = 10^3$. More information on running and using Tfit output is available at https://biof-git.colorado.edu/dowelllab/Tfit.

EM algorithm and bidirectional origin estimation

On its own, however, the template matching module of Tfit does not provide an exact estimate over Θ (the parameters associated with a single loading position). To perform optimization over Θ and specifically μ (the origin of bidirectional transcription), we derived the expectation maximization algorithm (outlined in detail in our previous publication) (Azofeifa and Dowell 2017) to optimize the likelihood function of Equation 4. In brief, we used the following EM-specific parameters at each loci: The number of random reinitializations per loci was set to 64, the threshold at which the EM was said to converge, $|ll_t - ll_{t+1}|$, was set to 10^{-5}. Finally for computational tractability, the EM algorithm halted after maximum of 5000 iterations.

At each window predicted by the sliding window algorithm, we perform inference over μ, σ, λ, and π by the EM algorithm. Details of the derivation, model selection, and algorithm design can be found in our previous report (Azofeifa and Dowell 2017).

Footprint estimation

Importantly, our previous effort at parameter estimation of the finite mixture model assumed that RNAP behaved as a point source (Azofeifa and Dowell 2017). Consequently, we could not incorporate a systematic approach to estimate observed gaps between the forward and reverse strand peaks, which deviate more than could be explained by an exponentially modified Gaussian density function. Here, we amend our earlier model only slightly to estimate this behavior. We call the distance between the forward and reverse strand peaks, the *footprint* of RNAP or *fp*. In brief, *fp* amounts to adding or removing a constant to z_i, the genomic position of RNAP after loading and initiation. Assuming that $fp > 0$ then the above equations remain valid by a simple transformation to z_i:

$$z_i := z_i - s_i \cdot fp.$$

As in our previous effort (Azofeifa and Dowell 2017), we insert this new parameter into the conditional expectation of the latent variables given the observed random variables and perform a gradient step. This allows us to optimize for *fp* (Equation 5):

$$\widehat{fp}_k := \frac{1}{r_k}\sum_{i=1}^{N}(s_i(z_i - \mu) - E[Y|z_i, s_i; \theta^g]) \cdot r_i^k. \quad (5)$$

The interested reader should refer to our previous paper (Azofeifa and Dowell 2017) where each parameter is explained fully; derivation of the EM algorithm and fitting of the Tfit model are discussed heavily. For complete clarity, the full expression of the expectation operators is given by Equation 6:

$$E[Y|g_i; \theta^t] = s_i(z-\mu) - \lambda\sigma^2 + \frac{\sigma}{R(\lambda\sigma - s_i(z_i-\mu)/\sigma)},$$
$$r_i^k = p(k|g_i; \theta_k^g) = \frac{w_k \cdot p(g_i; \theta_k^g)}{\sum_{k\in\boldsymbol{K}} w_k \cdot p(g_i; \theta_k^g)}, \quad (6)$$
$$r_k = \sum_{i=1}^{N} r_i^k.$$

TF binding site prediction via eRNA presence

We compute the receiver operating characteristic (ROC) curve to quantify the ability of bidirectional transcription to predict TF ChIP binding. ENCODE-called peaks within a TF's ChIP-seq data are considered truth, and randomly selected regions that do not overlap any previously seen ChIP-seq peak are considered a gold standard for noise. For each peak (truth or noise), a bidirectional model is fit using the expectation maximization algorithm. A Bayesian information criteria (BIC) score was calculated between the exponentially modified Gaussian mixture model and a simple uniform distribution with support across the entire peak. We record a true positive if the BIC score exceeds a threshold τ and the peak was one of the ENCODE peak calls. We record a false positive if the BIC score exceeds the threshold (τ) and the peak is a random noise interval. We vary the threshold τ to obtain the ROC curve of Figure 1 and compute an area under the curve (AUC).

Computation of bimodality

To assess whether the distribution of ChIP peaks or TF binding motif sequences around an eRNA origin is bimodal, we developed and employed a pairwise distribution test. We define the ΔBIC score (in Equation 8) to be the difference in BIC scores between a single Laplace-uniform mixture centered at zero (unimodal) and a two component Laplace-uniform mixture with displacement away from 0, i.e., c (bimodal). The density function of a Laplace distribution with parameters (c,b) is provided in Equation 7, and we use the formulation for the uniform distribution of Equation 2.

$$p(d; c, b) = \frac{1}{2b}\exp\ -\frac{|d-c|}{b}. \tag{7}$$

Here D refers to the set of distances, $d_i \in [-1500, 1500]$, either the center of the TF binding peaks obtained from MACS (Zhang et al. 2008) or the center of TF binding motif sequence from the PSSM scanner relative to eRNA origin. If $\Delta BIC \gg 0$, we assume bimodality in TF peak location relative to the eRNA origin:

$$\begin{aligned}
\mathcal{L}_0(D;\Theta^*) &= \prod_{i=1}^{N}\frac{1}{3000},\\
\mathcal{L}_1(\mathrm{D};\Theta^*) &= \prod_{i=1}^{N} w\frac{1}{2b}\exp\ \left\{-\frac{|d_i|}{b}\right\} + \frac{1-w}{3000},\\
\mathcal{L}_2(\mathrm{D};\Theta^*) &= \prod_{i=1}^{N} \frac{w}{4b}\exp\left\{-\frac{|d_i-c|}{b}\right\} + \frac{w}{4b}\exp\ \left\{-\frac{|d_i+\mu|}{b}\right\} + \frac{1-w}{3000}.\\
\Delta\mathrm{BIC} &:= -2(\log\mathcal{L}(D)_1 - \log\mathcal{L}(D)_2) + k\log\ (|D|).
\end{aligned} \tag{8}$$

Θ^* is optimized again by the Expectation Maximization algorithm where the update rules are given in Equation 9:

$$\begin{aligned}
d_{t+1} &= \frac{1}{2(r^a + r^b)}\left(\sum_{i=1}^{n} r_i^a d_i + \sum_{i=1}^{n} r_i^a d_i\right),\\
b_{t+1} &= \frac{1}{2(r^a + r^b)}\left(\sum_{i=1}^{n} r_i^a |d_i| + \sum_{i=1}^{n} r_i^b |d_i|\right),\\
w_{t+1} &= \frac{r^a + r^b}{r},\\
r_i^a &= \frac{p(d_i; c, b)}{p(d_i; c, b) + p(d_i; -c, b) + u(d_i; -1500, 1500)},\\
r_i^b &= \frac{p(d_i; -c, b)}{p(d_i; c, b) + p(d_i; -c, b) + u(d_i; -1500, 1500)},\\
r_i^u &= 1 - r_i^a + r_i^b \quad r^x = \sum_{i=1}^{N} r_i^x \quad r = r^a + r^b + r^u.
\end{aligned} \tag{9}$$

We refer to a signal as bimodal (i.e., not unimodal) when $\Delta BIC > 500$, estimated from the distribution in Supplemental Figure S5D.

MD-score hypothesis testing

The MD-score relates the proportion of significant motif instances within some window $2h$ divided by the total number of motif instances against some larger window $2H$ centered at all bidirectional origin events. It is calculated on a per PWM binding model basis.

Let $X_j = \{x_1, x_2, \ldots\}$ be the set of bidirectional origin locations genome-wide for some experiment j. Let $Y_i = \{y_1, y_2, \ldots\}$ be the set of all significant motif instances for some TF-DNA binding motif model i genome-wide, which is static as it only depends on the genome build of interest. Furthermore, because recent human genome builds vary little at the sequence level, the metric is not expected to change significantly between hg19 versus GRCh38. Therefore, the set of all MD-scores is calculated by Equation 10:

$$\begin{aligned}
g(X_j, Y_i; a) &= \sum_{x\in X_j}\sum_{y\in Y_i}\delta(|x-y| < a),\\
md_{j,i} &= g(X_j, Y_i; h)/g(X_j, Y_i; H),\\
md_{j,i} &\in [0, 1) \quad \text{if } h < H.
\end{aligned} \tag{10}$$

Here, δ(.) is a simple indicator function that returns one if the condition (.) evaluates true and zero if false. The double sum, i.e., $g(a)$, is naively $O(|X||Y|)$; however, data structures like interval trees reduce time to $O(|X|\log|Y|)$.

To be clear, there exist 641 TF-DNA binding models in the HOCOMOCO database, and therefore, 641 MD-scores exist for some experiment j. Let md_i be the MD-score computed for some TF-DNA binding motif model. Therefore, let $MD_j = \{md_1, md_2, \ldots, md_{641}\}$ be the vector of all MD-scores for some data set j.

MD-score significance under stationary model

If y_i and x_i are uniformly distributed throughout the genome, i.e., following a homogeneous Poisson point process, then $g(h)$ is distributed as a binomial distribution with parameters p,N (Equation 11):

$$\begin{aligned}
g(h) &\sim \mathrm{B}(n, p),\\
\mathrm{B}(k; n, p) &= \binom{n}{k}(p)^k(1-p)^{n-k},\\
\text{where} \quad n &= G(H) \quad \text{and } p = h/H.
\end{aligned} \tag{11}$$

In cases where $g(H) \gg 0$, the binomial is well approximated by a Gaussian distribution, and hypothesis testing under some α level can proceed in the typical fashion. In brief, significantly increased MD-scores (by a binomial test) is diagnostic of heightened motif frequency surrounding eRNA origins.

MD-score significance under a nonstationary background model

Motif instances, however, are not distributed uniformly throughout the genome. Specifically, particular regions, such as gene promoters of the genome, are known to exhibit significance sequence bias. Indeed, the localized GC content is highly nonstationary at eRNAs (Supplemental Fig. S9A). Consequently, a binomial test, which assumes a homogeneous Poisson process of motif locations genome-wide, may be a too liberal null model (e.g., the wrong background assumption).

To control for this nonstationarity, we propose a simulation-based method to compute P-values for MD-scores under an empirical CDF, i.e., a localized background model. Let p be a $4x2H$ matrix where each column corresponds to a position from an origin and each row corresponds to a probability distribution over the DNA alphabet $\{A,C,G,T\}$. To be clear, $p_{0,0}$ corresponds to the probability of an A at position $-H$ from any bidirectional origin, similarly $p_{2,1500}$ corresponds to the probability that a G occurs at exactly the point of the bidirectional origin.

Therefore, the simulation-based method of the background model is simple. Given an experiment of X_j bidirectional origin locations, we simulate $|X_j|$ sequences following this nonstationary GC content bias. We then iterate over all PWM models and look for significant motif hits. We then compute summary statistics about the displacement of the motif sequence relative to the set of synthetic sequences, i.e., $MD = \{md_1, md_2, \ldots, md_{641}\}$. It should be noted that, in this data set, any motif model match is by complete chance alone. We iterate this process 10,000 times to compute a random distribution over md_i, i.e., $\overline{md_i}$, and thus we can

assess the probability of our observed (i.e., from real data) md_i relative to our empirically simulated $\overrightarrow{md_i}$. Example simulations are shown in Supplemental Figure S9B.

Cell type and TF enrichment analysis

This section serves to outline the rational for determining if heightened MD-scores correlate with a specific cell type category. More traditional approaches such as a one-way ANOVA test (MD-scores computed from similar cell types are grouped and within group variance is assessed via a F-distribution) will not adequately account for MD-scores with little support (i.e., motif hits that overlap very few eRNAs). To overcome this, we propose a relatively straightforward method that relies on performing hypothesis testing on all pairwise experimental comparisons.

Let j and k be two nascent transcription data sets of interest, then $mds_{j,i}$ and $mds_{k,i}$ refer to MD-scores for some TF-motif model (i) for which we can perform hypothesis testing over as outlined in MD-Score Hypothesis Testing. If we let α be the threshold at which we consider $mds_{j,i} - mds_{k,i}$ to significantly increase, then we expect on average $\alpha \cdot N - 1$ false positives when considering a single experiment against the rest of the corpus of size N.

Put another way, if we let the random variable $S_{j,i}$ refer to the number of times we consider $mds_{j,i} - mds_{k,i}$ to significantly increase in a data set comparison, then $S_{j,i}$ is binomial distributed with parameters $N - 1$ and α (Equation 12), assuming that there is not a relationship between the motif model i and the experiment j:

$$S_{j,i} = \sum_{k=1}^{N} \mathbb{I}(\, p(mds_{j,i} > mds_{k,i}) < \alpha). \tag{12}$$

In practice we set α to 10^{-6}, and $\mathbb{I}$ refers to an indicator function that returns one in the case where the statement evaluates to truth, otherwise zero.

Naively, we could now ask for all the data sets annotated as some cell type ct and then perform hypothesis testing on S_{ct} (the sum of $S_{j,i}$'s where experiment j belongs to the ct cell type set). Importantly, we only consider data set pairs for which i and j belong to different cell type sets. Unfortunately, a single experiment within the cell type set might show strong association with a TF (i.e., 90% of the $N - 1$ comparisons significantly deviate from zero) where the rest of the cell types show small numbers of significant deviations. By a binomial test, this is unlikely—even when considering the expansion induced by the cell type set—but intuitively does not fit into our notion of cell type association.

To this end, we define a final random variable $A_{ct,i}$ to be the number of times motif model i is significantly enriched for a data set j and that data set j belongs to some cell type (Equation 13):

$$A = \sum_{j=1}^{N} p(S_{j,i} > S) < 10^{-6}\mathrm{I}(j \in \mathrm{CT}), \tag{13}$$

where CT refers to the set of experiments that are annotated as cell type ct. From there, it is easy to assess A across cell types and motif models under a contingency model using Fisher's exact test.

Transcription of the TF gene when the MD-score is elevated or depleted

To evaluate whether significantly altered (elevated or depleted) MD-scores reflect TF activity, we first calculate the nascent transcription levels over the gene encoding the TF. To this end, all RefSeq genes were downloaded from hg19. Samples with fewer than 5000 Tfit bidirectional regions were removed from subsequent consideration. FPKM was calculated for each gene in each human nascent transcription sample ($n = 491$) over the body of the gene, defined here as 1 kb to the end of the gene. For all TFs in HOCOMOCO >1 kb and with a RefSeq name ($n = 635$ TFs), the maximum FPKM of all annotated isoforms was utilized. All TF MD-scores were compared to expectation and classified on a per sample basis. Significant deviations from expectation were determined as passing both the stationary and nonstationary test (P-value $<10^{-6}$). TFs with significant deviation were subsequently labeled as elevated if they had a minimum MD-score of 0.1 and were above expectation or labeled as depleted if they had a maximum MD-score of 0.1 and below expectation. To identify samples in which the TF is at expectation, we labeled a third set as at-expectation if they pass the stationary and nonstationary test (P-value $<10^{-2}$). For the box plots of Supplemental Figure S14A, we excluded samples with fewer than 10 significant (depleted or elevated) or at-expectation samples. Across all samples, to avoid zero FPKM the minimum nonzero FPKM was utilized.

We next calculated the Spearman's rank correlation coefficient and P-value across all samples ($n = 491$; scipy v0.17.1) between MD-scores and the FPKM of the gene encoding the TF (Supplemental Fig. S14B). When shuffling the FPKMs across samples, we expect an average of 8.4 TFs to show correlation (permutation testing 100 times, standard deviation 2.4 TFs). For all eRNAs (MD-score from nonpromoter associated bidirectionals), 286 of 635 TFs show a correlation (P-value <0.01). For all bidirectionals (includes promoters), the same P-value cutoff finds 441 of 635 TFs with correlation (expectation 16.5, standard deviation 3.8).

We next examined regions evaluated by a functional assay, namely, CapStarr-seq (Vanhille et al. 2015), for their co-occurance with eRNA origins. In CapStarr-seq, they utilized mouse 3T3 cells, selected TF-bound regions (by ChIP), and determined whether the bound regions functioned as an enhancer using a GFP expression assay. Identified regions were moved to mm10 coordinates using LiftOver (Hinrichs et al. 2006). For comparison to nascent transcription, Tfit-called bidirectionals (both eRNA and promoter origins) for mouse samples (SRR1233867, SRR1233868, SRR1233869, SRR1233870, SRR1233871, SRR1233872, SRR1233873, SRR1233874, SRR1233875, SRR1233876) from the 3T3 cell lines were combined (Step et al. 2014). While 35.5% of regions classified as a strong enhancer ($n = 186$) by CapStarr-seq contained a bidirectional origin, only 7.9% of regions classified inactive ($n = 4406$) had a bidirectional origin. Generally, bidirectionals within strong enhancers (by CapStarr-seq) were identified by Tfit in multiple nascent transcription replicates, while bidirectionals within inactive regions were only in one nascent transcription replicate. Overall, regions defined as strong enhancers were four times more likely to contain an eRNA origin than regions defined as inactive enhancers.

MD-score significance between experiments

The MD-score constitutes a proportion, and as long as h is upper-bounded by H, then $md_{j,i}$ will always exist within the semi-open interval [0,1). An important question is whether $md_{j,i}$ has significantly shifted between two experiments: j,k as a function of X_j and X_k. This analysis is straightforward under the two proportion z-test. Specifically, we are testing the null and alternative hypothesis tests in Equation 14:

$$\begin{aligned} H_0 &: md_{j,i} = md_{k,i}, \\ H_1 &: md_{j,i} \neq md_{k,i}. \end{aligned} \tag{14}$$

We can then compute the pooled sample proportion (p_i) and standard error (SE) as shown in Equation 15. Therefore, our test statistic z (Equation 16) is normally distributed with mean 0 and

variance 1:

$$p_i = \frac{(md_{j.i} \cdot g(X_j, Y_i; H) + md_{k.i} \cdot g(X_k, Y_i; H))}{g(X_j, Y_i; H) + g(X_k, Y_i; H)}, \quad (15)$$

$$SE = p(1-p) \cdot (1/g(X_j, Y_i; H) + 1/g(X_k, Y_i; H)),$$

$$z = \frac{md_{j.i} - md_{k.i}}{\sqrt{SE}} \sim N(0, 1). \quad (16)$$

Computation of the *P*-value can be assessed in the normal fashion under some α level. In all comparisons, we utilize multiple hypothesis correction outlined by Storey et al. (2007).

Acknowledgments

This work was funded in part by a National Science Foundation (NSF) IGERT grant number 1144807 (J.G.A., R.D.D.), a National Institutes of Health (NIH) grant T32 GM008759 (J.D.R.), a Sie post-doctoral fellowship (M.A.A.), the Boettcher Foundation's Webb-Waring Biomedical Research program (R.D.D.), and an NSF ABI DBI-12624L0 (R.D.D.). We acknowledge the BioFrontiers Computing Core at the University of Colorado Boulder for providing high-performance computing resources (NIH 1S10OD012300) supported by BioFrontiers' IT.

References

Allen MA, Mellert H, Dengler V, Andryzik Z, Guarnieri A, Freeman JA, Luo X, Kraus WL, Dowell RD, Espinosa JM. 2014. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *eLife* **3:** e02200. doi: 10.7554/eLife.02200.

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* **507:** 455–461.

Azofeifa JG, Dowell RD. 2017. A generative model for the behavior of RNA polymerase. *Bioinformatics* **33:** 227–234.

Balwierz PJ, Pachkov M, Arnold P, Gruber AJ, Zavolan M, van Nimwegen E. 2014. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res* **24:** 869–884.

Core L, Lis J. 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319:** 1791–1792.

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.

Cusanovich DA, Pavlovic B, Pritchard JK, Gilad Y. 2014. The functional consequences of variation in transcription factor binding. *PLoS Genet* **10:** e1004226.

Danko C, Hah N, Luo X, Martins A, Core L, Lis J, Siepel A, Kraus W. 2013. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol Cell* **50:** 212–222.

Danko CG, Hyland SL, Core LJ, Martins AL, Waters CT, Lee HW, Cheung VG, Kraus WL, Lis JT, Siepel A. 2015. Identification of active transcriptional regulatory elements from GRO-seq data. *Nat Methods* **12:** 433–438.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.

Everett L, Hansen M, Hannenhalli S. 2010. *Regulating the regulators: modulators of transcription factor activity*, pp. 297–312. Humana Press, Totowa, NJ.

The FANTOM Consortium and Riken Omics Science Center. 2009. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* **41:** 553–562.

Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, Innocenti C, Ferrier P, Spicuglia S, Gut M, Gut I, et al. 2012. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res* **22:** 2399–2408.

Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB, et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc Natl Acad Sci* **109:** 21330–21335.

Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. 2013. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* **23:** 1210–1223.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459:** 108–112.

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, et al. 2006. The UCSC genome browser database: update 2006. *Nucleic Acids Res* **34:** D590–D598.

Jonkers I, Kwak H, Lis JT. 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3:** e02407. doi: 10.7554/eLife.02407.

Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, et al. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell* **51:** 310–325.

Kaikkonen MU, Niskanen H, Romanoski CE, Kansanen E, Kivelä AM, Laitalainen J, Heinz S, Benner C, Glass CK, Ylä-Herttuala S. 2014. Control of VEGF-A transcriptional programs by pausing and genomic compartmentalization. *Nucleic Acids Res* **42:** 12570–12584.

Kulakovskiy IV, Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, Makeev VJ. 2013. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res* **41:** D195–D202.

Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, Ba-Alawi W, Bajic VB, Medvedeva YA, Kolpakov FA, et al. 2016. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res* **44:** D116–D125.

Laitem C, Zaborowska J, Isa NF, Kufs J, Dienstbier M, Murphy S. 2015. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II–transcribed genes. *Nat Struct Mol Biol* **22:** 396–403.

Lam MTY, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M, et al. 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498:** 511–515.

Li Xy, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Hendriks CLL, et al. 2008. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* **6:** e27. doi: 10.1371/journal.pbio.0060027.

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, et al. 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498:** 516–520.

Li W, Notani D, Rosenfeld MG. 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17:** 207–223.

Liu X, Kelm RJ, Strauch AR. 2009. Transforming growth factor β1-mediated activation of the smooth muscle α-actin gene in human pulmonary myofibroblasts is inhibited by tumor necrosis factor-α via mitogen-activated protein kinase kinase 1-dependent induction of the Egr-1 transcriptional repressor. *Mol Biol Cell* **20:** 2174–2185.

Luo X, Chae M, Krishnakumar R, Danko CG, Kraus WL. 2014. Dynamic reorganization of the AC16 cardiomyocyte transcriptome in response to TNFα signaling revealed by integrated genomic analyses. *BMC Genomics* **15:** 155.

Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113:** 631–642.

Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161:** 526–540.

Oren M. 1999. Regulation of the p53 tumor suppressor protein. *J Biol Chem* **274:** 36031–36034.

Puc J, Kozbial P, Li W, Tan Y, Liu Z, Suter T, Ohgi KA, Zhang J, Aggarwal AK, Rosenfeld MG. 2015. Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell* **160:** 367–380.

Rackham OJL, Firas J, Fang H, Oates ME, Holmes ML, Knaupp AS, FANTOM Consortium, Suzuki H, Nefzger CM, Daub CO, et al.. 2016. A predictive computational framework for direct reprogramming between human cell types. *Nat Genet* **48:** 331–335.

Read T, Richmond PA, Dowell RD. 2016. A *trans*-acting variant within the transcription factor RIM101 interacts with genetic background to determine its regulatory capacity. *PLoS Genet* **12:** e1005746. doi: 10.1371/journal.pgen.1005746.

Savic D, Roberts BS, Carleton JB, Partridge EC, White MA, Cohen BA, Cooper GM, Gertz J, Myers RM. 2015. Promoter-distal RNA polymerase II binding discriminates active from inactive CCAAT/enhancer-binding protein beta binding sites. *Genome Res* **25:** 1791–1800.

Shimamoto T, Ohyashiki K, Ohyashiki J, Kawakubo K, Fujimura T, Iwama H, Nakazawa S, Toyama K. 1995. The expression pattern of erythrocyte/megakaryocyte-related transcription factors GATA-1 and the stem cell leukemia gene correlates with hematopoietic differentiation and is associated with outcome of acute myeloid leukemia. *Blood* **86:** 3173–3180.

Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA. 2015. Transcription factor trapping by RNA in gene regulatory elements. *Science* **350:** 978–981.

Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13:** 613–626.

Step SE, Lim HW, Marinis JM, Prokesch A, Steger DJ, You SH, Won KJ, Lazar MA. 2014. Anti-diabetic rosiglitazone remodels the adipocyte transcriptome by redistributing transcription to PPARγ-driven enhancers. *Genes Dev* **28:** 1018–1028.

Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. 2007. Gene-expression variation within and among human populations. *Am J Hum Genet* **80:** 502–509.

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

Teytelman L, Thurtle DM, Rine J, van Oudenaarden A. 2013. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci* **110:** 18602–18607.

Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LT, Fernandez N, Ballester B, Andrau JC, Spicuglia S. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* **6:** 6905. doi: 10.1038/ncomms7905.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10:** 252–263.

Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13:** 227–232.

Worsley Hunt R, Wasserman WW. 2014. Non-targeted transcription factors motifs are a systemic component of ChIP-seq datasets. *Genome Biol* **15:** 412.

Yadon AN, Van de Mark D, Basom R, Delrow J, Whitehouse I, Tsukiyama T. 2010. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. *Mol Cell Biol* **30:** 5110–5122.

Yao L, Berman BP, Farnham PJ. 2015. Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. *Crit Rev Biochem Mol Biol* **50:** 550–573.

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

Received May 30, 2017; accepted in revised form January 24, 2018.

Enhancer RNA profiling predicts transcription factor activity

Joseph G. Azofeifa, Mary A. Allen, Josephina R. Hendrix, et al.

Genome Res. published online February 15, 2018
Access the most recent version at doi:10.1101/gr.225755.117

Chapter 3

Transcription Factor Enrichment Analysis

3.1 Preamble

While the MD-Score was a big step not only as a method development paper but also as an eRNA characterization paper, there were several aspects of this technique that we sought to improve. First, the MD-Score relied on a binary metric for measuring TF activity. Across two conditions, the MD-Score created histograms of motif overlaps to detected bidirectionals and quantified separately these histograms for each condition. This meant that the MD-Score was blind to events where the transcriptional levels of bidirectionals were changing. Second, the MD-Score had no way of combining replicates which significantly limited its statistical power. We addressed both these issues by designing a metric with inspiration from the popular GSEA technique for measuring enrichment of gene sets.

3.1.1 Significance

As discussed in Chapter 2, measuring TF activity is a global readout of cellular function and current experimental techniques for measuring TF activity are low-throughput. In the previous chapter, I discussed the development of the MD-Score as a computational approach for measuring TF activity from sites of bidirectional transcription overlapped with TF motifs. Because the MD-Score used a binary metric and did not incorporate replicates, we developed Transcription Factor Enrichment Analysis (TFEA) based on the popular gene set enrichment analysis (GSEA). In addition to solving the outstanding issues with the MD-Score approach, TFEA also eliminated

the need for some thresholds and is essentially an asymmetry pattern recognizer. In contrast to the MD-Score, TFEA requires a set of conditions in which to perform its analysis whereas the MD-Score could detect TF activity in a single dataset.

While built originally for nascent transcription, TFEA has proven to be more broadly applicable. TFEA essentially is a motif enrichment strategy that balances both relative changes in transcription (or whatever the data measures) and position of the motif from a designated point of interest. Briefly, TFEA works by ranking regions of interest (ROI) by some metric (usually read depth), scanning for motifs over ROI, calculating an enrichment curve based on the motif positions relative to ROI centers, and quantifying the area between this curve and random expectation (the diagonal). This value is called the E-Score and represents the percentage enrichment of a TF motif in input ROIs across two conditions. Significance of the E-Score is calculated empirically, by shuffling the rank order of ROI and recomputing a simulated E-Score 1000 times.

In the work that follows, TFEA is first compared to the MD-Score and to the differential MD-Score (MDD-Score). The MDD-Score is a method developed by Margaret Gruca that sought to incorporate differential signal into the MD-Score approach. TFEA is then benchmarked with AME, a motif enrichment algorithm developed as part of the MEME suite. Finally, we use TFEA to analyze publicly available time series datasets and show that TFEA can temporally unravel complex regulatory networks. Additionally, we show that TFEA works on datasets other than nascent such as CAGE, H3K27ac ChIP, EP300 ChIP, DNAse-Seq, and ATAC-Seq.

The current TFEA manuscript is under revision at Nature Communications but available as a preprint on BioRxiv (doi: https://doi.org/10.1101/2020.01.25.919738).

3.1.2 Contributions

The TFEA project was a collaborative effort over the course of several years. The majority of this work, which included method development and benchmarking was done by myself with Dr. Stanley contributing key aspects to method development, including *muMerge*. Rutendo Sigauke performed initial testing and validation of the method, Zachary Maas assisted with data download

and husbandry, and Jessica Westfall designed figure layouts and presentation.

What follows is the TFEA manuscript as it was uploaded to BioRxiv.

Transcription factor enrichment analysis (TFEA): Quantifying the activity of hundreds of transcription factors from a single experiment

Jonathan D. Rubin[1], Jacob T. Stanley[2], Rutendo F. Sigauke[3], Cecilia B. Levandowski[1], Zachary L. Maas[2], Jessica Westfall[4], Dylan J. Taatjes[1], Robin D. Dowell[2,4,5,*]

[1] Department of Biochemistry, University of Colorado, Boulder CO 80309

[2] BioFrontiers Institute, University of Colorado, Boulder CO 80309

[3] Computational Bioscience Program, Anschutz Medical Campus, University of Colorado, Aurora, CO 80045

[4] Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder CO 80309

[5] Department of Computer Science, University of Colorado, Boulder CO 80309

* Corresponding author: robin.dowell@colorado.edu

1 Abstract

Detecting differential activation of transcription factors (TFs) in response to perturbation provides insight into cellular processes. Transcription Factor Enrichment Analysis (TFEA) is a robust and reliable computational method that detects differential activity of hundreds of TFs given any set of perturbation data. TFEA draws inspiration from GSEA and detects positional motif enrichment within a list of ranked regions of interest (ROIs). As ROIs are typically inferred from the data, we also introduce *muMerge*, a statistically principled method of generating a consensus list of ROIs from multiple replicates and conditions. TFEA is broadly applicable to data that informs on transcriptional regulation including nascent (eg. PRO-Seq), CAGE, ChIP-Seq, and accessibility (e.g. ATAC-Seq). TFEA not only identifies the key regulators responding to a perturbation, but also temporally unravels regulatory networks with time series data. Consequently, TFEA serves as a hypothesis-generating tool that provides an easy, rigorous, and cost-effective means to broadly assess TF activity yielding new biological insights.

2 Introduction

Transcription factors (TFs) are DNA-binding proteins that regulate transcription. When a cell is challenged by a change in the environment, it responds by altering the activity of one or more TFs. TFs, through transcriptional changes, are then responsible for altering cellular function and ultimately deciding cell fate. Because of their importance in global cellular programs, measuring differential TF activity between two conditions is a readout of high-level cellular biology and provides critical insight when details of the involved cellular processes are not known.

Experimental methods for measuring TF activity have largely focused on measuring protein-DNA binding, typically by chromatin immunoprecipitation (ChIP), resulting in high quality sequence recognition motifs for many TFs[19, 38]. Yet, not all binding sites lead to altered transcription activity[61, 55, 18]. Consequently, many of the approaches to inferring regulation by TFs combine ChIP data or motif hits with measures of gene expression[10, 28]. Relying on gene expression data, however, limits the effectiveness of these approaches. Gene expression assays, such as RNA-seq, are only indirect measures on actual transcription. RNA-seq is a steady state measure of RNA and reflects a combination of transcription and degradation[26, 57, 22]. Furthermore, the steady state nature of RNA-seq limits the response dynamics of the assay[25, 37, 42, 1, 36], as both newly created and long lived RNAs contribute to RNA measurements[49, 47]. Therefore, directly assaying transcription initiation improves on both the positional and temporal resolution when quantifying the activity of regulatory sites.

A large number of high throughput assays either directly or indirectly assay transcription initiation. Nascent transcription assays[16, 35] directly measure *bona fide* transcription, prior to RNA processing. Cap associated approaches, such as CAGE and GRO-CAP, target the 5′ cap of transcripts[4, 15, 58]. Transcription arises from a subset of nucleosome free regions, therefore chromatin accessibility data indirectly informs on the locations of transcription initiation. Likewise, some histone marks have been associated with actively transcribed regions, such as H3K27ac and H3K4me1/3 [11]. In principle, differential signals from these assays inform on the underlying mechanistic activity of TFs[6].

With differential regulatory data, the objective is to infer which transcription factors are causally responsible for the observed changes. With high quality motifs now residing in numerous databases[34, 43, 38], these catalogs can be leveraged to resolve the concurrent activity of many TFs. Historically, detecting motif enrichment in this way relied on sequences being classified into either signal or background and then calculating motif enrichment in signal sequences relative to background[11, 14]. More sophisticated approaches can take advantage of two additional factors: 1) positional information — where the motif is located relative to a region of interest[8, 6] and 2) differential information — the amount of change occurring within that region of interest[45, 12]. Relatively few techniques encode both types of information[39, 52, 24] and these currently provide no easily accessible software package or web-based application.

Our method, which we refer to as transcription factor enrichment analysis (TFEA) draws inspiration from the popular gene set enrichment analysis (GSEA) algorithm[56]. TFEA improves on our previous position based approach[6] and shows performance comparable to the state of the art motif enrichment approaches. Additionally, TFEA can be applied to a number of regulatory data types including PRO-seq, CAGE, DNAse-seq and ChIP-seq. Finally, TFEA is fast, computationally inexpensive, and designed with the user in mind, as we provide an easy to use web interface (https://tfea.colorado.edu), a command-line interface, and an importable Python 3 package. TFEA has the potential to become a transformative tool by providing easy downstream analysis aimed at distinguishing temporal and mechanistic details of complex regulatory networks.

3 Results

3.1 Overview

Conceptually, when a TF is active, it binds to a set of positions within the genome and alters transcription nearby, both at promoters and enhancers. Importantly, this process can both give rise to new transcripts and alter the levels of existing transcripts. Nascent transcription assays show that when a TF is activated, transcripts arise immediately proximal to the corresponding TF motif[1, 6]. In this work we introduce TFEA, which quantifies positional enrichment of TF motifs across an ordered list of regions (Figure 1). The key input into TFEA is a ranked list of regions of interest (ROIs) that typically are obtained independently from each replicate dataset but can also be a list of annotated regions, such as known promoters.

3.2 *muMerge*: Combining genomic features from multiple samples into consensus regions of interest

A key challenge in defining a set of consensus ROIs is retaining positional precision when combining region estimates that originate from different samples (replicates and conditions). To this end, we developed a statistically principled method of performing this combination called *muMerge* (See Supp. Fig. 3 and online method section 4.1.1 for details). In order to demonstrate the efficacy of *muMerge*, we compare its performance to two common methods for combining regions across multiple samples—merging all samples (e.g. with *bedtools merge*) and intersecting all samples (e.g. with *bedtools interesect*). We performed two tests using simulated data (Supp. Fig.4). For each replicate, we performed 10,000 simulations of sample regions for a single loci, and calculated the average performance.

Using the simulated regions, we first evaluate the methods' precision as the number of replicates increases. In Fig. 2a, we observe that as the number of replicates increases *muMerge* converges on the correct theoretical loci position

(μ) more quickly than the other two methods, while still maintaining the correct width for the region (Fig.2b).

The second test we performed sought to evaluate the accuracy of these methods when inferring two closely spaced loci, with increasing distance between those loci (Fig.2c). While closely spaced loci are challenging to distinguish, we observe that *muMerge* smoothly transitions from calling a single inferred loci (when μ_1 and μ_2 are too close to be resolved) to two distinct loci. In contrast, the merge and intersect methods show abrupt transitions that follow increasingly poor ROI width estimates (Fig.2d).

3.3 Transcription Factor Enrichment Analysis

Armed with the defined set of ROIs, the goal of TFEA is to determine if a given TF motif shows positional enrichment preferentially at regions with higher differential signal. Positional enrichment is consistent with the TF contributing to observed alterations. In prior work, we assessed the enrichment of motifs relative to positions of RNA polymerase initiation using a co-occurrence metric referred to as a motif displacement score (MD-Score; see Supp. Fig.5 for full details)[6]. Unfortunately, the MD-Score approach not only ignored alterations in transcript levels (See Supp. Fig.6) but also utilized an arbitrary distance threshold that classified motif proximity in a binary fashion.

To include transcript levels into the metric, we can rank ROIs by differential signal (e.g. transcription) before subsequently performing motif displacement calculations within these regions. The simplest approach to this problem is to compare the MD-Scores between the set of differentially transcribed regions and regions whose transcription is unchanged, a method we refer to as the differential motif displacement analysis (MDD, see Supp. Fig.7 for full details)[52, 24]. Unfortunately, the MDD method introduces an additional arbitrary threshold to classify regions as differentially transcribed or not and still uses the distance threshold set by the MD-Score approach.

In TFEA, we sought a non-binary enrichment metric that accounts for not only the underlying changes in transcription but also the positional enrichment of the motif (Fig.1). We begin by leveraging the statistically robust, gold standard DE-Seq package[2, 41] to rank regions based not only on the differential p-value but also the direction of fold change. Each region of interest then contributes positively to the enrichment curve in a weighted fashion. These weights are determined by the distance of the motif to the reference point using an exponential function to favor closer motifs. The subsequent enrichment score (E-Score in Fig. 1) is proportional to the integrated difference between the observed and background enrichment curves, calculated as the area under the curve (AUC) in Fig. 1 (see Eq. 8 for precise definition). The background (null) enrichment curve assumes uniform enrichment across all ROIs, regardless of differential signal.

By default, TFEA accounts for the known GC bias of enhancers and promoters by incorporating a correction to the enrichment score (Supp. Fig.8). Once E-Scores for all TFs have been calculated, we fit a linear regression to the

distribution of these scores as a function of motif GC-content. Corrected E-Scores are then calculated from the observed E-Score with the y-offset observed from the linear regression fit (see Eq. 11). This GC bias correction can be optionally turned off.

Subsequently, we assess the significance of the enrichment score by comparison to randomized ROI order, similar to GSEA[56]. To this end, we generate a null distribution of enrichment scores from random permutations, shuffling the rank order of regions and recalculating the E-Score for each shuffled permutation. The final significance of the enrichment score is then calculated from the Z-score, using the Bonferroni correction to account for multiple hypothesis testing. In this manner, TFEA provides a statistically robust and principled way of calculating the motif enrichment that accounts for both differential transcription and motif position in a manner that does not require arbitrary cutoffs.

3.4 Differential transcription signal improves motif inference over positional information alone

To assess the effectiveness of the TFEA method, we first compared its performance to both the MD-Score[6] and MDD-Score[52, 24] approaches. We examined a dataset in which a 1 hr Nutlin-3a treatment of HCT116 cells is used to activate TP53[1]. For all methods, sites of RNA polymerase loading and initiation were determined from GRO-seq data[1] using the Tfit algorithm[7] and combined using *muMerge* to identify ROIs. For all methods, the significance cutoff utilized was determined by comparing within treatment replicates (e.g. DMSO to DMSO) and identifying the cutoff at which no changes are detected (see Supp. Fig.9). Using these per method cutoffs, we recover TP53 from all three approaches (Fig. 3a). Notably, by including differential transcription information, the signal to noise ratio of TP53 detection is significantly improved—modestly in the case of MDD and dramatically for TFEA.

We next sought to determine whether TFEA could infer the responsible TF when the underlying changes in transcription were predominantly alterations in existing transcript levels. For this test, we relied on the fact that TP53 response in epithelial cells depends on the TP53 family member TP63[32]. Because TP53 and TP63 have nearly identical motifs, we reasoned that the presence of a constitutively active TP63 would result in elevated basal transcription proximal to TP53/TP63 motifs. To test this hypothesis, we performed PRO-seq on MCF10A cells after 1 hour treatment of either DMSO (control) or Nutlin-3a, and applied all three methods to the resulting data.

Consistent with the constitutive activity of TP63, we observed no change in the TP53 motif by MD-Score analysis (Figure 3b, left). This is due to a larger fraction of ROIs having pre-existing transcription prior to Nutlin-3a exposure in MCF10A relative to HCT116 cells (Figure 3c-e, Supp. Fig.10). While the MDD-Score method recovers TP53 (Fig.3b, middle), TFEA significantly improves the signal of the TP53 motif relative to the distribution of all other motifs (Fig.3b, right). For more detailed analysis of TP53 after Nutlin-3a in HCT116 and MCF10A, see Supp. Figs 11 and 12.

3.5 TFEA improves motif enrichment detection by incorporating positional information

We next sought to quantify the performance of TFEA with varying degrees of signal, background, and positional information. As a reference point, we leveraged the widely used MEME-Suite component AME, which quantifies motif enrichment by fitting a linear regression to ranked ROIs as a function of motif instances (Supp. Fig.13) [45]. To benchmark the two methods, we required biologically representative data sets with known motif enrichment so that error rates could be readily calculated. To this end, we utilized the sites of RNA polymerase initiation detected in untreated GRO-seq datasets of HCT116 cells[1] as the background ROIs. These regions were arbitrarily ordered to mimic a pattern of differential transcription. Subsequently, specific instances of the TP53 motif were generated from the position specific scoring matrix obtained from the HOCOMOCO database[33] and embedded via sequence replacement into the ordered ROI list.

We then varied the number of motifs across ROIs to simulate distinct signal to noise ratios and assess the accuracy of both TFEA and AME (Supp. Fig.14). Since the significance cutoff thresholds chosen for each method greatly influence the subsequent results, we first measured the mean false positive rate (FPR) and mean true positive rate (TPR) across tests of varying signal and background (Figure 4a). We found that AME detected many false positives (defined as all motifs besides TP53) at loose threshold cutoffs and therefore chose a strict cutoff of 1e-30 for AME. TFEA on the other hand, had a very low FPR even at loose thresholds with the TPR decreasing as the cutoff became stricter. We therefore chose a cutoff of 0.1 for TFEA. We next calculated an F1-Score based on the number of times each method correctly recovered the TP53 motif (and no other motifs) out of the 10 simulations for each test (Fig.4b).

We first measured F1-Scores for AME and TFEA with varying relative amounts signal and background (Figure 4b). We found that at high background levels (above 80%), AME was no longer able to detect the enrichment of TP53. TFEA on the other hand, was able to detect TP53 even at high background levels by incorporating positional information. Computing the differential F1-Scores between the two methods (Figure 4c) shows that TFEA performs well in cases where AME detects no enrichment of TP53 (26% of cases), whereas AME outperforms TFEA in 21% of cases.

To further determine how TFEA handles the loss of positional information, we chose the highest signal level tested and altered the variance (standard deviation of the signal position) and the background level (Figure 4d). As expected, AME shows consistent behavior regardless of the positional information of the motif. In contrast, TFEA is able to distinguish signal with differing levels of positional localization. In the extreme case of no positional localization (motifs embedded with a uniform distribution), TFEA performs only slightly worse than AME (Figure 4e).

Additionally, we sought to benchmark the runtime performance and memory usage of TFEA against AME. Here we leverage a first order Markov model (from

untreated DMSO samples[1]) to simulate increasing numbers of ROIs as input. Analyzing the core collection of HOCOMOCO TF motifs (n=401), we found that AME runtime increased exponentially while TFEA runtime increased linearly with a single processor (Supp. Fig.15a). Importantly, TFEA can utilize parallel processing leading to significantly faster runtimes. In terms of memory usage, although TFEA consumes more memory than AME, even in the worst case of 100,000 input regions, TFEA's memory footprint is less than 1Gb and therefore can still be run on a local desktop computer (Supp. Fig.15b).

Finally, we sought to examine the performance of TFEA and AME on real data and determine whether TFEA could identify biologically relevant signal in a dataset other than nascent RNA sequencing. Cap analysis of gene expression (CAGE) precisely defines the transcription start site (TSS) of individual transcripts[53, 21, 3]. We analyzed a CAGE-seq timeseries dataset from the FANTOM consortium[21, 9]. In this dataset, human derived monocytes were differentiated into macrophages and treated with lipopolysaccharide (LPS), a proxy for bacterial infection. Differential expression analysis was performed on each LPS time point comparing treatment to control to obtain a list of ranked ROIs.

TFEA recovered the immediate innate immune response, exemplified by the most rapid reported (within 15 min) activation of NF-$\kappa\beta$ (TF65/RELA, RELB, and NFKB1; Figure 5a). Additionally, TFEA temporally resolved the known secondary response that arises at later time points, which includes the activation of the IFN-stimulated gene factor 3 (ISGF3)[46] complex, comprising IRF9 and STAT1/2[48]. In contrast, AME did not recover the innate immune response at the earliest time point and provided less temporal resolution when distinguishing primary and secondary responses.

Concurrent with the immediate innate immune response, TFEA identified a set of TFs that exhibit a rapid decrease in E-Scores including ELF1/2[17], TYY1 [30][63], USF1/2[31], and GABPA[62]. The decreased E-Score set includes TYY1, a transcriptional inhibitor known to be activated directly by NFκB [54]. Reduction in the E-Score of TYY1 illustrates an important aspect of TFEA—namely, that it cannot distinguish between the activation of a repressor or the loss of an activator. Ultimately, we show with this proof of principle that if the cellular response to LPS was not known *apriori*, we could temporally resolve key aspects of the regulatory network using TFEA and dense time series CAGE data (Figure 5b and Supp. Figure 16).

3.6 TFEA works on numerous regulatory data types including ChIP and accessibility data

Though we developed *muMerge* and TFEA for the purpose of inferring TF activity from high resolution data on transcription initiation, this procedure can in principle be used on any assay that produces a localized readout on regulation, such as chromatin immunoprecipitation (ChIP) or DNA accessibility. Although these data sets are less precise and are not direct readouts of polymerase initiation, the popularity of these data make them readily available. To determine whether

TFEA could adequately infer TF activity from these datasets, we analyzed a timeseries dataset from ENCODE[19, 44] in which cells were treated with dexamethasone (Dex)—a known activator of the glucocorticoid receptor (GR).

TFEA correctly identifies GR as the key responding TF from the datasets that most closely capture RNA polymerase initiation (including p300, H3K27ac, and DNA accessibility), and does not identify GR for the transcriptionally repressive mark H3K9me3 (Figure 6a)[40, 44]. Surprisingly, the effects of p300 and H3K27ac are seen rapidly, as soon as 5min after dexamethasone treatment. As expected, H3K27ac deposition is temporally lagged behind its canonical acetyl-transferase p300[29, 60, 51]. Additionally, the enhancer marks H3K4me1 and H3K4me2 show strong enrichment of GR by 30min but the promoter mark H3K4me3 shows only modest enrichment, further supporting the finding that GR binds primarily at enhancers[44] (Supp. Fig. 17). Using the diversity of data types and dense time series, we can construct a temporally resolved mechanism of how GR effects changes in transcription (Figure 6b and c). In short, TFEA's results for this array of accessibility marks are exactly consistent with biological expectation.

3.7 Discussion

We present here transcription factor enrichment analysis (TFEA), a computational method that measures the global correlation between the position of a TF motif and its differential effects on transcription across the genome, following any given perturbation. We show that TFEA outperforms existing enrichment methods when positional data is available and is comparable to these methods in the absence of positional signal. Further, we show that TFEA, when leveraged with high resolution time series data, can provide mechanistic insight into the order of regulatory events responding to the perturbation.

A key aspect of TFEA is the incorporation of both positional and differential information in calculating TF activity. Most current motif enrichment algorithms use solely differential information, likely due to the poor positional resolution on historically popular techniques such as ChIP-Seq. Methods such as nascent transcription and CAGE provide higher resolution on the position of RNA polymerase initiation genome wide. To leverage the improved resolution of these methods, we introduce *muMerge*, a statistically principled way of combining ROIs across replicates and conditions that better captures position and length-scale information as compared to standard merging or intersecting approaches. The presence of improved positional information greatly increases the ability to detect biologically relevant TFs.

Although TFEA makes significant improvements in detecting the activity of TFs in response to perturbations, there are several aspects of this approach that could be improved. TFEA is dependent on having a collection of known motifs, yet some TFs have no known motif or one of poor quality. However, over time, the quality and numbers of TFs in the major databases have dramatically improved[38]. Furthermore, TFEA can only distinguish between paralogous motifs to the extent that they have distinct motifs. Importantly, motif scanning

still requires a fixed cutoff within TFEA. Future iterations of the method could conceivably eliminate this cutoff, but likely this will substantially increase runtimes for what may only be minor gains in performance. Genome-wide, sites of transcription initiation (both promoters and enhancers) show substantial GC bias. Often short high GC content motifs, which are exceedingly common in ROIs, appear to show significant changes with a perturbation. While we made some effort to account for this using linear regression, this approach is empirical and a more principled approach is desired.

Despite these caveats, TFEA recovers known TF dynamics across a broad range of data types in response to a variety of perturbations. Inevitably, the data type utilized influences the detection ability of TFEA. For example, while CAGE data provides precise resolution on the TSS, it must be deeply sequenced to reliably detect enhancer associate transcription events[15]. Consequently, TFs that predominantly regulate enhancers will likely be less detectable in poorly sequenced CAGE data. On the other hand, some methods are more capable of detecting immediate changes in RNA polymerase initiation, allowing for shorter more refined time points. As demonstrated here, TFEA is able to leverage the information from each data set by incorporating both its distinct positional and differential signal. Applying TFEA to diverse data types, using dense time series, can uncover a detailed mechanistic understanding of the key regulators that enact the cell's dynamic response to a perturbation.

4 Online Methods

4.1 TFEA

We have developed Transcription Factor Enrichment Analysis (TFEA) to identify transcription factors that demonstrate significant differential activity following a perturbation. It has been observed that, during a perturbation, the binding sites of active transcription factors co-localize with regulatory regions that exhibit strong differential RNA polymerase initiation[6]. TFEA leverages this observation to calculate an enrichment score that quantifies this activity and an associated significance for each TF.

Here we describe in detail the key steps of the TFEA pipeline (shown in Figure 1)—specifically, for each TF we describe how the main input (regions of interest—ROIs) are defined, how the ROIs are ranked, and how the enrichment score is subsequently calculated and GC-corrected.

4.1.1 Defining the Regions of Interest with *muMerge*

One input required for TFEA is a common set of regions of interest (ROIs) on which all experimental samples are evaluated. Each region (consisting of a genomic start and stop coordinate) represents a reference point (the midpoint of the region) and an uncertainty on that reference point (the width of the region). Biologically, the reference point is the presumed transcription start site. Regions can be derived from a number of data types, with varying degrees of

precision. For example, CAGE data provides a highly precise measure of a TSS while nascent sequencing is slightly less precise. Other assays like ChIP (for RNA polymerase or H3K4 methylation) or ATAC have much lower positional precision.

Regardless of the assay, most methods for identifying such regions fit each dataset independently (*e.g.*, a peak caller for ChIP data or Tfit for identifying sites of bidirectional transcription in nascent data). As a result, these regions will not be exactly consistent between samples (*e.g.* some sites are condition specific and even for shared sites boundaries may vary). Therefore, a method is needed to combine the regions from all the samples into a consensus set. To this end, we developed a probabilistic, principled method (hereafter referred to as *muMerge*) for determining consensus regions of interest, informed by the corresponding regions predicted from individual samples. *muMerge* was developed specifically for determining the set of consensus RNA polymerase loading and initiation sites observed in nascent sequencing data (by combining bidirectional calls from any number of samples) but it can also be applied to peak calls generated from numerous other regulatory data types (*e.g.*, ChIP, ATAC, or histone marks).

The basic assumption made by *muMerge* is that each sample is an independent observation of an underlying set of hypothetical loci—where each hypothetical loci has a precise critical point μ, of which the corresponding sample region ([*start*, *stop*]) is an estimate. We assume this loci is more likely to be located at the center of the sample region than at the edges, so *muMerge* represents the sample region by a standard normal probability distribution, centered on the region, whose standard deviation correlates with region width.

To calculate a best estimate (the ROI) for a given loci, *muMerge* calculates a joint probability distribution across all samples from all regions that are in the vicinity of the loci. This joint distribution is calculated by assuming:

1. replicates within a condition are independent and identically distributed (*i.i.d.*)

2. replicates *across* conditions are mutually exclusive (*i.e.*, a sample cannot represent multiple experimental conditions)

Hence *muMerge* computes the product of the normal distributions across all *replicates* within a condition and then sums these results across all *conditions*. The best estimates for the transcription loci μ are taken to be the maxima of this joint distribution—these are the ROI positions. Finally, to determine an updated width, or confidence interval, for each ROI, *muMerge* assumes that the original sample regions whose midpoints are closest to the new position estimate are the most informative for the updated width. Thus the ROI width is calculated by a weighted sum of the widths of the original regions, weighted by the inverse of the distance to each one.

***muMerge* mathematical description:** Principally, *muMerge* makes two probabilistic assumptions about sequenced samples:

- **Assumption A:** Replicate samples are independent measurements of *identical experimental conditions* and therefore any corresponding sample regions within them are independent and identically distributed (*i.i.d.*) observations of a common random variable (*i.e.*, the underlying hypothetical loci).
- **Assumption B:** Cross-condition samples are independent measurements of *mutually exclusive experimental conditions* and therefore any sample regions within them are observations of (potentially) disjoint random variables.

These two assumptions inform how *muMerge* accounts for each individual sample, when computing the most likely ROI for any given genomic location (see below for further details).

To start, the two inputs to *muMerge* are a set of regions for each sample (genomic coordinates: $\{[start, stop], ...\}$) that annotate the sequenced features present in the dataset, as well as an experimental conditions table that indicates the sample groupings (which samples are from which experimental condition). With these inputs, *muMerge* performs the following steps to compute a global set of ROIs:

1. Group overlapping sample regions, processing each group one at a time
2. Express each sample region as a positional probability distribution
3. Generate a joint distribution
4. Identify maximum likelihood ROI positions from the joint distribution
5. Compute ROI widths via weighted sum
6. Adjust the sizes of overlapping ROIs
7. Record final ROIs for the given group
8. Repeat 2–8 for all remaining groups

First, from the input samples, *muMerge* groups all sample regions that overlap in genomic coordinate (a region is grouped with all other regions it overlaps and, transitively, with any regions overlapping those). We denote a single group of overlapping regions as G_r. This grouping is done globally for all samples, resulting in a set of grouped regions $G = G_r$, such that every sample region is contained in exactly one grouping G_r (i.e., $G_r \cap G_s = \emptyset, \forall\, r \neq s$). Then each group of regions, G_r, is processed individually, as the remainder of this section describes. For a given group, we denote each sample region within it as the 2-tuple $(\mu_k, \sigma_k)_{ij} \in G_r$, where μ_k is the genomic coordinate (base position) of the center of the region and σ_k is the region half-width (number of bases). The indices denote the k-th sample region for replicate j in condition i.

muMerge then processes the regions in G_r as follows. Each region within the group is expressed as a standard normal distribution (ϕ) as a function of base position x,

$$(\mu_k, \sigma_k)_{ij} \rightarrow p_{ij}^{(k)}(x) = \phi\left(\frac{x - \mu_k}{\rho\,\sigma_k}\right) \tag{1}$$

where ρ is the "width ratio"— the ratio of the half-width sample region to the standard deviation of the normal distribution—with a default of $\rho = 1$ (user option). This distribution represents the probability of the location for the underlying hypothetical loci (μ), of which $(\mu_k, \sigma_k)_{ij}$ is an estimate. For those samples with no regions within G_r, the probability distribution is expressed as a uniform, $p_{ij}^{(k)}(x) = 1/\Delta$ where Δ is the full range encompassed by the overlapping sample regions. In other words, we assume that if the sample contains no data to inform the location of the underlying loci at that location, then all positions are equally likely for that sample. *muMerge* then calculates a joint distribution ($\mathcal{P}(x)$) by combining all $p_{ij}^{(k)}(x)$ for the group as follows:

$$\mathcal{P}(x) = \sum_i \left(\prod_j \left(\sum_k p_{ij}^{(k)}(x)\right)\right) \tag{2}$$

Here we are calculating the product of the replicate distributions (index j— those within a given experimental condition), consistent with our probabilistic assumption A, and the sum of the resulting distributions across experimental conditions (i index), consistent with our probabilistic assumption B. Though this function is not a normalized probability distribution, we are only interested in relative values of $\mathcal{P}(x)$. Specifically, we are interested in the maxima of this function. We identify the set of maxima (which we denote $\{\widehat{\mu}_k\}$) and rank them by the function value for each position, $\mathcal{P}(\widehat{\mu}_k)$. We then keep the top $M + 1$ from the ranked set, where M is the median number of regions per sample in G_r (user option). This is our final set of estimates on the hypothetical loci positions, μ—*i.e.*, the positions of our ROIs for group G_r.

For each $\widehat{\mu}_k$, we then calculate a width for the resulting ROI. We do so for each by calculating a weighted sum over the set of all original sample regions in the group, $\{(\mu_k, \sigma_k)_{ij}\}$, weighted by the inverse of the distance from the final position estimate to each μ_k. Thus the final ROI half-width, $\widehat{\sigma}_k$, is calculated as follows:

$$\widehat{\sigma}_k = \sum_i \frac{\sigma_i}{|\widehat{\mu}_k - \mu_i| + 1} \Bigg/ \sum_i \frac{1}{|\widehat{\mu}_k - \mu_i| + 1} \tag{3}$$

where i indexes all sample regions in the group $G_r = \{(\mu_k, \sigma_k)_{ij}\}$. Our rationale is that the width of those sample regions that are closer to the ROI position $\widehat{\mu}_k$, are more informative for the ROI width and therefore are given a larger weight. This results in a set of ROIs $\{(\widehat{\mu}_k - \widehat{\sigma}_k,\, \widehat{\mu}_k + \widehat{\sigma}_k)\}$.

Finally, we determine if there is overlap between any of the regions in this set of ROIs. If so, any two overlapping regions are reduced in size, symmetrically about their centers, until they no longer overlap. This is done so that any

genomic position can be uniquely associated with an ROI. The final ROIs for the group are then written to an output file to be used downstream in the pipeline. This process is repeated for all groups of overlapping sample regions (i.e., $\forall G_r \in G$).

4.1.2 Ranking ROIs

With a set of ROIs identified, the next step is to rank them by differential signal. Because the goal of TFEA is to identify transcription factors that are enriched during a perturbation and because the ROIs are associated with transcription factor activity, it follows that a ranking based on the differential signal at the ROIs would capture the regulatory behavior of the TF. For different types of datasets, the differential signal represents different biological processes—differential transcription for nascent (PRO-seq or GRO-seq), differential accessibility (DNAse or ATAC-Seq), and differential occupancy for ChIP. There are a number of ranking metrics one could use that are based on these differential signals—for example, difference in coverage, log-fold change, or a differential significance (p-value). For TFEA, we chose to rely on a well-established tool (*DESeq2*) to perform our ranking, since it was designed to model the statistical variation found in sequencing data[41].

For a set of ROIs, TFEA calculates read coverage for each replicate and condition using *bedtools multibamcov* (version 2.25.0)[50]. TFEA then inputs the generated counts table into *DESeq2*[41] (or *DESeq*[2] if no replicates are provided) to obtain differential read coverage for all ROIs. By default, these regions are then ranked by the *DESeq2* computed p-value, separated by positive or negative log-fold change (alternative user option to rank the ROIs by fold-change). In other words, the ROIs are ranked from the most significant positive fold-change to the most significant negative fold-change.

4.1.3 Identifying location of motif instances

Accurately identifying the locations of motif instances relative to each ROI is a critical step in the TFEA pipeline. By default TFEA uses the motif scanning method FIMO, which is a part of the MEME suite (version 5.0.3)[23]. FIMO represents each TF by a base-frequency matrix and uses a zero-order background model to score each position of the input sequences. For each ROI, we scan the 3kb sequence surrounding the ROI center ($\widehat{\mu}_i \pm 1.5$kb). This 3kb window was chosen primarily to reduce computation time and is also consistent with the window used for the MD-score method[6]. For each TF, we utilize a scoring threshold of 10^{-6} and keep the highest scoring position (denoted m_i), in the event more than one motif instance is identified. If no position score above the threshold, then no m_i is recorded for the ROI. Our background model is determined by calculating the average base frequency over all 3kb regions. For each TF, we use the frequency-matrix from the HOCOMOCO database[33] with a default psuedo-count of 0.1.

4.1.4 Enrichment Score

With the motif instances identified for each of the ranked ROIs, we now detail how TFEA calculates the enrichment score ("E-Score"—in Fig. 1) for each transcription factor. The procedure for calculating enrichment requires two inputs:

1. N-tuple sequence ($\widehat{\mu}_i$)—the genomic coordinates for reference points, assumed to be the centers of all ROIs (*e.g.*, consensus ROIs calculated by *muMerge*), ranked by *DESeq2* p-value (separated by the sign of the fold-change).

2. Sequence (m_i)—the genomic coordinates of each max-scoring motif instance (*e.g.*, motif loci generated by scanning with FIMO), for each ROI.

We first calculate the motif distance d_i for each ROI—the distance from each $\widehat{\mu}_i$ to the highest scoring motif instance m_i within 1.5kb of $\widehat{\mu}_i$. If no m_i exists within 1.5kb, then d_i is assigned a null value (Ø) (Eq. 4).

$$d_i = \begin{cases} |\widehat{\mu}_i - m_i|, & \text{if } m_i \text{ is present} \\ \text{Ø}, & \text{if } m_i \text{ is not present} \end{cases} \tag{4}$$

Next, we calculate the background distribution of motif distances. We assume the majority of the ROIs experience no significant fold-change—namely, those ROIs in the middle of the ranked list. Consequently, we calculate the mean, background motif distance (Eq. 5) for those ROIs whose rank is between the first and third quartiles of the sequence of ROI positions, ($\widehat{\mu}_i$), as follows

$$\bar{d} = \text{mean}\{d_i \,|\, \forall\, i, \text{if } Q_1 \leq i \leq Q_3 \text{ and } d_i \neq \text{Ø}\} \tag{5}$$

where Q_1 and Q_3 are the first and third quartiles, respectively. Our assumption is that the inter-quartile range of the sequence ($\widehat{\mu}_i$)—between indices Q_1 and Q_3—represents the background distribution of motif distances for the given transcription factor, and therefore defines the weighting scale for significant ROIs in our enrichment calculation. We found this to be essential since the background distribution varies between transcription factors. This variation in the background can be attributed to the similarity of a given motif to the base content surrounding the center of ROIs. For example, in the case of RNA polymerase loading regions identified in nascent transcription data (which demonstrate a greater GC-content proximal to μ as compared to genomic background[6]), GC-rich transcription factor motifs were more likely to be found proximal to each ROI by chance and thus resulted in a smaller $\bar{d}$ than would be the case for a non-GC-rich motif.

Having calculated the mean background motif distance, we proceed to calculate the enrichment contribution (*i.e.*, weight—Eq. 6) for each ROI in the sequence (see "Weight Calculation" in Fig. 1).

$$w_i = \begin{cases} e^{-d_i/\bar{d}}, & \text{if } d_i \neq \text{Ø} \\ 0, & \text{if } d_i = \text{Ø} \end{cases} \tag{6}$$

In order to calculate the E-Score, we first generate the enrichment curve for the given TF (solid line in "Enrichment Curve" in Fig. 1) and the background (uniform) enrichment curve (dashed line in "Enrichment Curve" in Fig. 1). We define the E-Score as the integrated difference between these two (scaled by a factor of 2, for the purpose of normalization). The enrichment curve (Eq. 7), which is the normalized running sum of the ROI weights, and the E-Score (Eq. 8) are calculated as follows:

$$e(i) = \frac{\sum_{k=0}^{i} w_k}{\sum_{k=0}^{N} w_k} \tag{7}$$

$$E = \frac{2}{N} \sum_{i} \left(e(i) - \frac{i}{N} \right) \tag{8}$$

where i is the index for the ROI rank and i/N represents the uniform, background enrichment value for the ith of N ROIs. The background enrichment assumes every ROI contributes an equal weight w_i, regardless of its ranking position. Therefore, the enrichment curve (Eq. 7) will deviate significantly from background if there is correlation between the weight and ranked position of the ROIs. In this case, the E-Score will significantly deviate from zero, with $E > 0$ indicating either increased activity of an activator TF or decreased activity of a repressor TF. Likewise, $E < 0$ indicates either a decrease in an activator TF or an increase in a repressor TF. By definition, the range of the E-Score is -1 to $+1$.

Unlike GSEA, which uses a Kolmogorov–Smirnov-like statistic to calculate its enrichment score[56], the TFEA E-Score is an area-based statistic. GSEA was designed to identify if a predetermined, biologically related subset of genes is over-represented at the extremes of a ranked gene list. Therefore, the KS-like statistic is a logical choice for measuring how closely clustered are the elements of the subset, since it directly measures the point of greatest clustering and otherwise is insensitive to the ordering of the remaining elements. Conversely, because TFEA's ranked list does not contain two categories of elements (the ROIs) and all elements can contribute to the E-Score, we wanted a statistic that was sensitive to how all ROI in the list were ranked—for this reason, we chose the area-based statistic. The null hypothesis for TFEA assumes all ROI contribute equally to enrichment, regardless of their motif co-localization and rank. Hence the uniform background curve, to which the enrichment curve is compared.

In order to determine if the calculated E-Score (Eq. 8) for a given transcription factor is significant, we generate a E-Score null distribution from random permutations of $(\widehat{\mu}_i)$. We generate a set of 1000 null E-Scores $\{E'_i\}$, each calculated from an independent random permutation of the ranked ROIs, $(\widehat{\mu}_i)$. Our E-Score statistic is zero-centered and symmetric, therefore we assume $\{E'_i\} \sim \mathcal{N}(E_0, \sigma_E^2)$. The final E-Score for the transcription factor is compared to this null distribution to determine the significance of the enrichment.

Prior to calculating the E-Score p-value, we apply a correction to the E-Score based on the GC-content of the motif relative to that of all other motifs to be

tested (user configurable). This correction was derived based on the observation that motifs at the extremes of the GC-content spectra were more likely to called as significant across a variety of perturbations. We calculate the E-Scores for the full set of transcription factors as well as the GC-content of each motif, $\{(g_i, E_i)\}$. We then calculate a simple linear regression for the relationship between the two,

$$\widehat{b} = \bar{E} - \widehat{m}\bar{g} \tag{9}$$

$$\widehat{m} = \frac{\sum_{i=1}^{n}(g_i - \bar{g})(E_i - \bar{E})}{\sum_{i=1}^{n}(g_i - \bar{g})} \tag{10}$$

$$E_{GC}(g) = \widehat{b} + \widehat{m}g \tag{11}$$

where $\bar{E}$ and $\bar{g}$ are the average E-Score and average GC-content. $E_{GC}(g)$ is the amount of the E-Score attributed to the GC-bias for a motif with GC-content g. Thus the final E-Score for the transcription factor is given by $E_{TF} = E - E_{GC}(g_{TF})$, the difference between Eq. 8 and 11. If GC-content correction is not performed, then Eq. 8 is taken to be the final E-Score. The p-value for the final TF E-Score is then calculated from the Z-score, $Z_{TF} = (E_{TF} - E_0)/\sigma_E$.

4.2 Software Availability

TFEA is available for download at https://github.com/Dowell-Lab/TFEA and comes with muMerge integrated. Alternatively, *muMerge* can be downloaded independently at https://github.com/Dowell-Lab/mumerge. Additionally, TFEA can be utilized through the web interface at https://tfea.colorado.edu/.

4.3 Benchmarking

In order to benchmark the performance of *muMerge* and TFEA, we performed a number of simulations that isolate the different parameters of *muMerge* and TFEA, comparing the performance to that of some commonly used alternatives. Here we describe how the data for each test was generated.

4.3.1 *muMerge*: Simulating replicates for calculation of ROIs

To test the performance of *muMerge* in a principled manner, we first generate replicate data in a way that simulates the uncertainty present in individual samples. For each replicate, we perform 10,000 simulations of sample regions for a single loci, and calculate the average performance. For each simulation we assume a precise position and width for the hypothetical loci and model the uncertainty of each sample region with a binomial and Poisson distribution, respectively. The position of each sample region, $\widehat{\mu}$, is pulled from a symmetric binomial distribution $\widehat{\mu} \sim B(n = 100, p = 0.5)$, centered at zero. The half-width of each sample region, W, is pulled from a Poisson distribution $W \sim Pois(\lambda = 100)$. The specific distributions utilized to generate the sample regions are as follows:

$$\text{loci estimate} \equiv \begin{cases} \text{position:} & \widehat{\mu} \sim \mu + B(n = 100, p = 0.5) - np \\ \text{half-width:} & \widehat{\sigma} \sim Pois(\lambda = 100) \end{cases} \tag{12}$$

Here $B(\cdot)$ is the binomial distribution centered at np with success probability 0.5 and variance $np(1-p) = 25$. Thus, the position estimator $\widehat{\mu}$ is centered at μ. $Pois(\cdot)$ is the Poisson distribution, thus, the half-width for each sample region have mean and variance of $\lambda = 100$.

The first test consisted of inferring a single loci (located at $\mu = 0$) from an increasing number of replicates. A sample region for each replicate was generated from Eq. 12. This simulation was repeated 10,000 times for each number of replicates being combined. The methods *muMerge*, *bedtools merge* and *bedtools intersect* were applied to each of the 10,000 simulations. The average error on the midpoint (its deviation from the true loci position, $\mu = 0$) and region width were calculated for the regions generated from each method, averaged over all 10,000 simulations. The behavior of the average positional error and region width as a function of number of combined replicates is shown in Fig. 2a, b.

The second test consisted of inferring two loci ($\mu_1 = -x$ and $\mu_2 = +x$) as the distance between those loci was increased (from $x = 0$ to 200). This simulation was repeated 10,000 times for each value of x (with 3 replicates). The distribution of the inferred positions and widths were plotted, using *muMerge*, *bedtools merge* and *bedtools intersect*. The distribution of positions and widths as a function of the distance between μ_1 and μ_2 are shown in Fig. 2c, d.

4.3.2 TFEA: Simulated motif enrichment

To generate test sequences for benchmarking, we randomly sampled 10,000 sequences from detected bidirectionals in untreated HCT116 cells [1]. We then embedded instances of the TP53 motif in the highest ranked sequences with a normal distribution with $\mu = 0$ and $\sigma = 150$ (representative of signal). To simulate background noise, we embedded instances of the TP53 motif with a uniform distribution to a percentage of the remaining sequences (chosen randomly). To calculate an F1-Score, for each scenario of varying signal to background we generated 10 simulations. We then calculated the harmonic mean of precision and recall with the aggregate p-values of all 10 simulations measuring all 401 TF motifs within the HOCOMOCO database (total 4010 TF motifs). True positives, in this case, were the 10 instances of the TP53 motif that should be significantly enriched. Any other significantly enriched TF motifs were considered false positives. We performed two sets of tests: 1) varying the amount of signal and the amount of background and 2) varying the standard deviation of the highest signal tested (10% signal; with the last scenario being uniform signal distribution) and the amount of background.

4.3.3 TFEA: Testing compute performance

The base (ATGC) content of regulatory regions was calculated from the sites of RNA polymerase initiation inferred in HCT116 DMSO (using Tfit; described in [6]). One million 3kb sequences were generated based on the empirical probability of the positional base composition. We then randomly sampled an increasing number of sequences (up to 100,000) to be used in the computational processing tests. Run time and compute resources were measured using the Linux *time* command on a single node of a 70-node mixed-platform high-memory compute cluster running CentOS 7.4. To compute the runtime for a single processor, we added the *systime* and *usertime*. To compute memory usage for a single processor, we reran TFEA using only a single processor.

4.4 Datasets Utilized

We generated PRO-seq libraries for MCF10A cells with and without Nutlin-3a. Additionally, a number of publicly available datasets were utilized, including: Allen 2014 (Nutlin-3a, GC-correction), ENCODE (GGR: Reddy - Dex/GR) and FANTOM (Baillie - Macrophage/LPS). See supplemental material for a full list of accession codes.

4.4.1 PRO-Seq in MCF10A

Cas9RNP formation: sgRNA was formed by adding tracrRNA (IDT cat# 1072533) and crRNA (TP53 exon 2, positive strand, AGG PAM site, sequence: GATCCACTCACAGTTTCCAT) in a 1:1 molecular ratio together and then heating to 95°C and then allowing to slowly cool to room temperature over 1 hour. Cas9RNP was then formed by adding purified Cas9 protein to sgRNA at a ratio of 1:1.2. 3.7μL of purified Cas9 protein at 32.4μM was added to 2.9μL of 50μM sgRNA. This was then incubated at 37°C for 15 minutes, and used at 10μM concentration within the hour.

Donor Plasmid Construction: Vector Builder was used to construct plasmid. Insert was flanked by 1.5kb homology arms, and mCherry was inserted as a selection marker.

CRISPR/Cas9 Genome Editing: MCF10A cells cultured in DMEM/F12 (Invitrogen #11330-032) media containing 5% horse serum (LifeTech #16050-122), 20ng/mL EGF ((Peprotech #AF-100-15), 0.5μg/mL Hydrocortisone (Sigma #H0888-1g), 100ng/mL Cholera toxin (Sigma #C8052-2mg), 10μg/mL insulin (Sigma #I1882-200mg), and 1x Gibco 100x Antibiotic-Antimycotic (Fisher Sci, 15240062) penicillin-streptomycin. Cells were split 24 hours prior to experiment and grown to approximately 70% confluency on a 15cm plate. Media was aspirated, and the cells were washed with PBS. 4ml of trypsin per plate were used to harvest adherent cells, after which 8mL of resuspension medium (DMEM/F12 containing 20% horse serum and 1x pen/strep) was added to each plate to

neutralize the trypsin. Cells were collected in a 15ml centrifuge tube and spun down at 1,000xg for 5 minutes, then washed in PBS and spun down again at 1,000xg for 5 minutes. Cells were counted using a hemocytometer and $5x10^5$ cells were put in individual 1.5mL eppedorph tubes for transfection. Cells were re-suspended in 4.15μL Buffer R, 10μM Cas9RNP (6.6μL), 1μg WTp53 donor plasmid (1.25μL). Mixture was drawn up into a 10μL Neon pipet tip, electroporated using the Neon Transfection Kit with 10μL tips (1400V, 20ms width, 2 pulse). Transfected cells were then pipetted into 2mL of antibiotic free media. After 1 week of recovery, cells were then single cell sorted into 96 well plate based on mCherry expression. Clones were then verified with sequencing, PCR, and western blot.

Nuclei Preparation: MCF10A WTp53 cells were seeded on three 25cm dishes (1x107 cells per dish) for each treatment 24 hours prior to the experiments (70% confluency at the time of the experiment). Cells were treated simultaneously with 10μM Nutlin3a or 0.1% DMSO for 1 hour. After treatment, cells were washed 3x with ice cold PBS, and then treated with 10 ml (per 15 cm plate) ice-cold lysis buffer (10 mM Tris–HCl pH 7.4, 2 mM MgCl2, 3 mM CaCl2, 0.5% NP-40, 10% glycerol, 1 mM DTT, 1x Protease Inhibitors (1mM Benzamidine (Sigma B6506-100G), 1mM Sodium Metabisulfite (Sigma 255556-100G), 0.25mM Phenylmethylsulfonyl Fluoride (American Bioanalytical AB01620), and 4U/mL SUPERase-In). Cells were centrifuged with a fixed-angle rotor at 1000xg for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1.5 mL lysis buffer to a homogenous mixture by pipetting 20-30X before adding another 8.5 mL lysis buffer. Suspension was centrifuged with a fixed-angle rotor at 1000xg for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1 mL of lysis buffer and transferred to a 1.7 mL pre-lubricated tube (Costar cat. No. 3207). Suspensions were then pelleted in a microcentrifuge at 1000xg for 5 min at 4°C. Next, supernatant was removed and pellets were resuspended in 500 μL of freezing buffer (50 mM Tris pH 8.3, 40% glycerol, 5 mM MgCl2, 0.1 mM EDTA, 4U/ml SUPERase-In). Nuclei were centrifuged 2000xg for 2 min at 4°C. Pellets were resuspended in 100 μL freezing buffer. To determine concentration, nuclei were counted from 1 μL of suspension and freezing buffer was added to generate 100 µL aliquots of $10x10^6$ nuclei. Aliquots were flash frozen in liquid nitrogen and stored at −80°C.

Nuclear run-on and RNA preparation: Nuclear run-on experiments were performed as described (Mahat et al., 2016) with the following modifications: the final concentration of non-biotinylated CTP was raised from 0.25 μM to 25 μM, a clean-up and size selection was performed using 1X AMPure XP beads (1:1 ratio) (Beckman) prior to test PCR and final PCR amplification, and the final library clean-up and size selection was accomplished using 1X AMPure XP beads (1:1 ratio) (Beckman).

Sequencing: Sequencing of PRO-Seq libraries was performed at the BioFrontiers Sequencing Facility (UC-Boulder). Single-end fragment libraries (75 bp) were sequenced on the Illumina NextSeq 500 platform (RTA version: 2.4.11, Instrument ID: NB501447), demultiplexed and converted BCL to fastq format using bcl2fastq (bcl2fastq v2.20.0.422); sequencing data quality was assessed using FASTQC (v0.11.5) and FastQ Screen (v0.11.0), both obtained from https://www.bioinformatics.babraham.ac.uk/projects/. Trimming and filtering of low-quality reads was performed using BBDUK from BBTools (v37.99) (Bushnell, n.d.) and FASTQ-MCF from EAUtils (v1.05) [5].

Availability: MCF10A PRO-seq data is available in GEO with accession numbers GSE142419.

4.4.2 Data Processing

GRO/PRO-Seq data: All GRO-Seq and PRO-Seq data were processed using the Nextflow[20] NascentFlow pipeline (v1.1 [59]) specifying the '–tfit' flag. Subsequent Tfit bed files from all samples were combined with *muMerge* to obtain a consensus list of ROIs.

ENCODE data: Raw bed and bam files were downloaded directly from ENCODE (encodeproject.org). These files were inputted directly into the TFEA pipeline for processing and analysis. AME analysis was performed on the ranked ROI list produced as an optional output from TFEA.

FANTOM data: Raw expression tables for the Macrophage LPS time series were downloaded using the table extraction tool (TET) from the FANTOM Semantic catalogue of Samples, Transcription initiation, And Regulations (SSTAR; http://fantom.gsc.riken.jp/5/sstar/Macrophage_response_to_LPS). Because the annotations for regions within hg38 counts tables contained "hg19", we performed this analysis in the hg19 genome with the hg19 counts table instead of the hg38 counts table. We then performed DE-Seq analysis (since there were no replicates) on each time point compared to control and ranked the annotated regions within the counts table similar to Figure 1. We then ran TFEA and AME with default settings on each of the three donors. We displayed only data for donor 2, as this sample had the most complete time series data.

Clustering FANTOM data: We retained TFs with at least 15 significant (p-adj < 0.1) time points (representing 2/3 of all timepoints) from the TFEA output and applied K-means clustering. Clustering of the time series data was performed on the first two hours only, in order to distinguish the early responses to LPS infection. K-means clustering was conducted using the Hartigan and Wong algorithm with 25 random starts and 10 iterations for $k = 3$ [27]. The optimal number of clusters was selected using the Elbow method [13].

String database analysis: Protein names from TFs that were found to be significant in at least 15 time points were taken from the HOCOMOCO database. These proteins were inputted directly into the String database (https://string-db.org). Clusters were formed by selecting the MCL clustering option with an inflation parameter of 3 (default). Network edges were selected to indicate the strength of the data support. Finally, nodes disconnected from the network were hidden.

5 Figure Captions

Figure 1: **TFEA calculates motif enrichment using differential and positional information.** The TFEA pipeline requires, minimally, a ranked list of ROIs. Optionally, a user may provide raw read coverage and regions, in which case TFEA will perform ranking using DE-Seq [2, 41] analysis. With a set of ranked ROIs, TFEA analyzes motif enrichment for each motif provided as a .meme database. For each motif, positions are determined by FIMO scans and an enrichment curve is calculated by weighting each motif instance (using an exponential decay function) and adding this value to a running sum. An E-Score is calculated as 2 * the AUC, e.g. the area under the curve between the running sum and a uniform background (line). For statistical significance, the ROI rank is shuffled 1000 times, and E-scores are recalculated for each shuffle. The true E-Score is then compared to the distribution of E-Scores obtained from the shuffling events. For example output of TFEA see Supp Fig 1 and Supp Fig 2.

Figure 2: ***muMerge* precisely combines multiple samples into consensus ROIs.** Here we show a comparison of three methods (*bedtools merge*, *bedtools intersect*, and *muMerge*) for generating ROIs from multiple samples (See Supp Fig 4 for schematic of each method). Test 1 demonstrates the position and width accuracy of a calculated ROI for a single loci, μ, as the number of sample replicates are increased (from one to ten). With *muMerge* (a) the positional uncertainty decreases quickly while the (b) estimated ROI width remains relatively constant. Standard error, indicated by shading, is less than the line width. Test 2 demonstrates the precision of the calculated ROI for two closely spaced loci, μ_1 and μ_2, as the spacing between them is increased. In this case, *muMerge* (c) transitions from a single loci to two distinct loci more gradually and (d) the estimated ROI widths do not deviate from the expected value. In all cases, expected value and variance for simulations is indicated by grey lines and shading, respectively. For further detail on the results of Test 1 and 2 and how the simulations were performed, see Supp Fig 4 and methods section 4.3.1.

Figure 3: **TFEA improves the detection of p53 following Nutlin-3a treatment.** (a) Application of the MD-Score, MDD-Score, and TFEA to GRO-Seq data in HCT116 cells with 1hr Nutlin or DMSO treatment [1]. Cutoffs determined by comparing untreated replicates (see Supp Fig 9). (b) Application of the MD-Score, MDD-Score, and TFEA to PRO-Seq data in MCF10A cells with 1hr Nutlin or DMSO treatment. (c) Motif displacement distribution (as heatmap) of TP53 motif instances within 1.5kb of all ROI in either DMSO (blue) or Nutlin-3a (red). (d) Percentage overlap of TP53 motifs within 150bp in DMSO and Nutlin-3a ROIs. (e) Similar to (c) but in MCF10A cells.

Figure 4: **TFEA balances TF positional and differential signal.** (a) Optimal cutoffs are determined using the mean true positive rate (TPR; green) and mean false positive rate (FPR; orange) across different signal and background levels as a function of varying the threshold cutoff. (b) F1-score of AME and TFEA for varied signal and background, using optimal AME cutoff 1e-30 and TFEA cutoff 0.1. (c) Difference in F1-score across all simulations (n=121). TFEA outperforms AME in 26% of cases (red) whereas AME outperforms TFEA in 21% of cases (blue). (d) F1-scores and (e) difference in scores for highest signal tested (10% signal), now varying the standard deviation of the signal and background. See Supp Fig 14 for more details on simulations.

Figure 5: **TFEA dissects the temporal dynamics of infection.** (a) Analysis of LPS timeseries CAGE data[21, 9] using AME (top) or TFEA (bottom). Trajectories of activity profiles shows LPS triggers immediate activation of the NF-$\kappa\beta$ complex (TF65/RelB/NFKB1; yellow), observable at 15min (blue arrow). TFEA detects a concomitant down regulation of a set of transcription factors, exemplified here by TYY1 (purple). TFEA also resolves subsequent dynamics (green bracket) of ISGF3 activation (containing IRF9/STAT1/STAT2; red lines). (b) Schematic depicting the molecular insights gained from TFEA analysis. See Supp Fig 16 for more analysis.

Figure 6: **TFEA captures rapid dynamics of glucocorticoid receptor (GR) following treatment with dexamethasone.** (a) TFEA correctly identifies GR from time series ChIP data on the histone acetyl-transferase p300, H3K27ac and DNase I. No signal is observed in the negative control H3K9me3. TFEA correctly shows a temporal lag in H3K27ac signal (yellow arrow). (b) Known cellular dynamics of GR induced by dexamethasone. (c) Mechanistic and temporal insights gained by performing TFEA analysis, question marks indicate datasets where earlier time points were not available to resolve temporal information.

Acknowledgments

This work was funded in part by a National Science Foundation (NSF) ABI grant number 1759949, a National Institutes of Health (NIH) grant RO1 GM125871, and an NIH training grant T32 GM008759. We acknowledge the BioFrontiers Computing Core at the University of Colorado Boulder for providing high-performance computing resources (NIH 1S10OD012300) supported by BioFrontiers' IT. In particular we thank Matt Hynes-Grace, Jon DeMasi, and Ethan Kern for assistance in the development of the TFEA website. Finally, we also thank the BioFrontiers Institute Next-Gen Sequencing Core and the Biochemistry Shared Cell Culture Facility for their invaluable contributions to this study.

References

[1] M. A. Allen, H. Mellert, V. Dengler, Z. Andryzik, A. Guarnieri, J. A. Freeman, X. Luo, W. L. Kraus, R. D. Dowell, and J. M. Espinosa. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *eLife*, 3:e02200, 2014. doi: 10.7554/eLife.02200.

[2] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106, 2010.

[3] R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jorgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Muller, T. F. Consortium, A. R. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–461, 03 2014.

[4] R. Andersson, P. Refsing Andersen, E. Valen, L. J. Core, J. Bornholdt, M. Boyd, T. Heick Jensen, and A. Sandelin. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun*, 5, 11 2014.

[5] E. Aronesty. Comparison of Sequencing Utility Programs. *The Open Bioinformatics Journal*, 7(1), Jan. 2013.

[6] J. G. Azofeifa, M. A. Allen, J. R. Hendrix, T. Read, J. D. Rubin, and R. D. Dowell. Enhancer RNA profiling predicts transcription factor activity. *Genome research*, feb 2018.

[7] J. G. Azofeifa and R. D. Dowell. A generative model for the behavior of RNA polymerase. *Bioinformatics*, 33(2):227–234, 09 2016.

[8] T. L. Bailey and P. Machanick. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Research*, 40(17):e128–e128, Sept. 2012.

[9] J. K. Baillie, E. Arner, C. Daub, M. De Hoon, M. Itoh, H. Kawaji, T. Lassmann, P. Carninci, A. R. R. Forrest, Y. Hayashizaki, G. J. Faulkner, C. A. Wells, M. Rehli, P. Pavli, K. M. Summers, and D. A. Hume. Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. *PLoS Genetics*, 13(3), Mar. 2017.

[10] P. J. Balwierz, M. Pachkov, P. Arnold, A. J. Gruber, M. Zavolan, and E. van Nimwegen. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Research*, 24(5):869–884, May 2014.

[11] A. J. Bannister and T. Kouzarides. Regulation of chromatin by histone modifications. *Cell Research*, 21(3):381–395, Mar. 2011.

[12] I. Berest, C. Arnold, A. Reyes-Palomares, G. Palla, K. D. Rasmussen, K. Helin, and J. B. Zaugg. Quantification of differential transcription factor activity and multiomics-based classification into activators and repressors: diffTF. *bioRxiv*, page 368498, Dec. 2018.

[13] P. Bholowalia and A. Kumar. Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9), 2014.

[14] V. Boeva. Analysis of genomic sequence motifs for deciphering transcription factor binding and transcriptional regulation in eukaryotic cells. *Frontiers in Genetics*, 7:24, 2016.

[15] L. J. Core, A. L. Martins, C. G. Danko, C. T. Waters, A. Siepel, and J. T. Lis. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet*, 46(12):1311–1320, 12 2014.

[16] L. J. Core, J. J. Waterfall, and J. T. Lis. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909):1845–1848, December 2008.

[17] A. Curina, A. Termanini, I. Barozzi, E. Prosperini, M. Simonatto, S. Polletti, A. Silvola, M. Soldi, L. Austenaa, T. Bonaldi, S. Ghisletti, and G. Natoli. High constitutive activity of a broad panel of housekeeping and tissue-specific cis-regulatory elements depends on a subset of ETS proteins. *Genes & Development*, 31(4):399–412, Feb. 2017.

[18] D. A. Cusanovich, B. Pavlovic, J. K. Pritchard, and Y. Gilad. The functional consequences of variation in transcription factor binding. *PLoS Genet*, 10(3):e1004226, 03 2014. PMCID: PMC3945204.

[19] C. A. Davis, B. C. Hitz, C. A. Sloan, E. T. Chan, J. M. Davidson, I. Gabdank, J. A. Hilton, K. Jain, U. K. Baymuradov, A. K. Narayanan, K. C. Onate, K. Graham, S. R. Miyasato, T. R. Dreszer, J. S. Strattan, O. Jolanki, F. Y. Tanaka, and J. M. Cherry. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research*, 46(D1):D794–D801, 2018.

[20] P. Di Tommaso, M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo, and C. Notredame. Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4):316–319, Apr. 2017.

[21] A. R. R. Forrest, H. Kawaji, M. Rehli, J. Kenneth Baillie, M. J. L. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jørgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. A. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. C. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. Maxwell Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drabløs, A. S. B. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furuno, J.-i. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. Peter Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. J. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-sim, R.-i. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohmiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. D. Prendergast, O. J. L. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato,

H. Satoh, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. C. 't Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyoda, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verardo, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci, Y. Hayashizaki, and The FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature*, 507(7493):462–470, Mar. 2014.

[22] I. Gallego Romero, A. A. Pai, J. Tung, and Y. Gilad. Rna-seq: impact of rna degradation on transcript quantification. *BMC Biology*, 12(1):42, 2014.

[23] C. E. Grant, T. L. Bailey, and W. S. Noble. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018, Apr. 2011.

[24] M. A. Gruca, M. A. Gohde, and R. D. Dowell. Annotation agnostic approaches to nascent transcription analysis: Fast read stitcher and transcription fit. *Methods in Molecular Biology*, to appear, 2019.

[25] N. Hah, S. Murakami, A. Nagari, C. G. Danko, and W. L. Kraus. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Research*, 23(8):1210–1223, 2013.

[26] S. Hao and D. Baltimore. The stability of mrna influences the temporal order of the induction of genes encoding inflammatory molecules. *Nature Immunology*, 10(3):281–288, 2009.

[27] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[28] S. Jiang and A. Mortazavi. Integrating ChIP-seq with other functional genomics data. *Briefings in Functional Genomics*, 17(2):104–115, 03 2018.

[29] Q. Jin, L.-R. Yu, L. Wang, Z. Zhang, L. H. Kasper, J.-E. Lee, C. Wang, P. K. Brindle, S. Y. R. Dent, and K. Ge. Distinct roles of GCN5/PCAF-mediated H3k9ac and CBP/p300-mediated H3k18/27ac in nuclear receptor transactivation. *The EMBO Journal*, 30(2):249–262, Jan. 2011.

[30] M. Joo, J. G. Wright, N. N. Hu, R. T. Sadikot, G. Y. Park, T. S. Blackwell, and J. W. Christman. Yin Yang 1 enhances cyclooxygenase-2 gene expression in macrophages. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 292(5):L1219–1226, May 2007.

[31] S. Jüttner, T. Cramer, S. Wessler, A. Walduck, F. Gao, F. Schmitz, C. Wunder, M. Weber, S. M. Fischer, W. E. Schmidt, B. Wiedenmann, T. F. Meyer, M. Naumann, and M. Höcker. Helicobacter pylori stimulates host cyclooxygenase-2 gene transcription: critical importance of MEK/ERK-dependent activation of USF1/-2 and CREB transcription factors. *Cellular Microbiology*, 5(11):821–834, 2003.

[32] G. Karsli Uzunbas, F. Ahmed, and M. A. Sammons. Control of p53-dependent transcription and enhancer activity by the p53 family member p63. *Journal of Biological Chemistry*, 2019.

[33] I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, and V. J. Makeev. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Research*, 46(D1):D252–D259, Jan. 2018.

[34] I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, A. V. Soboleva, A. S. Kasianov, H. Ashoor, W. Ba-alawi, V. B. Bajic, Y. A. Medvedeva, F. A. Kolpakov, et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic acids research*, 44(D1):D116–D125, 2016.

[35] H. Kwak, N. J. Fuda, L. J. Core, and J. T. Lis. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, 339(6122):950–953, 2013.

[36] M. T. Lam, W. Li, M. G. Rosenfeld, and C. K. Glass. Enhancer {RNAs} and regulated transcriptional programs. *Trends in Biochemical Sciences*, 39(4):170 – 182, 2014.

[37] M. T. Y. Lam, H. Cho, H. P. Lesch, D. Gosselin, S. Heinz, Y. Tanaka-Oishi, C. Benner, M. U. Kaikkonen, A. S. Kim, M. Kosaka, C. Y. Lee, A. Watt, T. R. Grossman, M. G. Rosenfeld, R. M. Evans, and C. K. Glass. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 498(7455):511–515, Jun 2013.

[38] S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch. The Human Transcription Factors. *Cell*, 172(4):650–665, 2018.

[39] T. Lesluyes, J. Johnson, P. Machanick, and T. L. Bailey. Differential motif enrichment analysis of paired ChIP-seq experiments. *BMC Genomics*, 15(1), Sept. 2014.

[40] Q. Li, A. Su, J. Chen, Y. A. Lefebvre, and R. J. G. Hache. Attenuation of Glucocorticoid Signaling through Targeted Degradation of p300 via the 26S Proteasome Pathway. *Molecular Endocrinology*, 16(12):2819–2827, 12 2002.

[41] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550, 2014.

[42] X. Luo, M. Chae, R. Krishnakumar, C. G. Danko, and W. L. Kraus. Dynamic reorganization of the AC16 cardiomyocyte transcriptome in response to TNFα signaling revealed by integrated genomic analyses. *BMC Genomics*, 15:155–155, 2014.

[43] A. Mathelier, O. Fornes, D. J. Arenillas, C.-y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, and W. W. Wasserman. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 44(D1):D110–D115, 11 2015.

[44] I. C. McDowell, A. Barrera, A. M. D'Ippolito, C. M. Vockley, L. K. Hong, S. M. Leichter, L. C. Bartelt, W. H. Majoros, L. Song, A. Safi, D. D. Koçak, C. A. Gersbach, A. J. Hartemink, G. E. Crawford, B. E. Engelhardt, and T. E. Reddy. Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. *Genome Research*, Aug. 2018.

[45] R. C. McLeay and T. L. Bailey. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics*, 11(1):165, Apr. 2010.

[46] C. Molle, M. Goldman, and S. Goriely. Critical Role of the IFN-Stimulated Gene Factor 3 Complex in TLR-Mediated IL-27p28 Gene Expression Revealing a Two-Step Activation Process. *The Journal of Immunology*, 184(4):1784–1792, Feb. 2010.

[47] M. Muhar, A. Ebert, T. Neumann, C. Umkehrer, J. Jude, C. Wieshofer, P. Rescheneder, J. J. Lipp, V. A. Herzog, B. Reichholf, D. A. Cisneros, T. Hoffmann, M. F. Schlapansky, P. Bhat, A. von Haeseler, T. Köcher, A. C. Obenauf, J. Popow, S. L. Ameres, and J. Zuber. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science*, 360(6390):800–805, May 2018.

[48] J. Nan, Y. Wang, J. Yang, and G. R. Stark. IRF9 and unphosphorylated STAT2 cooperate with NF-κB to drive IL6 expression. *Proceedings of the National Academy of Sciences*, 115(15):3906–3911, Apr. 2018.

[49] M. T. Paulsen, A. Veloso, J. Prasad, K. Bedi, E. A. Ljungman, B. Magnuson, T. E. Wilson, and M. Ljungman. Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods*, 67(1):45–54, May 2014.

[50] A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.

[51] R. Raisner, S. Kharbanda, L. Jin, E. Jeng, E. Chan, M. Merchant, P. M. Haverty, R. Bainer, T. Cheung, D. Arnott, E. M. Flynn, F. A. Romero, S. Magnuson, and K. E. Gascoigne. Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3k27 Acetylation. *Cell Reports*, 24(7):1722–1729, Aug. 2018.

[52] S. K. Sasse, M. Gruca, M. A. Allen, V. Kadiyala, T. Song, F. Gally, A. Gupta, M. A. Pufall, R. D. Dowell, and A. N. Gerber. Nascent transcript analysis of glucocorticoid crosstalk with tnf defines primary and cooperative inflammatory repression. *Genome Research*, 2019.

[53] T. Shiraki, S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, R. Kodzius, A. Watahiki, M. Nakamura, T. Arakawa, S. Fukuda, D. Sasaki, A. Podhajska, M. Harbers, J. Kawai, P. Carninci, and Y. Hayashizaki. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences*, 100(26):15776–15781, 2003.

[54] J. Siednienko, A. Maratha, S. Yang, M. Mitkiewicz, S. M. Miggin, and P. N. Moynagh. Nuclear Factor κB Subunits RelB and cRel Negatively Regulate Toll-like Receptor 3-mediated β-Interferon Production via Induction of Transcriptional Repressor Protein YY1. *Journal of Biological Chemistry*, 286(52):44750–44763, Dec. 2011.

[55] M. Spivakov. Spurious transcription factor binding: Non-functional or genetically redundant? *BioEssays*, 36(8):798–806, 2014. PMCID: PMC4230394.

[56] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550, Oct. 2005.

[57] H. Tani, R. Mizutani, K. A. Salam, K. Tano, K. Ijiri, A. Wakamatsu, T. Isogai, Y. Suzuki, and N. Akimitsu. Genome-wide determination of rna stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Research*, 22(5):947–956, 2012.

[58] J. M. Tome, N. D. Tippens, and J. T. Lis. Single-molecule nascent rna sequencing identifies regulatory domain architecture at promoters and enhancers. *Nature Genetics*, 50(11):1533–1541, 2018.

[59] I. J. Tripodi and M. A. Gruca. Nascent-Flow, Dec. 2018.

[60] B. T. Weinert, T. Narita, S. Satpathy, B. Srinivasan, B. K. Hansen, C. Schölz, W. B. Hamilton, B. E. Zucconi, W. W. Wang, W. R. Liu, J. M. Brickman, E. A. Kesicki, A. Lai, K. D. Bromberg, P. A. Cole, and C. Choudhary. Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. *Cell*, 174(1):231–244.e12, June 2018.

[61] T. W. Whitfield, J. Wang, P. J. Collins, E. C. Partridge, S. F. Aldred, N. D. Trinklein, R. M. Myers, and Z. Weng. Functional analysis of transcription factor binding sites in human promoters. *Genome Biology*, 13(9):R50, 2012.

[62] H.-H. Xue, J. Bollenbacher-Reilley, Z. Wu, R. Spolski, X. Jing, Y.-C. Zhang, J. P. McCoy, and W. J. Leonard. The Transcription Factor GABP Is a Critical Regulator of B Lymphocyte Development. *Immunity*, 26(4):421–431, Apr. 2007.

[63] X.-C. Zhang, H.-F. Liang, X.-D. Luo, H.-J. Wang, A.-P. Gu, C.-Y. Zheng, Q.-Z. Su, and J. Cai. YY1 promotes IL-6 expression in LPS-stimulated BV2 microglial cells by interacting with p65 to promote transcriptional activation of IL-6. *Biochemical and Biophysical Research Communications*, 502(2):269–275, 2018.

Figure 1

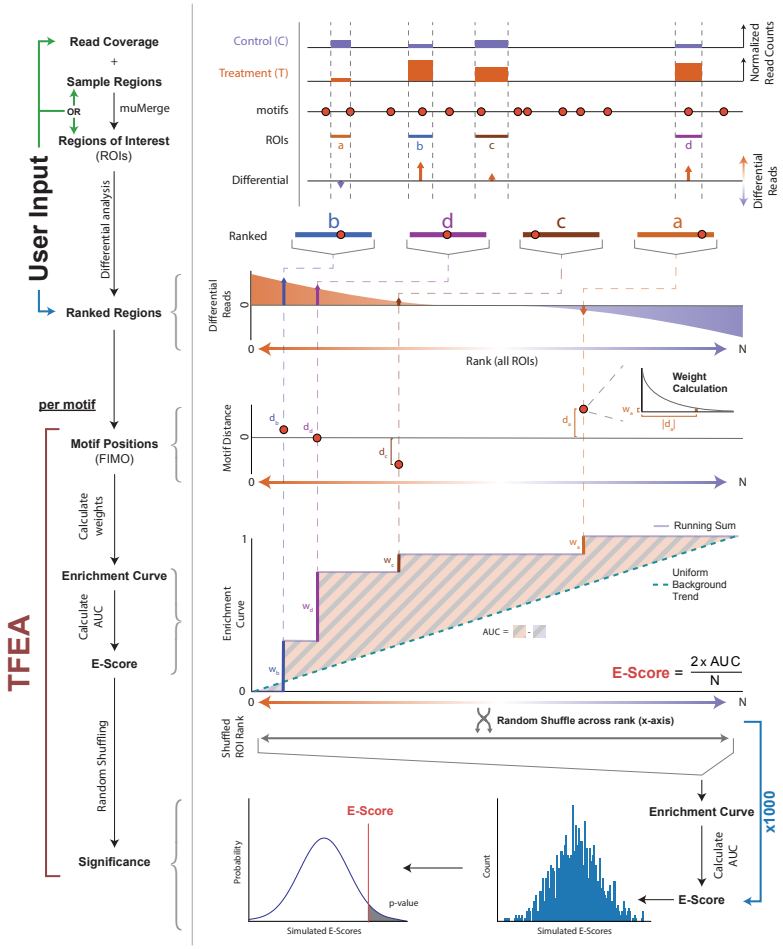


Figure 2

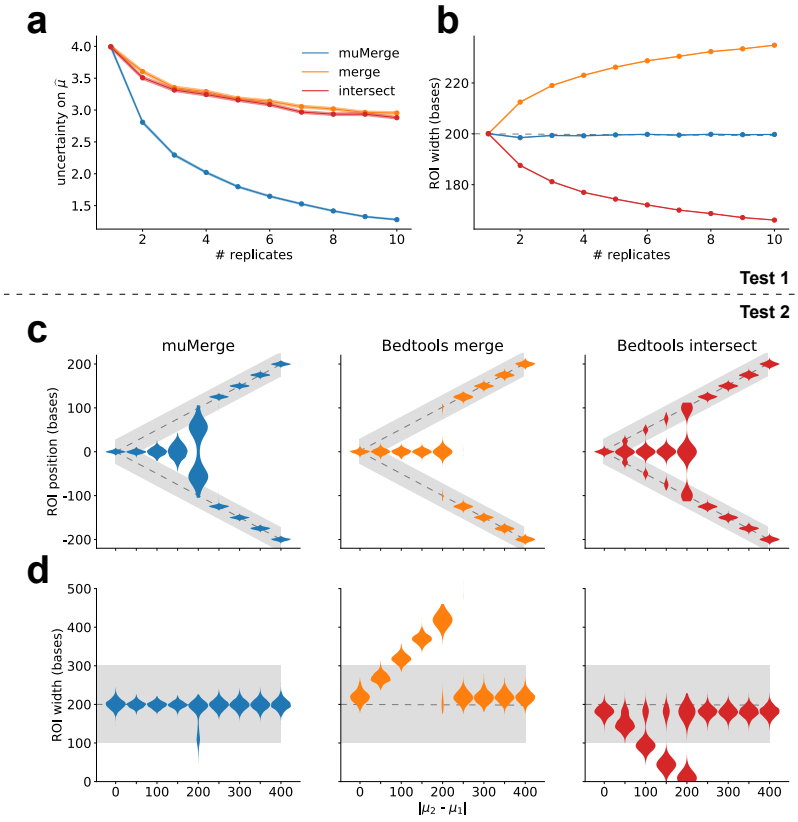


Figure 3

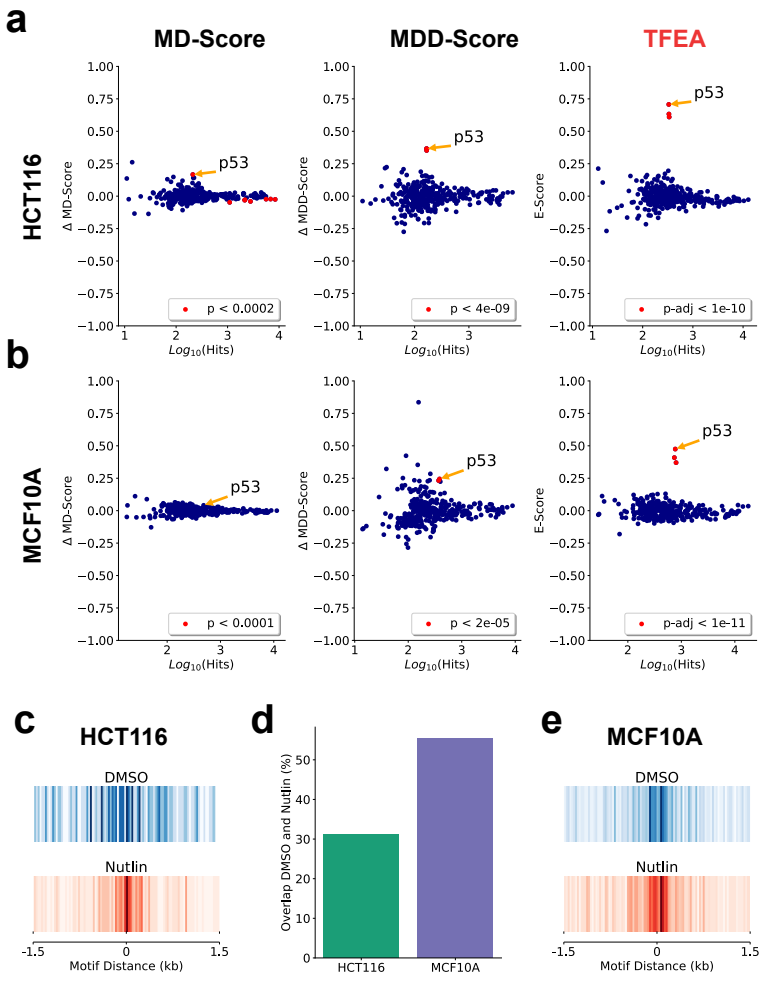


Figure 4

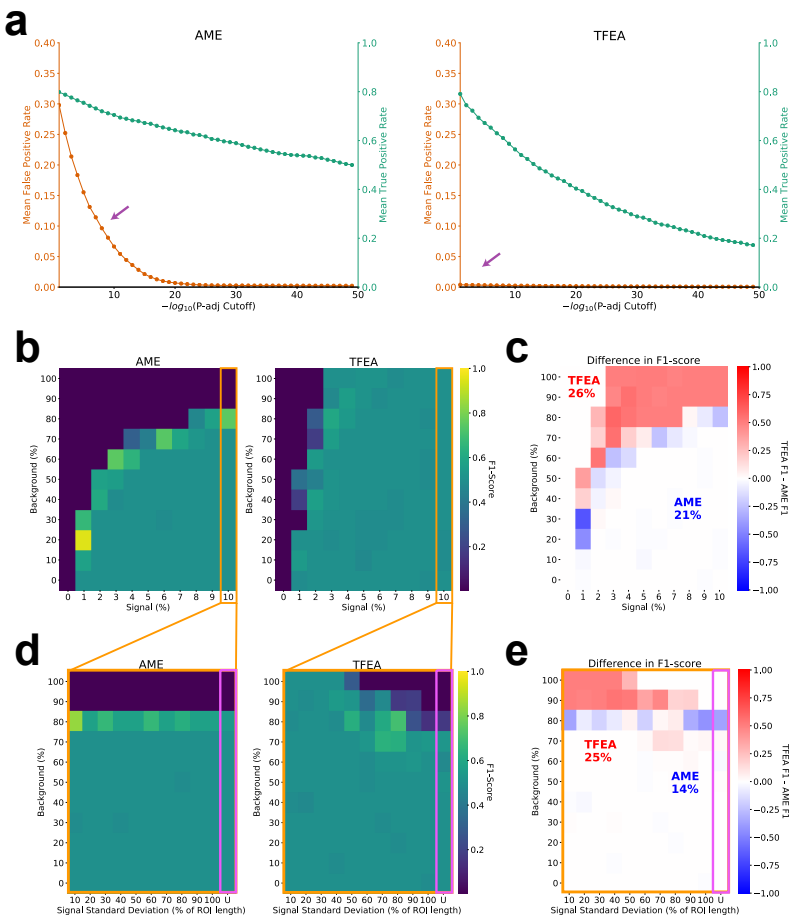


Figure 5

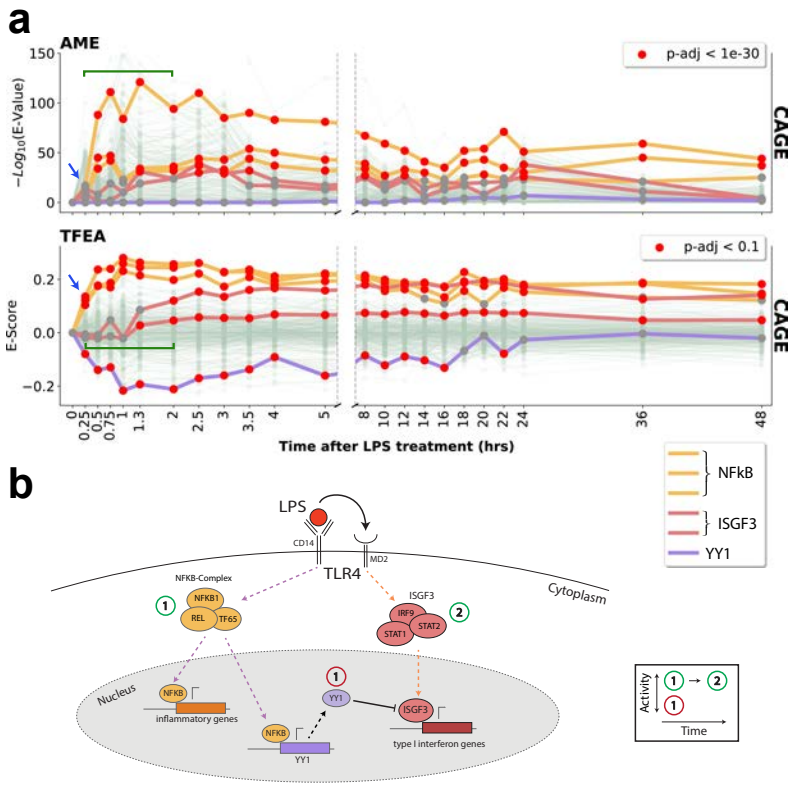
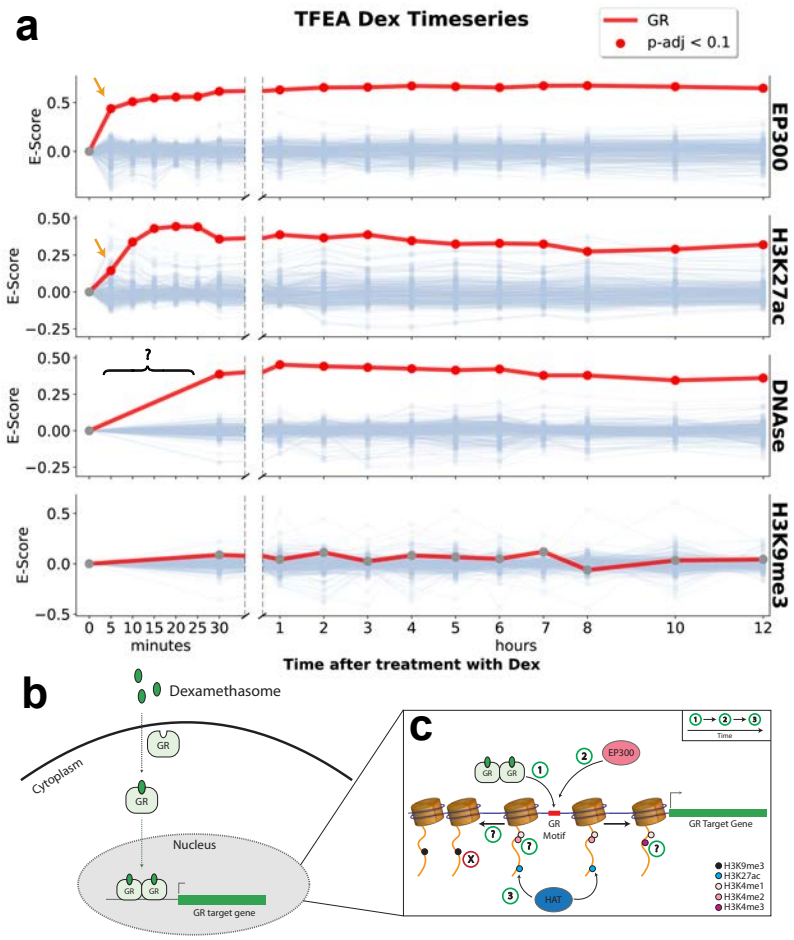


Figure 6



Chapter 4

Mediator Kinases and the Interferon Response

4.1 Preamble

While Mediator kinases have been shown to be important in disease, the role of Mediator kinase activity (in disease or normal cellular function) is not well understood. Furthermore, the extent to which both Mediator kinases (CDK8 and CDK19) serve redundant or non-redundant roles is an active area of study. Previously, studies had shown that Mediator kinases can serve as downstream regulators of signaling pathways. In support of this, transcription factors (TFs) are a large portion of known mediator targets (Poss 2016). Among the TF targets of Mediator kinases is STAT1, the main downstream effector of the IFN-γ signaling pathway. This study thus sought to probe the kinase dependent and independent roles of CDK8 and CDK19 within the IFN-γ response.

4.1.1 Significance

The IFN-γ signaling pathway is involved in the response to bacterial and viral infection. This essential cellular process is highly regulated as hyper or hypo activation can lead to autoimmune disease or immunocompromise. Furthermore, this pathway plays essential roles in cancer as this represents a major method for tumor surveillance. The IFN-γ signalling pathway works by detection of IFN-γ through cell surface receptors which causes a signaling cascade ultimately leading to the phosphorylation of TF signal transducer and activator of transcription 1 (STAT1) which triggers its translocation into the nucleus where it effects changes in target gene transcription.

Mediator is a large protein complex responsible for integrating signals from TF binding events

to the pre-initiation complex (PIC) containing RNA Polymerease II (RNAPII). Mediator kinases (CDK8 and CDK19) can reversibly associate with Mediator and are known to play a role in disease, specifically cancer. This study probed the transcriptional and gene expression effects of inhibiting Mediator kinase activity through small molecule addition or chemical genetics and depleting protein abundance using inducible knockouts or knockdowns. Surprisingly, CDK8 and CDK19 played non-redundant roles within the IFN-γ signaling pathway with the response affected by CDK8 activity and/or CDK19 protein. Further, this study showed that CDK19 knockdown significantly affected the cellular response to infection.

The results of this study were published on September 5, 2019 in Molecular Cell Volume 76, Issue 3, Pages 485-499 (https://doi.org/10.1016/j.molcel.2019.07.034).

4.1.2 Contributions

The majority of this work was performed by Dr. Steinparzer. I began working on this project when Dr. Steinparzer visited the lab for several months (a year?). She was a graduate student at the time interested in performing nascent sequencing on mouse embryonic fibroblasts (MEFs) treated with IFN-γ and the Mediator kinase inhibitor CA. I loosely guided Dr. Steinparzer on how to perform nuclear isolation and the GRO-Seq protocol, mainly answering questions related to established protocols. Following her departure from the lab and the sequencing of all her samples, I began a preliminary analysis on the MEF GRO-Seq data. I performed DE-Seq analysis and pause analysis via the quantification of a pause index (PI). I found that the PI of IFN-γ genes was elevated in cells treated with IFN-γ and CA compared to those treated with IFN-γ and DMSO. Dr. Sedlyarov then performed the final analysis confirming these results.

Dr. Steinparzer then performed many experiments including an array of shRNA knockdown RNA-Seq experiments showing that the IFN-γ response could be affected by CDK8 inhibition and CDK19 protein knockdown. Additionally, Dr. Levandowski along with Dr. Galbraith, and Dr. Andyrsik (others?) performed PRO-Seq experiments in human HCT116 cells. I, along with Dr. Sedlyarov, processed the PRO-Seq data (with Dr. Sedlyarov also processing the RNA-Seq data).

Using this data and the previous MEF GRO-Seq data, I performed a modified version of the MD-Score approach using only the top 20% of significantly differentially transcribed bidirectionals (more details in method section). I found that upon induction with IFN-γ both MEFs and HCT116 cells exhibited activation of the canonical STAT and IRF motifs. Upon simultaneous treatment with CA, these TFs were no longer activated (or activated to a lesser extent) after IFN-γ treatment. Finally, I performed all of the GSEA analyses in the paper and produced raw data snapshots of key genes that appear in the supplement.

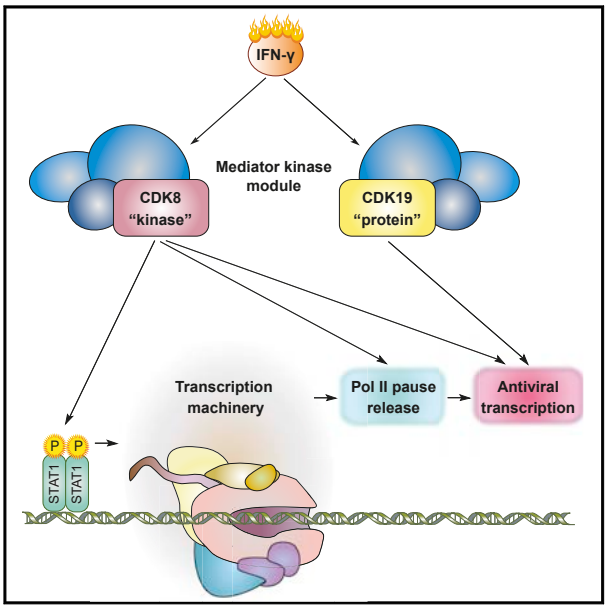
This project was an international collaboration of many scientists resulting in a rigorous probe into the functions of both Mediator kinases within the IFN-γ response. The published manuscript follows.

Article

Molecular Cell

Transcriptional Responses to IFN-γ Require Mediator Kinase-Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions

Graphical Abstract



Highlights

- Mediator kinases CDK8 and CDK19 are distinct transcription regulators in the IFN-γ pathway
- CDK8 acts as a kinase while CDK19 functions as a scaffold in driving IFN-γ gene profiles
- CDK8 and CDK19 activate distinct gene sets
- Mediator kinase CDK8 promotes polymerase II pause release during the IFN-γ response

Authors

Iris Steinparzer, Vitaly Sedlyarov, Jonathan D. Rubin, ..., Robin D. Dowell, Dylan J. Taatjes, Pavel Kovarik

Correspondence

taatjes@colorado.edu (D.J.T.), pavel.kovarik@univie.ac.at (P.K.)

In Brief

Steinparzer et al. discover that the Mediator kinases CDK8 and CDK19 are functionally and mechanistically distinct transcription regulators in the IFN-γ antiviral program: CDK8 and CDK19 activate distinct gene sets, and they do so in kinase-dependent (CDK8) and kinase-independent (CDK19) ways. Further, CDK8 kinase activity promotes RNAPII pause release.

Steinparzer et al., 2019, Molecular Cell *76*, 485–499
November 7, 2019
https://doi.org/10.1016/j.molcel.2019.07.034

CellPress

Molecular Cell
Article

Transcriptional Responses to IFN-γ Require Mediator Kinase-Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions

Iris Steinparzer,[1] Vitaly Sedlyarov,[2] Jonathan D. Rubin,[3] Kevin Eislmayr,[1] Matthew D. Galbraith,[4,5] Cecilia B. Levandowski,[3] Terezia Vcelkova,[1] Lucy Sneezum,[1] Florian Wascher,[1] Fabian Amman,[1,6] Renata Kleinova,[1] Heather Bender,[5] Zdenek Andrysik,[5] Joaquin M. Espinosa,[4,5] Giulio Superti-Furga,[2,7] Robin D. Dowell,[8,9] Dylan J. Taatjes,[3,*] and Pavel Kovarik[1,10,*]

[1]Max Perutz Labs, University of Vienna, Vienna Biocenter (VBC), Dr. Bohr-Gasse 9, Vienna, Austria
[2]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria
[3]Department of Biochemistry, University of Colorado, Boulder, CO 80303, USA
[4]Linda Crnic Institute for Down Syndrome, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
[5]Department of Pharmacology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
[6]Department of Theoretical Chemistry of the University of Vienna, 1090 Vienna, Austria
[7]Center for Physiology and Pharmacology, Medical University of Vienna, Vienna, Austria
[8]BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA
[9]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA
[10]Lead Contact
*Correspondence: taatjes@colorado.edu (D.J.T.), pavel.kovarik@univie.ac.at (P.K.)
https://doi.org/10.1016/j.molcel.2019.07.034

SUMMARY

Transcriptional responses to external stimuli remain poorly understood. Using global nuclear run-on followed by sequencing (GRO-seq) and precision nuclear run-on sequencing (PRO-seq), we show that CDK8 kinase activity promotes RNA polymerase II pause release in response to interferon-γ (IFN-γ), a universal cytokine involved in immunity and tumor surveillance. The Mediator kinase module contains CDK8 or CDK19, which are presumed to be functionally redundant. We implemented cortistatin A, chemical genetics, transcriptomics, and other methods to decouple their function while assessing enzymatic versus structural roles. Unexpectedly, CDK8 and CDK19 regulated different gene sets via distinct mechanisms. CDK8-dependent regulation required its kinase activity, whereas CDK19 governed IFN-γ responses through its scaffolding function (i.e., it was kinase independent). Accordingly, CDK8, not CDK19, phosphorylates the STAT1 transcription factor (TF) during IFN-γ stimulation, and CDK8 kinase inhibition blocked activation of JAK-STAT pathway TFs. Cytokines such as IFN-γ rapidly mobilize TFs to "reprogram" cellular transcription; our results implicate CDK8 and CDK19 as essential for this transcriptional reprogramming.

INTRODUCTION

Understanding how pathway-specific transcriptional responses are controlled remains an important but challenging endeavor. The transcriptional response to the ubiquitously acting cytokine interferon-γ (IFN-γ) protects against bacterial and viral infection and is indispensable for tumor surveillance (Schneider et al., 2014). Whereas the basic signaling components of the IFN-γ pathway are defined, the chromatin-associated processes that adjust the transcriptional output to physiological needs are incompletely understood.

Upon IFN-γ binding to its receptor, the tyrosine kinases JAK1 and JAK2 phosphorylate the transcription factor (TF) signal transducer and activator of transcription 1 (STAT1), causing its nuclear translocation (Schneider et al., 2014). STAT1 is tightly controlled to prevent pathologies associated with hyper- or hypo-activation. An important control mechanism targets the JAKs in the cytoplasm (Schneider et al., 2014); another layer of control occurs in the nucleus. For example, STAT1 promoter occupancy is limited by negative feedback that initiates once productive transcription complexes have been established (Wiesauer et al., 2015). Chromatin-associated STAT1 is also regulated by phosphorylation of its activation domain (AD) at residue S727 (Sadzak et al., 2008; Wen et al., 1995). This physiologically important modification adjusts the transcriptional output in gene-specific ways (Bancerek et al., 2013).

The RNA polymerase II (RNAPII) enzyme transcribes all protein-coding and most non-coding RNAs in the human genome, and Mediator appears to be required to activate RNAPII function genome-wide (Allen and Taatjes, 2015). The mammalian Mediator complex consists of 26 subunits and a 600-kDa, four-subunit kinase module consisting of MED12, MED13, CCNC, and CDK8 (or CDK19) that can reversibly associate with Mediator (Allen and Taatjes, 2015). Consequently, CDK8 and CDK19 are considered Mediator-associated kinases. CDK8 and CDK19 are highly similar (77% amino acid sequence identity) paralogs, and each appears to associate in a mutually exclusive fashion with the Mediator kinase module (Galbraith et al., 2013).

The transcriptional effects of Mediator kinases are generally cell-type and context specific (Galbraith et al., 2013; Johannessen et al., 2017). These characteristics are consistent with the fact that sequence-specific, DNA-binding TFs represent a major class of proteins that are targeted by CDK8 and CDK19 (Poss et al., 2016), and CDK8 and/or CDK19-dependent phosphorylation of these TFs has been shown to alter TF activity in a few well-studied cases (Bancerek et al., 2013; Nitulescu et al., 2017). Notably, STATs appear to represent a more common target of Mediator kinases, as they have been shown to be phosphorylated by CDK8 and/or CDK19 in diverse cell lineages (Bancerek et al., 2013; Nitulescu et al., 2017; Poss et al., 2016). These findings suggest that Mediator kinase regulation of STAT TF function may be a common theme in biology.

Prior work has implicated Mediator kinases as promising targets for therapeutic manipulation of cytokine responses (Bancerek et al., 2013; Chen et al., 2017; Johannessen et al., 2017; Nitulescu et al., 2017). These studies have focused on CDK8 and used knockdown approaches. Thus, the role of CDK19 versus CDK8 was not assessed, and the requirement for the kinase activity per se (i.e., in contrast to the physical presence of CDK8) remained undefined. Among the Mediator kinase inhibitors that have been characterized (Dale et al., 2015; Johannessen et al., 2017; Koehler et al., 2016), the natural product cortistatin A (CA) stands out based upon its potency and high selectivity for CDK8 and CDK19 (Pelish et al., 2015). Kinome-wide screens revealed essentially no off-target kinase inhibition, even at concentrations that were orders of magnitude higher than its *in-vitro*-measured K_d (0.2 nM) (Pelish et al., 2015). Here, we used CA in combination with chemical genetics and conditional knockout or knockdown methods to thoroughly evaluate the regulatory roles of CDK8 and CDK19 in the context of IFN-γ signaling and antiviral responses. Our results define kinase-specific and kinase-independent roles for each and establish CDK8 and CDK19 as essential but nonredundant regulators of IFN-γ transcriptional responses.

RESULTS

Mediator Kinase Activity Is a Gene-Selective Regulator of IFN-γ-Stimulated Transcription

Our previous study demonstrated that CDK8 controlled transcriptional responses to IFN-γ (Bancerek et al., 2013), but it remained unclear whether its effects were kinase dependent or kinase independent. To address this question, we employed CA, a potent and highly selective inhibitor of the Mediator kinases CDK8 and CDK19 (Pelish et al., 2015). As expected, IFN-γ-stimulated phosphorylation of STAT1 at Y701 was not affected by CA treatment, whereas CA inhibited IFN-γ-induced phosphorylation of STAT1 S727 (Figure 1A), consistent with previous findings in human cells (Bancerek et al., 2013; Nitulescu et al., 2017; Pelish et al., 2015).

To investigate the effects of CDK8 and CDK19 kinase activity on the transcriptome, mouse embryonic fibroblasts (MEFs) were treated for 1 h with CA (100 nM, or DMSO control) followed by stimulation with IFN-γ for 6 h. Ribosomal RNA-depleted total RNA from these cells was subjected to RNA sequencing (RNA-seq); expression (fragments per kilobase million [FPKMs]) for exonic (i.e., mRNA) and intronic (i.e., pre-mRNA) reads were calculated separately, as described previously (Madsen et al., 2015), with modifications (see STAR Methods). Principal-component analysis (PCA) and analysis of counts normalized for library size and composition (Love et al., 2014) revealed that replicate 1 of an IFN-γ-stimulated triplicate was an outlier (Figures S1A, left panel, and Figure S1B) and was omitted from subsequent analyses.

Treatment of MEFs for 6 h with IFN-γ increased the mRNA levels of 221 genes ($\log_2$-fold-change [lfc] ≥ 1; adjusted p value [padj] < 0.05; FPKM ≥ 1) and downregulated mRNA levels of only 3 genes (lfc ≤ -1; padj < 0.05; FPKM ≥ 1; Figure 1B; Table S1A). Similar numbers and largely overlapping sets of genes were induced by IFN-γ at the pre-mRNA level (Figures 1C and S1C; Table S1B), indicating that the increased mRNA levels resulted not from changes in mRNA stability but rather from transcriptional stimulation by IFN-γ, as reported previously (Bancerek et al., 2013; Dolken et al., 2008). Of the 221 genes induced by IFN-γ, 38 were less induced (lfc < 0, padj < 0.05) and 7 more strongly induced (lfc > 0, padj < 0.05) in cells treated with CA (Figure 1D; Table S1A). Comparable results were obtained at the pre-mRNA level (31 and 5, respectively; Figure 1E; Table S1B). As expected, CA treatment did not affect cell viability within the time frame of these experiments (Figure S1D). Together, these results indicated that inhibition of Mediator kinase activity primarily reduced IFN-γ-stimulated transcription but had gene-selective effects.

Mediator Kinases Act in Part through STAT1 S727 Phosphorylation

STAT1 S727 is an established CDK8 kinase target (Bancerek et al., 2013; Pelish et al., 2015). To find out whether the effects of CA on induction of IFN-γ-regulated genes were dependent on STAT1 S727 phosphorylation by Mediator kinases, we compared responses of MEFs derived from STAT1 S727A knockin mice (Bancerek et al., 2013) with those expressing wild-type (WT) STAT1. Stimulation of S727A MEFs with IFN-γ (3 h) induced a robust response (Figures 1F and S1E; Table S2). Compared to WT MEFs, the induction of many IFN-γ target genes was altered in S727A MEFs (Figure 1G; Table S2), consistent with previous studies (Bancerek et al., 2013). CA treatment predominantly decreased induction in both WT and S727A MEFs (Figures S1F and S1G; Table S2), and most of the CA-sensitive genes were similarly affected in both WT and S727A MEFs (Figure 1H; Table S2). These data indicate that Mediator kinases act not only through STAT1 but also through other proteins to help regulate the transcriptional response to IFN-γ.

Mediator Kinase Inhibition Increases RNAPII Pausing at IFN-γ-Regulated Genes

Given that CA treatment modulated IFN-γ-stimulated transcription (Figure 1), we asked whether CA effects could be linked to any specific stage of transcription. We employed global nuclear run-on followed by sequencing (GRO-seq), a well-tested method that detects actively transcribing polymerases and measures nascent transcription, genome-wide (Core et al., 2008). A short stimulation (30 min) with IFN-γ allowed us to assess the effects

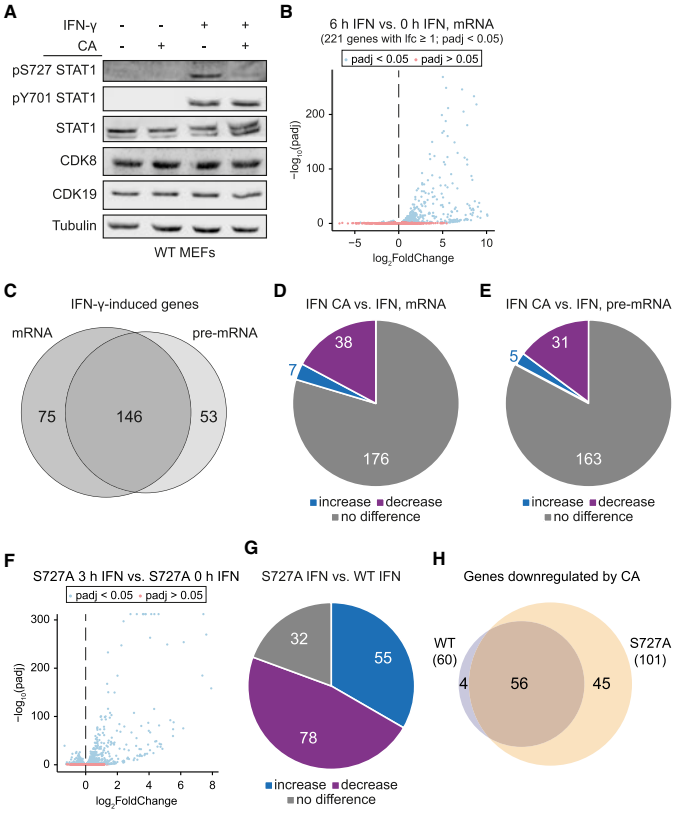


Figure 1. Mediator Kinase Inhibition Impairs IFN-γ-Stimulated Transcription in Gene-Selective Ways

(A) CA inhibits IFN-γ-induced STAT1 AD phosphorylation at S727. WT MEFs (±100 nM CA, 1-h pretreatment) were subjected to 45-min IFN-γ stimulation followed by western blot against phosphorylated STAT1 at S727 (pS727) or Y701 (pY701) and total STAT1, CDK8, CDK19, or tubulin. STAT1 bands correspond to STAT1α or STAT1β isoforms.

(B) Gene expression changes (mRNA) in WT MEFs upon 6-h IFN-γ treatment (blue, padj < 0.05; red, padj > 0.05). Genes with padj < 0.05, $\log_2$ fold change (lfc) ≥ 1, FPKM stimulated ≥ 1 were regarded as IFN-γ induced (221 genes).

(C) Overlap of genes induced after 6-h IFN-γ at mRNA versus pre-mRNA levels.

(D and E) Effects of CA on expression of IFN-γ-induced genes. WT MEFs (±100 nM CA, 1-h pretreatment) were stimulated with IFN-γ (6 h) followed by RNA-seq. The numbers of differentially expressed (upregulated: lfc > 0, padj < 0.05; downregulated: lfc < 0, padj < 0.05) IFN-γ-induced genes (defined in B) at mRNA (D) and pre-mRNA (E) level are shown.

(F–H) Mediator kinases act in part through STAT1 S727 phosphorylation.

(F) Gene expression (mRNA) changes in S727A MEFs upon 3-h IFN-γ treatment.

(G) Effects of STAT1 S727A mutation on IFN-γ-induced genes: genes induced by IFN-γ in WT MEFs are upregulated (55 genes, lfc > 0, padj < 0.05), downregulated (78 genes, lfc < 0, padj < 0.05), or unaffected (32 genes) in S727A MEFs.

(H) Overlap of IFN-γ-induced genes downregulated (lfc < 0, padj < 0.05) by CA (100 nM CA, 1-h pretreatment) in WT MEFs versus S727A MEFs.

of Mediator kinase activity on the primary (i.e., STAT1-driven) IFN-γ response.

We completed biological replicate GRO-seq experiments in MEFs following 30-min treatment with IFN-γ upon pretreatment with CA (100 nM, 1 h, or DMSO control), as with mRNA analyses. Gene set enrichment analysis (GSEA) confirmed induction of expected IFN-γ-responsive gene sets, with significant enrichment (false discovery rate [FDR] q-value < 0.05) of IFN-γ and JAK-STAT signaling pathways (Figure 2A; Table S3A). The total number of genes induced after 30-min IFN-γ stimulation was 200 (lfc > 0; padj < 0.05; Table S4A), similar to the number of genes induced at the mRNA level after 6 h (Figure 1B). The overlap between IFN-γ-induced genes in GRO-seq (30 min IFN-γ) and RNA-seq (6 h IFN-γ) experiments comprised 51 genes (Figure S2A). CA treatment resulted in reduced induction of a large proportion of IFN-γ target genes (Figure 2B; Table S4B), consistent with CA effects on mRNA levels during 6-h stimulation with IFN-γ (Figures 1D and 1E).

The GRO-seq read counts across *Irf1* (RefSeq GenBank: NM_008390), a STAT1 target gene (Bancerek et al., 2013), revealed RNAPII pausing prior to IFN-γ stimulation, as evident from pausing index (PI) of 2.03 calculated from pooled replicates (see STAR Methods) (Figure 2C; Table S4C). Overlay of reads confirmed similar changes in both replicates (Figure S2B). Paused RNAPII was released upon IFN-γ stimulation, as revealed by almost evenly distributed read counts across the *Irf1* gene and concomitant reduction in PI (PI = 0.69; Figures 2C and S2B). However, Mediator kinase inhibition with CA blocked this process (PI = 1.78; Figures 2D and S2B; Table S4C). The CA-mediated increase in PI correlated with lower induction of *Irf1* transcripts (199.5 versus 117.5 transcripts per million reads [TPM]; Figure 2D) and IRF1 protein (Figure S2C). Similar regulation was observed also for Tap1, another STAT1 target gene (Bancerek et al., 2013) (PI = 5.6 for unstimulated control, PI = 1.82 for 30-min IFN-γ, and PI = 2.83 for 30-min IFN-γ plus CA; Table S4C).

The CA-mediated PI increase at *Irf1* suggested that Mediator kinase inhibition prevented release of RNAPII pausing during IFN-γ stimulation. In agreement, transcriptome-wide analysis of CA-treated IFN-γ-stimulated samples versus IFN-γ-stimulated

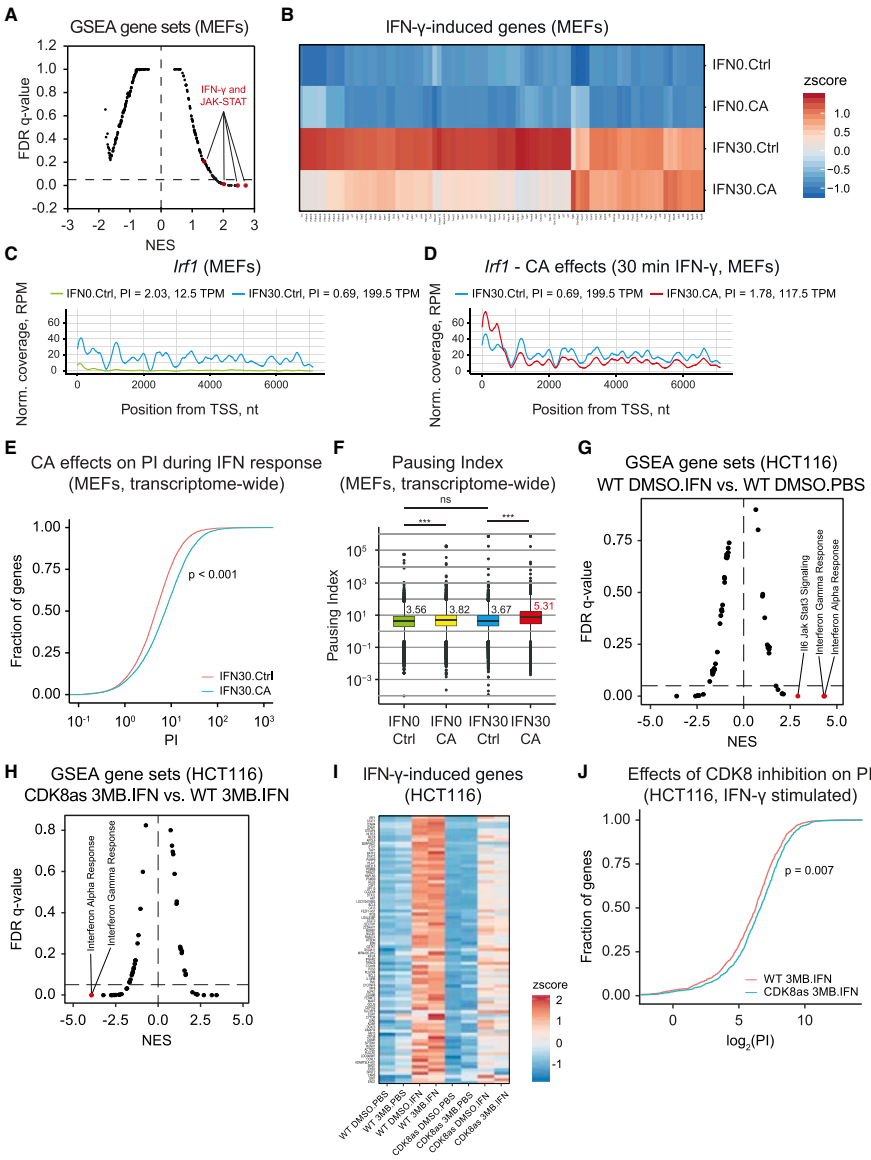


Figure 2. Mediator Kinase Inhibition Increases RNAPII Pausing

(A–F) MEFs pretreated with CA (100 nM, 1 h) or DMSO (Ctrl) were stimulated with IFN-γ (30 min; IFN30) or unstimulated (IFN0) and subjected to GRO-seq.

(A) Moustache plot of false discovery rate (FDR) versus normalized enrichment score (NES) based upon GSEA of GRO-seq data for IFN30 versus IFN0. Dashed line, 0.05 FDR. Only positively enriched gene sets are found at FDR < 0.05. Gene sets for IFN-γ and JAK-STAT pathways are highlighted.

(B) Effects of CA on induction of IFN-γ target genes (lfc ≥ 1, padj < 0.05).

(legend continued on next page)

DMSO controls revealed a shift to higher PIs in CA-treated cells (Figures 2E and S2D; Table S4C). Analysis of PIs transcriptome-wide by the Mann-Whitney *U* test (R Core Team, 2017) also confirmed that CA treatment significantly increased PIs during IFN-γ stimulation (PI = 5.31 versus 3.82; Figure 2F). PI analysis of the gene group induced by IFN-γ revealed decreased pausing upon stimulation (Figure S2E), consistent with PI changes at *Irf1* (Figure 2C). Interestingly, 19 of the 38 IFN-γ-induced genes whose mRNA levels were downregulated by CA (Figure 1D) showed a corresponding increase in PI with CA treatment (Table S4D). These data demonstrate Mediator kinase-dependent RNAPII pause release during IFN-γ-induced transcription.

Chemical Genetics Confirms RNAPII Pause Regulation by CDK8 Kinase

We next employed a CDK8-analog-sensitive (CDK8as) HCT116 human cell line (Galbraith et al., 2017), which enabled us to define CDK8-specific effects while testing in a different model system (i.e., human versus mouse). WT and CDK8as cells were induced with IFN-γ for 45 min, followed by nuclei isolation for PRO-seq analysis (Kwak et al., 2013). In addition to unstimulated controls, we completed experiments in the presence or absence of the CDK8as inhibitor 3MB-PP1 (Galbraith et al., 2017). As expected, IFN-γ induced IFN-responsive gene sets in HCT116 cells (Figure 2G; Table S3B). The CDK8as cells behave as hypomorphs (Galbraith et al., 2017); consistently, IFN-γ response in CDK8as cells was diminished compared to WT cells (Figure S2F; Table S3B). Furthermore, inhibition of CDK8 kinase activity with the ATP analog 3MB-PP1 strongly reduced the IFN-γ response in CDK8as cells as compared to WT cells treated with 3MB-PP1 (Figure 2H; Table S3B). Reduced induction of IFN-γ response genes with inhibition of CDK8 was similar to the effects in IFN-γ-stimulated MEFs treated with CA. Heatmaps from the PRO-seq data demonstrated that 3MB-PP1 treatment did not impair the induction of IFN-γ response genes in WT cells; however, 3MB-PP1 caused reduction of the hypomorph IFN-γ response in CDK8as cells (Figure 2I; Tables S5A and S5B).

In IFN-γ stimulated MEFs, we observed a correlation between increased PI and reduced gene expression in CA-treated cells (Table S4D). Consistent with these results, a PI analysis from the PRO-seq data in IFN-γ stimulated HCT116 cells showed a similar trend upon CDK8 kinase inhibition during the IFN-γ response (Figures 2J and S2G; Table S5C). In the absence of IFN-γ stimulation, CDK8 inhibition did not increase the PI in this group of genes (Figure S2H; Table S5C). Taken together, the data summarized in Figure 2 suggest that (1) Mediator kinase activity contributes to release of paused RNAPII at IFN-γ-induced genes, (2) inhibition of Mediator kinases increases RNAPII pausing in human cells and MEFs, and (3) reduced induction of IFN-γ target genes upon Mediator kinase inhibition correlates with defects in RNAPII pause release.

Enhancer RNA (eRNA) Transcription Implicates Specific TFs and Mediator Kinases in the IFN-γ Response

Enhancer-associated transcription appears to represent the most rapid transcriptional response to a stimulus (Arner et al., 2015). Enhancer activity correlates with the expression of unstable, bidirectional transcripts defined as eRNAs. Although the function of eRNAs remains unclear, their abundance can infer TF activity (Azofeifa et al., 2018). For instance, bidirectional eRNAs originate around sites of TF binding, and if a consensus sequence is defined, these sites can reliably predict active TFs (or TFs being repressed) at the time of the analysis (Azofeifa et al., 2018).

The eRNA profile associated with IFN-γ response has not been addressed, nor has the role of Mediator kinases in eRNA expression. To this end, we measured TF motif displacement (MD) across eRNAs in MEFs, as described previously (Azofeifa et al., 2018), with minor improvements (see STAR Methods). Quantifying TF activity associated with the IFN-γ response (no IFN-γ versus 30-min IFN-γ), we observed a significant increase in MD score of IFN-related TF motifs (STAT1 and STAT5B; Figures 3A and S3A; Table S6A). The data indicated that eRNA transcripts originating from these TF motifs were induced following IFN-γ treatment. In contrast, eRNAs associated with these same TF motifs (i.e., STAT1 and STAT5) were not induced by IFN-γ if Mediator kinase activity was inhibited by CA (Figures 3B and S3A; Table S6B). These results suggest that IFN-γ-responsive STAT TFs are activated, at least in part, through Mediator kinase function. In agreement, induction of IFN-γ response genes is reduced in CA-treated MEFs (Figure 1).

Similar to results in MEFs, PRO-seq data collected in control or IFN-γ-stimulated (45 min) HCT116 cells revealed activation of TFs known to be involved in the IFN response (e.g., STAT1/2 and IRF1-3; Figures 3C and S3B; Table S6C); moreover, eRNAs associated with these TFs were not induced by IFN-γ

(C and D) Plot of GRO-seq reads (pooled replicates) at *Irf1* locus for IFN0.Ctrl and IFN30.Ctrl (C) as well as IFN30.Ctrl and IFN30.CA (D). Pausing indices (PIs) and transcripts per million reads (TPM; nt 501 to transcript end) are indicated.

(E) Empirical cumulative density function (ECDF) plot of PI distribution (transcriptome-wide) under IFN30.Ctrl (red) and IFN30.CA (blue) conditions. Kolmogorov-Smirnov test p value < 2.2e-16.

(F) Median PI and statistical assessment of PI changes for all expressed genes (RefSeq). Median PI value is shown for each condition (red is the highest). Mann-Whitney *U* test, p value not significant (ns) ≥ 0.05; *p < 0.05; **p < 0.01; ***p < 0.001.

(G–J) HCT116 WT and CDK8as cells were stimulated with IFN-γ (IFN) for 45 min (or unstimulated, PBS) and simultaneously treated with 10 μM 3MB-PP1 (3MB) or DMSO, and subsequently subjected to PRO-Seq.

(G) GSEA for IFN-γ response of WT HCT116 cells (DMSO.IFN versus DMSO.PBS).

(H) GSEA for CDK8 inhibition in IFN-γ-stimulated HCT116 cells (CDK8as 3MB.IFN versus WT.3MB.IFN).

(I) Effects of CDK8 inhibition (CDK8as, 3MB) on expression of IFN-γ-induced genes compared to WT cells. Both IFN-γ-stimulated (IFN) and unstimulated (PBS) as well as 3MB- and control (DMSO)-treated cells were analyzed. IFN-γ-induced genes (N = 83): padj < 0.1, lfc > 1 for WT DMSO.IFN versus WT DMSO.PBS using gene body (+301 to end) counts.

(J) PI distribution during IFN-γ response upon CDK8 inhibition (CDK8as 3MB.IFN, blue) versus uninhibited control (WT 3MB.IFN, red). Distribution shown for genes downregulated by CDK8 inhibition (N = 956, padj < 0.1, $\log_2$ fold change < 1 for CDK8as 3MB.IFN versus WT 3MB.IFN).

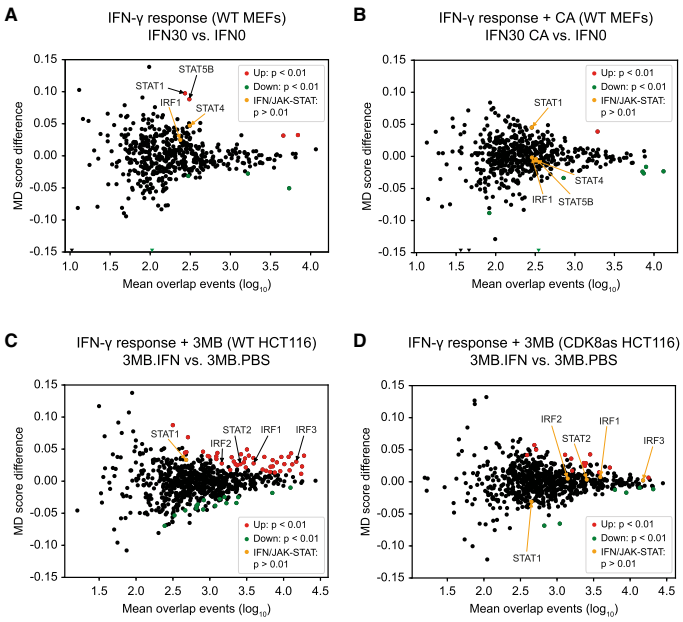


Figure 3. eRNA Transcription Predicts Activation of Specific TFs and Role of Mediator Kinases in the IFN-γ Response

(A–D) Motif displacement (MD) score during the IFN-γ (30 min) response in MEFs (A and B) and HCT116 cells (C and D), derived from GRO-seq (MEFs) and PRO-seq (HCT116) data.

(A) MD score difference for TFs in IFN-γ-stimulated (IFN30) versus unstimulated (IFN0) MEFs (IFN30 versus IFN0). STAT1 and STAT5b motifs are significantly enriched upon IFN-γ stimulation.

(B) Effect of CA treatment on MD score during the IFN-γ response (IFN30.CA versus IFN0). STAT1 and STAT5b motifs enriched in (A) are not enriched upon CA treatment.

(C) MD score difference for TFs in IFN-γ-stimulated (IFN) versus unstimulated (PBS) WT HCT116 cells with 3MB-PP1 treatment (3MB.IFN versus 3MB.PBS). TF motifs for IFN and JAK-STAT pathways (IRF1, IRF2, IRF3, and STAT2) are significantly enriched upon IFN-γ stimulation.

(D) Effect of CDK8 inhibition on MD score during the IFN-γ response (CDK8as 3MB.IFN versus 3MB.PBS). TF motifs for the IFN and JAK-STAT pathways (IRF1, IRF2, IRF3, and STAT2) enriched in (C) are not enriched upon CDK8 inhibition.

in CDK8-inhibited cells (Figures 3D and S3B; Table S6D). These results suggest that Mediator kinase function during IFN-γ response is conserved in mouse and human cells and specifically identifies a role for CDK8 kinase activity in human cells. In contrast to MEFs, numerous other TFs (i.e., beyond STATs and IRFs) were significantly activated with IFN-γ treatment in HCT116 cells, and these also lacked evidence of activation upon CDK8 kinase inhibition (Figures 3C and 3D; Tables S6C and S6D). Such results suggest novel roles for these TFs in the IFN-γ response. Collectively, the data from Figure 3 (1) implicate a set of TFs that are rapidly activated upon IFN-γ stimulation in mouse and human cells and (2) suggest that Mediator kinase activity is required for proper activation of these IFN-γ-induced TFs.

CDK8, Not CDK19, Is the Major IFN-γ-Activated STAT1 AD Kinase

CA revealed Mediator-kinase-dependent effects on IFN-γ-induced genes (Figures 1 and 2) and experiments with CDK8as cells confirmed the involvement of CDK8 kinase activity (Figure 2); however, the role of CDK19 remained unclear, because CA inhibits both CDK8 and CDK19 (Pelish et al., 2015). To address the individual contribution of CDK8 and CDK19, we used immortalized MEFs derived from CreERT2-*Cdk8*$^{fl/fl}$ animals. These cells endogenously express CDK8 and CDK19 and allow conditional depletion of each (see below).

To address the role of CDK8, CreERT2-*Cdk8*$^{fl/fl}$ MEFs (referred to as CDK8fl-MEF) were treated with 4-hydroxytamoxifen (4OHT) for 3 h followed by 48-h recovery without 4OHT. This resulted in inducible *Cdk8* knockout (CDK8-iKO) (Figure S4A) and loss of the CDK8 protein (Figure 4A). As expected, CDK8fl-MEFs treated with 4OHT exhibited decreased IFN-γ-induced S727 phosphorylation of STAT1 (Figure 4A), whereas STAT1 Y701 phosphorylation was not affected, consistent with previous results (Pelish et al., 2015). CDK19 protein levels remained unchanged in *Cdk8*-deleted cells (Figure 4A), suggesting it was not effectively compensating for CDK8 as a STAT1 AD kinase.

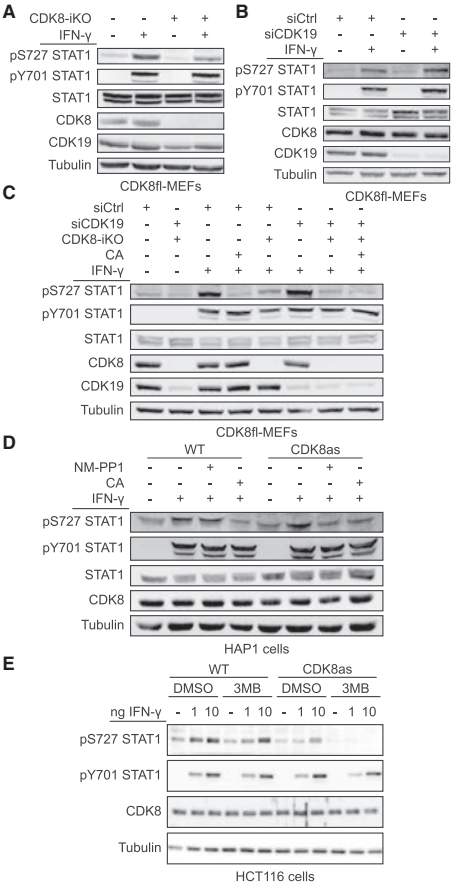


Figure 4. CDK8, Not CDK19, Is the Major STAT1 AD Kinase in the IFN-γ Response

(A) Inducible CDK8 knockout (CDK8-iKO) in CDK8fl-MEFs using 4-hydroxytamoxifen (4OHT) treatment. Cells were 4OHT treated to activate CreERT2 or control treated, followed by IFN-γ stimulation (45 min) and subsequent western analysis using antibodies against phosphorylated STAT1 (pS727 or pY701 STAT1) and total STAT1, CDK8, CDK19, and tubulin.

(B) siRNA knockdown of CDK19 (siCDK19). CDK8fl-MEFs were treated with siCDK19 or non-targeting siCtrl followed by IFN-γ stimulation and immunoblotting as in (A). Quantitative evaluation of blot is shown in Figure S4C.

(C) Effects of inducible CDK8 knockout (CDK8-iKO), CDK19 knockdown (siCDK19), and Mediator kinase inhibition (CA) on IFN-γ-induced STAT1 S727 phosphorylation. Treatments and immunoblotting as in (A) and (B). Note siCDK19 had no effect on IFN-γ-induced STAT1 S727 phosphorylation (lane 6 versus lane 3).

(D and E) Effects of CDK8 inhibition (analog-sensitive mutant CDK8as) on IFN-γ-induced STAT1 S727 phosphorylation.

(D) HAP1 cells expressing WT or CDK8as from the endogenous locus were treated with NM-PP1 (10 μM, 4 h) or CA (100 nM, 1 h) or control treated before stimulation with IFN-γ (45 min). Extracts were analyzed as in (A). Note that inhibition of IFN-γ-induced STAT1 S727 phosphorylation by NM-PP1 was comparable to that by CA.

(E) HCT116 cells expressing WT or CDK8as from the endogenous locus were simultaneously treated with 3MB-PP1 (or DMSO control) and IFN-γ for 45 min. Extracts were analyzed as in (A).

The role of CDK19 in IFN-γ response was assessed via small interfering RNA (siRNA)-mediated *Cdk19* knockdown in CDK8fl-MEFs. This approach was adopted instead of *Cdk19* knockout for several reasons. First, it allowed the same (i.e., isogenic) cells to be used for CDK8 knockout and CDK19 depletion, thereby avoiding differences arising from undefined genetic heterogeneities in different cell lines. Second, this inducible depletion system minimized risks of long-term effects (e.g., adaptation and compensation) in cell populations with stable deletions (Rossi et al., 2015). Third, this strategy was justified by our attempts to generate MEFs bearing a stable *Cdk19* deletion; in some cases, deletion of *Cdk19* in CDK8fl-MEFs caused elevation of CDK8 protein levels (Figure S4B). By contrast, such compensatory effects were not observed if cells were only transiently depleted of CDK19 (Figure 4B).

Knockdown of *Cdk19* by treatment of CDK8fl-MEFs with CDK19-siRNA (referred to as siCDK19) resulted in >90% reduction of CDK19 protein levels (Figures 4B and S4C). Notably, IFN-γ-induced STAT1 S727 phosphorylation was not reduced by *Cdk19* knockdown, suggesting that CDK19 plays no role in phosphorylation of STAT1 AD in the presence of CDK8 (Figure 4B). To more directly address a potential redundancy of CDK8 and CDK19, CDK8fl-MEFs were treated with 4OHT, siCDK19, or both, and IFN-γ-induced STAT1 S727 phosphorylation was examined (Figure 4C). Similar to experiments shown in Figures 4A and 4B, 4OHT (i.e., CDK8-iKO), but not siCDK19 treatment, consistently reduced IFN-γ-induced S727 phosphorylation (Figure 4C; replicate and quantitation in Figure S4D). Treatment with both 4OHT and siCDK19 was only slightly more efficient in inhibition of IFN-γ-induced S727 phosphorylation, similar to treatment with CA (Figures 4C and S4D). Residual S727 phosphorylation is likely caused by kinases not relevant for the IFN-γ pathway (e.g., p38 mitogen-activated protein kinase [MAPK]; Kovarik et al., 1999). Neither the inducible deletion of *Cdk8* nor knockdown of *Cdk19* caused major changes in the expression of other subunits of the Mediator kinase module (i.e., MED12, MED13, and CCNC) (Figure S4E).

The importance of CDK8 compared to CDK19 in IFN-γ-induced STAT1 S727 phosphorylation was further examined using two different human cell lines expressing analog-sensitive CDK8 (CDK8as; F97G) instead of WT CDK8: HCT116-CDK8as (Galbraith et al., 2017) and a HAP1-CDK8as cell line (see STAR Methods). The CDK8as version in HAP1 cells was sensitive to the ATP analog NM-PP1 (Figures S4F and S4G), similar to the reported inhibition of CDK8as in HCT116 cells by 3MB-PP1 (Galbraith et al., 2017). As expected, treatment of HAP1-CDK8as and HCT116-CDK8as cells with different inhibitory ATP analogs (NM-PP1 and 3MB-PP1, respectively) blocked IFN-γ-induced STAT1 S727 phosphorylation, whereas the analogs had no effect in WT cells (Figures 4D, 4E, and S4H). Taken together, the data in Figure 4 implicate CDK8, but not CDK19, as the STAT1 AD kinase in the IFN-γ pathway in both mouse and human cells.

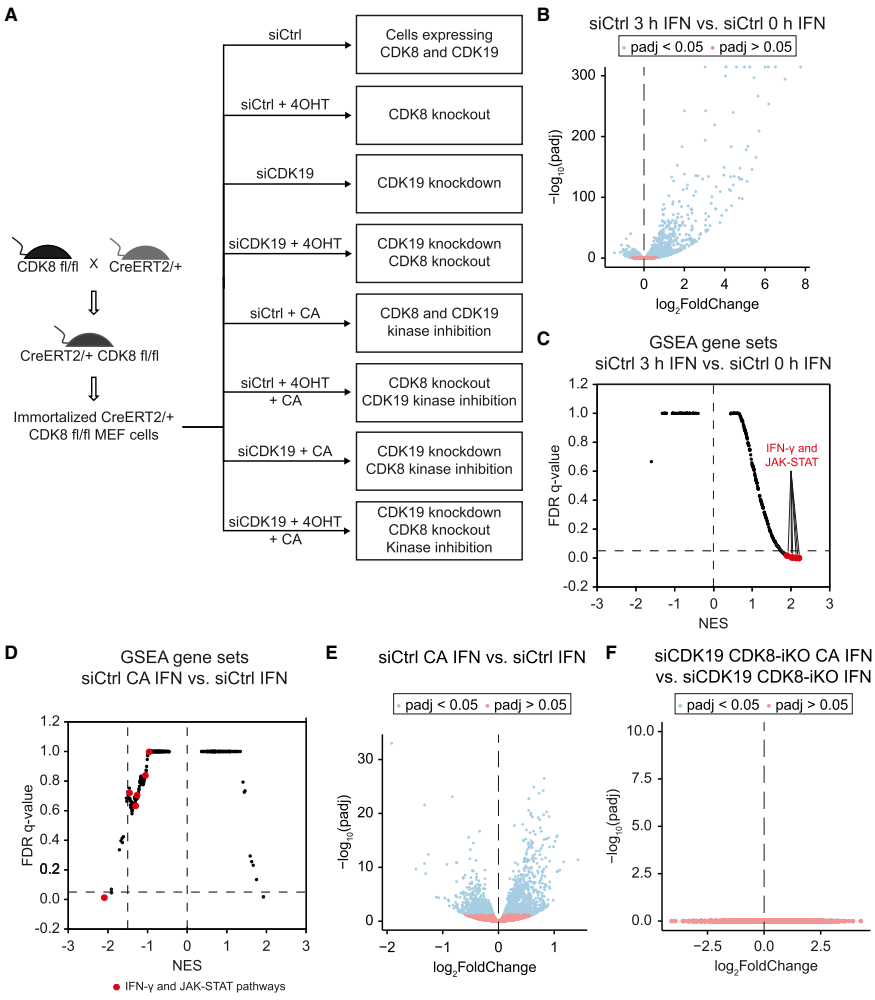


Figure 5. CA Has No Effect on IFN-γ-Regulated Transcription in the Absence of CDK8 and CDK19

(A) Experimental overview. RNA-seq experiments were completed using 3 replicates for each condition (8 conditions total; siCtrl, control condition).

(B) IFN-γ-induced gene expression changes (mRNA) in siCtrl cells (siCtrl 3 h IFN versus siCtrl 0 h IFN).

(C) GSEA of IFN-γ-induced changes in siCtrl cells (siCtrl 3 h IFN versus siCtrl 0 h IFN). Gene sets representing IFN and JAK-STAT pathways are highlighted.

(D) GSEA of CA effects on IFN-γ-induced (3 h) changes in siCtrl cells (siCtrl CA IFN versus siCtrl IFN).

(E) CA effects on gene expression changes (mRNA) upon IFN-γ stimulation (3 h) in siCtrl cells (siCtrl CA IFN versus siCtrl IFN).

(F) CA effects on gene expression changes (mRNA) in the absence of CDK8 and CDK19 during IFN-γ stimulation (3 h) (siCDK19 CDK8-iKO CA IFN versus siCDK19 CDK8-iKO IFN). Note that no genes were significantly (blue) up- or downregulated by CA.

A Strategy to Decouple the Effects of CDK8 versus CDK19: Kinase Function versus Protein Presence

The data in Figure 4 revealed distinct roles for CDK8 versus CDK19 in STAT1 S727 phosphorylation. We next addressed whether we could decouple CDK8- versus CDK19-dependent effects on IFN-γ-induced gene expression. We used the strategy summarized in Figure 5A, which enabled inducible knockout of CDK8 (CDK8-iKO) by 4OHT treatment, knockdown of CDK19 (siCDK19), or both (CDK8-iKO + siCDK19). CA treatment was implemented as needed to decouple protein presence (CDK8

and/or CDK19) from enzymatic function. As before, CA treatment (or DMSO controls) occurred 1 h prior to IFN-γ stimulation. For these experiments, IFN-γ treatment lasted 3 h prior to analysis.

After treatments, mRNA and pre-mRNA levels (three biological replicates) were determined by RNA-seq (Table S7), as described for 6-h IFN-γ treatment (Figure 1). IFN-γ treatment almost exclusively upregulated gene expression, with induction of 274 genes (lfc $\geq$ 1; padj < 0.05; FPKM in IFN-γ stimulated samples $\geq$ 1) and only 4 genes downregulated (lfc $\leq$ 1; padj < 0.05; FPKM in IFN-γ stimulated samples $\geq$ 1) (Figure 5B). The numbers of genes regulated at the pre-mRNA level were similar and largely overlapping with mRNA gene sets implicating IFN-γ effects at the transcriptional level rather than post-transcriptional regulation (Figure S5A). GSEA for IFN-γ-stimulated versus unstimulated datasets confirmed induction of IFN-γ and JAK-STAT pathways (Figure 5C; Table S3C). Moreover, the gene sets revealed by GSEA (normalized enrichment score [NES] > 1.5; FDR < 0.05) in IFN-γ-stimulated versus unstimulated cells showed a large overlap between GRO-seq (t = 30-min IFN-γ) and RNA-seq (t = 3-h IFN-γ) experiments and consisted mostly of IFN-γ or IFN-γ-related pathways (e.g., JAK-STAT, interleukin-2 [IL-2], IL-4, and IL-6) (Figure S5B; Table S3D).

CA Does Not Affect Transcription in the Absence of CDK8 and CDK19

To test whether CA effects on transcription were dependent on the Mediator kinases, we analyzed CA effects during the IFN-γ response in MEFs devoid of both CDK8 and CDK19 (siCDK19 CDK8-iKO) compared to control MEFs (siCtrl). As expected, CA negatively affected gene sets associated with the IFN-γ and JAK-STAT pathways and altered the genome-wide IFN-γ response in control MEFs (Figures 5D and 5E; Table S3E). In contrast, CA did not significantly change the transcriptome in the absence of CDK8 + CDK19, with or without IFN-γ stimulation (Figures 5F and S5C). Similar results were seen in CDK8 knockout cells (see below). These results reflect the high degree of CA selectivity (Pelish et al., 2015) and also imply that CA does not alter the biological function of other (i.e., non-kinase) transcriptional regulatory proteins.

CDK8 and CDK19 are Nonredundant and Mechanistically Distinct Transcriptional Regulators of the IFN-γ Response

To analyze effects of CDK8 and CDK19 during IFN-γ response and to distinguish kinase-dependent versus kinase-independent functions, we generated RNA-seq comparisons among IFN-γ-induced genes from IFN-γ-stimulated control (siCtrl) cell populations. Comparisons were completed between the experiments summarized in Figure 5A. A heatmap depicting genes (N = 178) induced by IFN-γ in control cells (siCtrl) (lfc $\geq$ 1; padj < 0.05; FPKM in IFN-γ stimulated samples $\geq$ 1) that were differentially expressed (padj < 0.05) in at least one of the conditions is shown in Figure 6A. The data revealed a similar, but not identical, pattern of gene expression changes for Mediator kinase inhibition in control cells (siCtrl CA versus siCtrl; Figure 6A, line 1) and in the absence of CDK19 (siCDK19 CA versus siCDK19; Figure 6A, line 2); furthermore, similar gene expression changes were seen in the absence of CDK8 (siCtrl CDK8-iKO versus siCtrl; Figure 6A, line 3). These results (Figure 6A, lines 1–3) suggested that Mediator kinase activity regulated IFN-γ-induced transcription specifically through CDK8, not CDK19. The lack of significant effect of Mediator kinase inhibition in the absence of CDK8 (siCtrl CDK8-iKO CA versus siCtrl CDK8-iKO; Figure 6A, line 4) further supports this conclusion.

The pattern for CDK19 depletion was markedly different from other conditions (siCDK19 versus siCtrl; Figure 6A, line 5), suggesting that CDK19 plays distinct roles in regulation of transcriptional responses to IFN-γ. This was supported by limited overlap of IFN-γ-induced genes downregulated by CDK8 deletion versus CDK19 knockdown (Figure S5D). Volcano plots corresponding to heatmap comparisons are shown in Figures 6B–6F.

The distinct transcriptional effects of CDK19 knockdown (siCDK19) on the IFN-γ transcriptional response were apparent also from GSEA comparisons (Table S3E). Genes downregulated by CDK8-iKO (18 gene sets with NES < −1.5) versus siCDK19 (19 gene sets) shared only 2 gene sets (Figure S5E; Table S3E). Whereas IFN-γ and JAK-STAT pathway gene sets were identified in the CDK8-iKO (Figure S5F) and CA-treated datasets (Figure 5D), these pathways were not observed in siCDK19 gene sets; these gene sets included metabolic and other inflammatory pathways (Figure S5G; Table S3E). Thus, in agreement with the gene expression analyses summarized in Figure 6, GSEA implicated CDK19 in the regulation of different gene sets in IFN-γ-treated cells. We note, however, that expression of several well-known IFN-γ target genes such as *Irf1* and *Gbp2* was similarly affected by CDK8-iKO, siCDK19, or CA (Table S7).

The data summarized in Figures 5 and 6 revealed that the IFN-γ response was differentially regulated by CDK8 and CDK19, by distinct mechanisms. Regulation by CDK8 depended primarily on its kinase activity whereas the effects of CDK19 appeared to be kinase independent.

CDK19 Is a Kinase-Independent Driver of the IFN-γ Antiviral Response

CCNC is required for Mediator kinase activity (Knuesel et al., 2009; Li et al., 2014; Turunen et al., 2014), and we previously established CCNC as important for cellular response to viral infection (Bancerek et al., 2013). Because CCNC interacts with CDK8 or CDK19 in a mutually exclusive manner (Galbraith et al., 2013), these results did not distinguish between CDK8 and CDK19. However, the data shown in Figures 4 and S4 revealed that STAT1 AD phosphorylation was mediated through CDK8. Because STAT1 AD phosphorylation is required for the induction of antiviral state by IFN-γ (Bromberg et al., 1996; Horvath and Darnell, 1996), this implicates CDK8 as an essential activator of antiviral response, which is further supported by CDK8-dependent effects on IFN-γ-induced genes (Figures 6A and S5F).

To test the role of CDK19 in the antiviral response, we used methods similar to those used previously for CCNC (Bancerek et al., 2013); specifically, we infected siCDK19 or siCtrl MEFs (pre-stimulated for 24 h with serial dilutions of IFN-γ) with vesicular stomatitis virus (VSV) and assessed cell viability. Cells depleted for CDK19 reproducibly displayed increased sensitivity

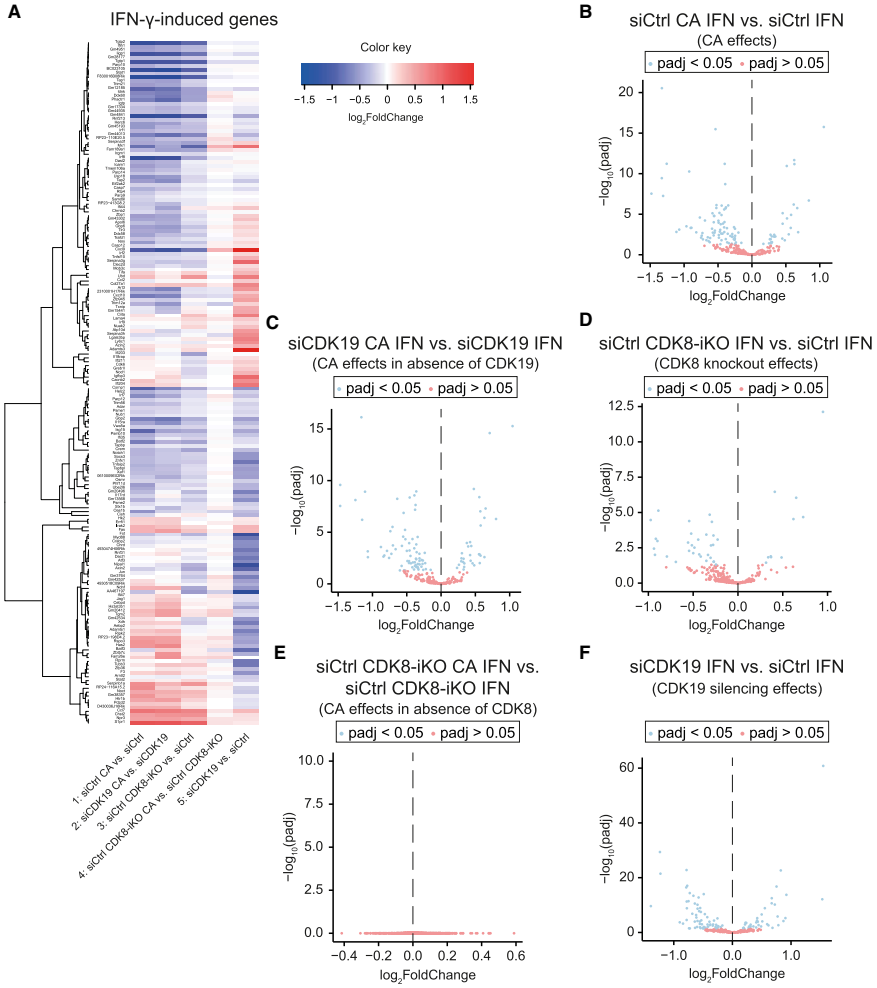


Figure 6. Transcriptional Response to IFN-γ Is Predominantly Executed by Kinase-Dependent Effects of CDK8 and Kinase-Independent Effects of CDK19

(A–F) CDK8fl-MEFs were treated as described in Figure 5A and assessed by differential mRNA expression (3 biological RNA-seq replicates each).

(A) Heatmap summarizing mRNA expression changes caused by CA, CDK8 knockout, and CDK19 knockdown in a group of IFN-γ-induced genes. Genes induced by IFN-γ (lfc ≥ 1, padj < 0.05, FPKM stimulated ≥ 1) in siCtrl-treated CDK8fl-MEFs (siCtrl) were analyzed under the following conditions: CA treatment (CDK8 + CDK19 inhibition), lane 1; CA treatment in the absence of CDK19 (i.e., CDK8 inhibition), lane 2; inducible CDK8 knockout (CDK8-iKO), lane 3; CA treatment in absence of CDK8 (i.e., CDK19 inhibition), lane 4; CDK19 knockdown, lane 5. Only genes that changed (padj < 0.05) in at least 1 of the 5 conditions are shown.

(B–F) Volcano plots corresponding to treatments shown in (A). Panel order in (B)–(F) corresponds to lane order in (A). No significant (padj < 0.05) gene expression changes upon CA treatment of CDK8 knockout cells (CDK19 inhibition: siCtrl CDK8-iKO CA versus siCtrl CDK8-iKO) (E).

to VSV infection compared to siCtrl-treated samples (Figure S6A). Calculation of the EC_{50} (IFN-γ needed to prevent 50% cell death) revealed that cells depleted of CDK19 were 5-fold more sensitive to VSV infection than controls (EC_{50} 0.925 versus 0.181; Figure 7A), demonstrating an important role for the CDK19 protein in the antiviral response.

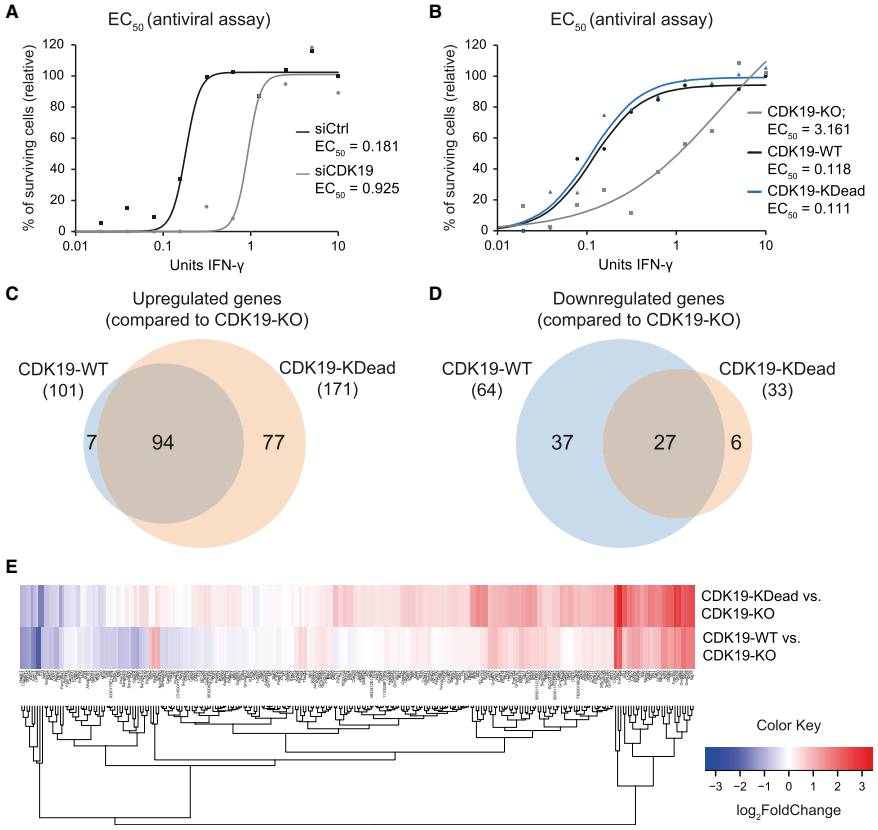


Figure 7. CDK19 Drives the IFN-γ Antiviral Response in Kinase-Independent Ways

(A) The IFN-γ-dependent antiviral response in the absence of CDK19. siCDK19 and siCtrl cells were pretreated with various concentrations of IFN-γ followed by infection with vesicular stomatitis virus (VSV). Percentages of surviving cells and EC_{50} values (IFN-γ concentration needed to prevent 50% cell death) are shown (as means of duplicate experiments).

(B) Assessment of IFN-γ-dependent antiviral response of CDK19-KO cells rescued with CDK19-WT or CDK19-KDead. CDK19-KO, CDK19-WT, and CDK19-KDead cells were IFN-γ treated and VSV infected and analyzed as in (A).

(C and D) Overlap of IFN-γ-induced genes upregulated (C) or downregulated (D) upon rescue of CDK19-KO cells with CDK19-WT or CDK19-KDead. Data are derived from RNA-seq of CDK19-KO, CDK19-WT, and CDK19-KDead stimulated with IFN-γ for 3 h.

(E) Heatmap summarizing expression changes of IFN-γ-induced genes (lfc ≥ 1, padj < 0.05, FPKM stimulated ≥ 1) upon rescue of CDK19-KO cells with CDK19-WT or CDK19-KDead.

To further probe kinase versus protein functions of CDK19, we generated CDK19 knockout (CDK19-KO) MEFs using CRISPR-Cas9 in CDK8fl-MEFs and "rescued" them (without clonal selection) with expression of WT CDK19 (CDK19-WT) or kinase-dead CDK19 (CDK19-KDead). Expression levels of CDK19-WT and CDK19-KDead were comparable in rescue cells (Figure S6C); importantly, the cell pools did not correspondingly upregulate CDK8.

VSV infection assays showed that rescue expression of CDK19-WT or CDK19-KDead had comparable effects, with ~30-fold better IFN-γ-dependent survival compared with CDK19-KO MEFs (EC_{50} 0.118, 0.111, and 3.161, respectively; Figures 7B and S6B). Consistent with the VSV infection assays, RNA-seq data (Table S8) showed similar effects of CDK19-WT and CDK19-KDead rescue expression on IFN-γ-induced genes, as depicted in Venn diagrams (Figures 7C and 7D), heatmaps (Figure 7E), and PCA plots (Figure S6C). These data further support a kinase-independent, structural and/or scaffolding role for the CDK19 protein. Combined with the gene expression data summarized in Figures 5 and 6, these results demonstrated

that CDK8 and CDK19 are essential but nonredundant regulators of IFN-γ responses; moreover, unlike CDK8, CDK19 drives the IFN-γ-induced antiviral defense by a kinase-independent mechanism.

DISCUSSION

The IFN-γ signaling pathway is relevant in most mammalian cell types. The Mediator kinase CDK8 has been shown to regulate IFN-γ-stimulated transcription partially through phosphorylation of the STAT1 AD at S727 (Bancerek et al., 2013; Nitulescu et al., 2017). Potential functions for the highly conserved CDK8 paralog CDK19, however, have not been thoroughly addressed. The development of selective inhibitors of CDK8 and CDK19 has enabled a delineation of kinase-specific functions (Dale et al., 2015; Johannessen et al., 2017; Koehler et al., 2016; Pelish et al., 2015). These studies have established that the transcriptional effects of Mediator kinase inhibition can be distinct from CDK8 or CDK19 knockdown (Poss et al., 2016), which reflects basic structural or scaffolding roles for the kinase protein itself. In this study, we set out to define and decouple (1) the regulatory roles of CDK19 versus CDK8 and (2) the effects of their enzymatic activity versus the structural and/or scaffolding function for each kinase. To our knowledge, this IFN-γ-focused study represents the most thorough analysis of Mediator kinase structural and enzymatic function to date.

The GRO-seq and PRO-seq data have solidified a role for CDK8 in RNAPII pause regulation, which could be inferred from chromatin immunoprecipitation sequencing (ChIP-seq) data in human cells (Galbraith et al., 2013; Pelish et al., 2015). Inhibition of Mediator kinase activity increased pausing at hundreds of genes in MEFs and human HCT116 cells. This effect was prominent at IFN-γ-induced genes, and we noted a correlation between increased PI and genes negatively regulated by CA. These results point to a general role for Mediator kinase activity in RNAPII pausing and/or pause release. Although the kinase substrates for CDK8 and CDK19 have not been identified in IFN-γ-stimulated cells, experiments in unstimulated HCT116 cells identified high-confidence targets that may contribute to Mediator-kinase-dependent RNAPII pause regulation, including AFF4, NELFA, and POLR2M (Poss et al., 2016). CDK8-dependent STAT1 S727 phosphorylation may also impact RNAPII promoter-proximal pausing upon IFN-γ stimulation. Future experiments will seek to delineate phospho-site specific regulatory roles; however, we anticipate that many CDK8 substrates contribute to transcriptional regulation upon activation of IFN-γ signaling cascades.

The most rapid transcriptional response to a stimulus appears to be expression at enhancers (Arner et al., 2015). The MD score is an unbiased means to assess eRNA transcriptional changes, and mapping changes (positive or negative) to consensus TF binding motifs can reliably infer TF activity (Azofeifa et al., 2018). Using the MD score method, we identified eRNAs that were strongly induced by IFN-γ (t = 30 min); as expected, the "epicenters" of the induced eRNAs mapped to consensus motifs of JAK-STAT pathway TFs (e.g., STAT1 and IRF1) in both mouse and human cells. Notably, Mediator kinase inhibition blocked eRNA induction at these loci, suggesting reduced IFN-γ-responsive TF activity. Whereas the MD score data showed expected TF induction upon IFN-γ stimulation, eRNAs associated with other TFs were identified as well. This may reflect uncharacterized biological roles for these factors (e.g., ESR2 and MEF2) in the IFN-γ response, which could be explored in future studies.

In the context of IFN-γ stimulation, we completed comparative analyses in a cell line (CDK8fl-MEFs) that allowed inducible deletion or depletion of either CDK8 or CDK19, thereby avoiding compensatory effects that commonly arise from clonal selection of knockout cell lines (Rossi et al., 2015). Although both CDK8 and CDK19 were shown to govern the IFN-γ transcriptional response, they regulated distinct sets of genes via distinct mechanisms. The impact of CDK8 derived primarily from its kinase activity. Using a combination of approaches, we observed that phosphorylation of the STAT1 AD at S727 is mediated by CDK8, but not CDK19, extending previous results in IFN-γ-stimulated MEFs (Bancerek et al., 2013). In contrast to CDK8, we observed that the kinase activity of CDK19 was largely inconsequential. Rather, a structural role was evident for CDK19, as its inducible knockdown triggered stark transcriptional effects during IFN-γ stimulation that were not affected by CA treatment.

Both the kinase-dependent (CDK8-mediated) and kinase-independent (CDK19-mediated) effects appeared to be essential for the IFN-γ response. STAT1 AD phosphorylation (CDK8 dependent) is required for efficient induction of the antiviral state (Bromberg et al., 1996; Horvath and Darnell, 1996). In this study, we demonstrated that the CDK8 paralog CDK19 is a similarly essential component of the IFN-γ-induced antiviral program. Importantly, CDK19 kinase-independent function was verified in rescue experiments; expression of either WT or CDK19-KDead in a null background restored gene expression and VSV resistance toward WT levels. A structural and/or scaffolding role for CDK19 (i.e., kinase independent) was also observed during p53 response in human osteosarcoma cells (Audetat et al., 2017), suggesting that CDK19 mediates structural interactions not shared by CDK8. Because CDK8 and CDK19 are mutually exclusive subunits of the Mediator kinase module (Galbraith et al., 2013), these results suggest the assembly of distinct CDK8- and CDK19-Mediator complexes at select genomic loci during IFN-γ stimulation.

Because IFN-γ almost exclusively upregulated gene expression, this served as a means to study Mediator-kinase-dependent effects on transcription activation. We observed that CA had minimal impact on gene expression in unstimulated cells, whereas CA suppressed induction of dozens of genes activated by IFN-γ. These findings have parallels with work from the Roninson group, in which they noted that Mediator kinase inhibition (with Senexin A) suppressed transcriptional activation by nuclear factor κB (NF-κB) but had little effect on basal expression (Chen et al., 2017). Similarly, Johannessen et al. noted that Mediator kinase inhibition did not grossly perturb the transcriptome of quiescent macrophages (Johannessen et al., 2017). Given the selective effects on inducible (i.e., not basal) gene expression, the Roninson group concluded that CDK8 and/or CDK19 may represent key mediators of transcriptional reprogramming (Chen et al., 2017). In this context, transcriptional reprogramming refers to the initial stimulus response and is distinct from reprogramming associated with cell differentiation.

Rapid transcriptional changes are important for immune system activation during infection and likely contribute to longer-term shifts in gene expression patterns (e.g., through epigenetic changes; Ivashkiv, 2018). Our results implicate Mediator kinases in these processes, but future study is needed to characterize their impact across longer time frames. It is notable that CDK8 or CDK19 knockdown or CDK8 and CDK19 inhibition is not generally cytotoxic under normal growth conditions (Donner et al., 2007; Galbraith et al., 2013; Pelish et al., 2015), implicating Mediator kinase function as more important for adaptive transcriptional responses (i.e., reprogramming). The regulation of IFN responses has broad physiological relevance, ranging from inflammation to aging to tumor cell clearance. Our results reveal that the Mediator kinases activate distinct transcriptional programs, through distinct mechanisms, in response to the ubiquitous inflammatory cytokine IFN-γ, suggesting that separately targeting CDK8 kinase activity or CDK19 protein levels (e.g., with proteolysis targeting chimeras [PROTACs]) may have diverse biomedical applications.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - MEF (mouse embryonic fibroblasts) cell lines established in this study:
- METHOD DETAILS
  - Cell culture
  - Cytokines and inhibitors
  - RNA-Seq
  - GRO-Seq
  - PRO-Seq
  - Modified motif displacement score (MD Score) analysis
  - Inducible CDK8 knockout
  - Knockdown of CDK19
  - Knockout of CDK19
  - HAP1 cells expressing analog-sensitive CDK8 (CDK8as)
  - CDK19-WT expression construct
  - CDK19-KDead expression construct
  - Rescue of CDK19-KO cells with CDK19-WT or CDK19-KDead
  - Generation of CDK8as mutant for *in vitro* kinase assays
  - *In vitro* kinase assays
  - Whole cell extracts and western blotting
  - Cytotoxicity assay
  - Vesicular stomatitis virus (VSV) infection assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2019.07.034.

ACKNOWLEDGMENTS

We thank T. Read, K. Audetat, and M. Allen for helpful discussions and M. Shair for cortistatin A. C. Larson and A. Vogt are acknowledged for sequencing advice. We thank T.M. Decker and T. Lee for antibody testing with immunoprecipitation-mass spectrometry (IP-MS) and P. Fischer for advice on RNA-seq library scaling. We thank K. Chylinski and VBCF Protein Technologies Facility (https://www.viennabiocenter.org/facilities) for help with CRISPR-Cas9 genome editing. This work was supported by the Austrian Science Fund (FWF) grants P31848-B, P27538-B21, SFB-F43, and W1261 (to P.K.); NIH grants GM117370 (to D.J.T.), GM008759 (to J.D.R.), AG051335 (to C.B.L.), and GM120109 and CA117907 (to J.M.E.); NSF grant MCB-1817582 (to J.M.E.); the Fonfara-LaRose Family Fund (J.M.E.); and FWF grant I2192-B22 ERASE and European Research Council grant AdG 695214 GameofGates (to G.S.-F.). We acknowledge the BioFrontiers Computing Core at UC-Boulder, supported by NIH grant OD012300.

AUTHOR CONTRIBUTIONS

I.S., D.J.T., and P.K. conceptualized and designed the experiments. I.S. established experimental systems and conducted GRO-seq, RNA-seq, and VSV assays. V.S. analyzed RNA-seq and GRO-seq data. J.D.R. performed GSEA and MD score analysis; M.D.G. completed PRO-seq data analysis. L.S. and F.A. performed differential expression analysis for RNA-seq data. K.E. generated constructs for CDK19-WT and CDK19-KDead rescue, established rescue cell lines, and together with R.K. generated analog-sensitive CDK8 cells. R.K. produced the graphical abstract. T.V. generated CDK8fl-MEFs and CDK19-KO MEFs. C.B.L. performed *in vitro* kinase assays and together with H.B. and Z.A. performed PRO-seq. Funding and mentoring were provided by J.M.E., R.D.D., G.S-F., D.J.T., and P.K. All authors contributed to data interpretation. I.S., D.J.T., and P.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 16, 2019
Revised: July 3, 2019
Accepted: July 25, 2019
Published: September 5, 2019

REFERENCES

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Allen, M.A., Andrysik, Z., Dengler, V.L., Mellert, H.S., Guarnieri, A., Freeman, J.A., Sullivan, K.D., Galbraith, M.D., Luo, X., Kraus, W.L., et al. (2014). Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. eLife *3*, e02200.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., et al.; FANTOM Consortium (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science *347*, 1010–1014.

Audetat, K.A., Galbraith, M.D., Odell, A.T., Lee, T., Pandey, A., Espinosa, J.M., Dowell, R.D., and Taatjes, D.J. (2017). A kinase-independent role for cyclin-dependent kinase 19 in p53 response. Mol. Cell. Biol. *37*, e00626-16.

Azofeifa, J.G., Allen, M.A., Hendrix, J.R., Read, T., Rubin, J.D., and Dowell, R.D. (2018). Enhancer RNA profiling predicts transcription factor activity. Genome Res. Published online February 15, 2018. https://doi.org/10.1101/gr.225755.117.

Bancerek, J., Poss, Z.C., Steinparzer, I., Sedlyarov, V., Pfaffenwimmer, T., Mikulic, I., Dölken, L., Strobl, B., Müller, M., Taatjes, D.J., and Kovarik, P. (2013). CDK8 kinase phosphorylates transcription factor STAT1 to selectively regulate the interferon response. Immunity *38*, 250–262.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168.

Bromberg, J.F., Horvath, C.M., Wen, Z., Schreiber, R.D., and Darnell, J.E., Jr. (1996). Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. Proc. Natl. Acad. Sci. USA *93*, 7673–7678.

Cadiñanos, J., and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. Nucleic Acids Res. *35*, e87.

Carette, J.E., Raaben, M., Wong, A.C., Herbert, A.S., Obernosterer, G., Mulherkar, N., Kuehne, A.I., Kranzusch, P.J., Griffin, A.M., Ruthel, G., et al. (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature *477*, 340–343.

Chen, M., Liang, J., Ji, H., Yang, Z., Altilia, S., Hu, B., Schronce, A., McDermott, M.S.J., Schools, G.P., Lim, C.U., et al. (2017). CDK8/19 Mediator kinases potentiate induction of transcription by NFκB. Proc. Natl. Acad. Sci. USA *114*, 10208–10213.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Dale, T., Clarke, P.A., Esdar, C., Waalboer, D., Adeniji-Popoola, O., Ortiz-Ruiz, M.J., Mallinger, A., Samant, R.S., Czodrowski, P., Musil, D., et al. (2015). A selective chemical probe for exploring the role of CDK8 and CDK19 in human disease. Nat. Chem. Biol. *11*, 973–980.

Decker, T., Kovarik, P., and Meinke, A. (1997). GAS elements: a few nucleotides with a major impact on cytokine-induced gene expression. J. Interferon Cytokine Res. *17*, 121–134.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dolken, L., Ruzsics, Z., Radle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., and Koszinowski, U.H. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972.

Donner, A.J., Szostek, S., Hoover, J.M., and Espinosa, J.M. (2007). CDK8 is a stimulus-specific positive coregulator of p53 target genes. Mol. Cell *27*, 121–133.

Galbraith, M.D., Allen, M.A., Bensard, C.L., Wang, X., Schwinn, M.K., Qin, B., Long, H.W., Daniels, D.L., Hahn, W.C., Dowell, R.D., and Espinosa, J.M. (2013). HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell *153*, 1327–1339.

Galbraith, M.D., Andrysik, Z., Pandey, A., Hoh, M., Bonner, E.A., Hill, A.A., Sullivan, K.D., and Espinosa, J.M. (2017). CDK8 kinase activity promotes glycolysis. Cell Rep. *21*, 1495–1506.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. *17*, 148.

Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and Bioconductor. Methods Mol. Biol. *1418*, 335–351.

Hameyer, D., Loonstra, A., Eshkind, L., Schmitt, S., Antunes, C., Groen, A., Bindels, E., Jonkers, J., Krimpenfort, P., Meuwissen, R., et al. (2007). Toxicity of ligand-dependent Cre recombinases and generation of a conditional Cre deleter mouse allowing mosaic recombination in peripheral tissues. Physiol. Genomics *31*, 32–41.

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

Horvath, C.M., and Darnell, J.E., Jr. (1996). The antiviral state induced by alpha interferon and gamma interferon requires transcriptionally active Stat1 protein. J. Virol. *70*, 647–650.

Ivashkiv, L.B. (2018). IFNgamma: signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. Nat. Rev. Immunol. *18*, 545–558.

Johannessen, L., Sundberg, T.B., O'Connell, D.J., Kolde, R., Berstler, J., Billings, K.J., Khor, B., Seashore-Ludlow, B., Fassl, A., Russell, C.N., et al. (2017). Small-molecule studies identify CDK8 as a regulator of IL-10 in myeloid cells. Nat. Chem. Biol. *13*, 1102–1108.

Kechin, A., Boyarskikh, U., Kel, A., and Filipenko, M. (2017). cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. J. Comput. Biol. *24*, 1138–1143.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Knuesel, M.T., Meyer, K.D., Donner, A.J., Espinosa, J.M., and Taatjes, D.J. (2009). The human CDK8 subcomplex is a histone kinase that requires Med12 for activity and can function independently of mediator. Mol. Cell. Biol. *29*, 650–661.

Koehler, M.F., Bergeron, P., Blackwood, E.M., Bowman, K., Clark, K.R., Firestein, R., Kiefer, J.R., Maskos, K., McCleland, M.L., Orren, L., et al. (2016). Development of a potent, specific CDK8 kinase inhibitor which phenocopies CDK8/19 knockout cells. ACS Med. Chem. Lett. *7*, 223–228.

Kovarik, P., Stoiber, D., Eyers, P.A., Menghini, R., Neininger, A., Gaestel, M., Cohen, P., and Decker, T. (1999). Stress-induced phosphorylation of STAT1 at Ser727 requires p38 mitogen-activated protein kinase whereas IFN-gamma uses a different signaling pathway. Proc. Natl. Acad. Sci. USA *96*, 13956–13961.

Kulakovskiy, I.V., Medvedeva, Y.A., Schaefer, U., Kasianov, A.S., Vorontsov, I.E., Bajic, V.B., and Makeev, V.J. (2013). HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Res. *41*, D195–D202.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Li, N., Fassl, A., Chick, J., Inuzuka, H., Li, X., Mansour, M.R., Liu, L., Wang, H., King, B., Shaik, S., et al. (2014). Cyclin C is a haploinsufficient tumour suppressor. Nat. Cell Biol. *16*, 1080–1091.

Li, D., Zhang, B., Xing, X., and Wang, T. (2015). Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. Methods *72*, 29–40.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

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

Madsen, J.G., Schmidt, S.F., Larsen, B.D., Loft, A., Nielsen, R., and Mandrup, S. (2015). iRNA-seq: computational method for genome-wide assessment of acute transcriptional regulation from total RNA-seq data. Nucleic Acids Res. *43*, e40.

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat. Protoc. *11*, 1455–1476.

Mancino, A., and Natoli, G. (2016). Specificity and function of IRF family transcription factors: insights from genomics. J. Interferon Cytokine Res. *36*, 462–469.

Nitulescu, I.I., Meyer, S.C., Wen, Q.J., Crispino, J.D., Lemieux, M.E., Levine, R.L., Pelish, H.E., and Shair, M.D. (2017). Mediator kinase phosphorylation of STAT1 S727 promotes growth of neoplasms with JAK-STAT activation. EBioMedicine *26*, 112–125.

Pelish, H.E., Liau, B.B., Nitulescu, I.I., Tangpeerachaikul, A., Poss, Z.C., Da Silva, D.H., Caruso, B.T., Arefolov, A., Fadeyi, O., Christie, A.L., et al. (2015).

Mediator kinase inhibition further activates super-enhancer-associated genes in AML. Nature *526*, 273–276.

Poss, Z.C., Ebmeier, C.C., Odell, A.T., Tangpeerachaikul, A., Lee, T., Pelish, H.E., Shair, M.D., Dowell, R.D., Old, W.M., and Taatjes, D.J. (2016). Identification of Mediator kinase substrates in human cells using cortistatin a and quantitative phosphoproteomics. Cell Rep. *15*, 436–450.

R Core Team. (2017). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nat. Genet. *38*, 500–501.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature *524*, 230–233.

Sadzak, I., Schiff, M., Gattermeier, I., Glinitzer, R., Sauer, I., Saalmüller, A., Yang, E., Schaljo, B., and Kovarik, P. (2008). Recruitment of Stat1 to chromatin is required for interferon-induced serine phosphorylation of Stat1 transactivation domain. Proc. Natl. Acad. Sci. USA *105*, 8944–8949.

Schneider, W.M., Chevillotte, M.D., and Rice, C.M. (2014). Interferon-stimulated genes: a complex web of host defenses. Annu. Rev. Immunol. *32*, 513–545.

Todaro, G.J., and Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. J. Cell Biol. *17*, 299–313.

Turunen, M., Spaeth, J.M., Keskitalo, S., Park, M.J., Kivioja, T., Clark, A.D., Mäkinen, N., Gao, F., Palin, K., Nurkkala, H., et al. (2014). Uterine leiomyoma-linked MED12 mutations disrupt mediator-associated CDK activity. Cell Rep. *7*, 654–660.

Wen, Z., Zhong, Z., and Darnell, J.E., Jr. (1995). Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. Cell *82*, 241–250.

Wiesauer, I., Gaumannmüller, C., Steinparzer, I., Strobl, B., and Kovarik, P. (2015). Promoter occupancy of STAT1 in interferon responses is regulated by processive transcription. Mol. Cell. Biol. *35*, 716–727.

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
pS727 STAT1	Kovarik et al., 1999	N/A
pS727 STAT1	Cell Signaling Technology	Cat#9177; RRID: AB_2197983
pY701 STAT1 (58D6)	Cell Signaling Technology	Cat#9167S; RRID: AB_561284
pY701 STAT1 (D4A7)	Cell Signaling Technology	Cat#7649; RRID: AB_10950970
STAT1 p84/p91 (E23)	Santa Cruz Biotechnology	Cat#sc-346; RRID: AB_632435
CDK8	Santa Cruz Biotechnology	Cat#sc-1521; RRID: AB_2260300
CDK8	Cell Signaling Technology	Cat#4101S; RRID: AB_1903934
CDK19	Sigma-Aldrich	Cat# HPA007053; RRID: AB_1846369
Alpha-Tubulin	Cell Signaling Technology	Cat#2144S; RRID: AB_2210548
Alpha-Tubulin	Sigma-Aldrich	Cat#T9026; RRID: AB_477593
MED12	Bethyl	Cat#A300-774A; RRID: AB_669756
MED13, TRAP240(E-12)	Santa Cruz Biotechnology	Cat# sc-515557
CCNC	Bethyl	Cat#A301-989A; RRID: AB_1576505
IRF1 (D5E4)	Cell Signaling Technology	Cat#8478T; RRID: AB_10949108
Peroxidase-conjugated AffiniPure Goat Anti-Rabbit IgG (H+L)	Jackson Immunoresearch	Cat#111-035-003; RRID: AB_2313567
Peroxidase-conjugated AffiniPure Goat Anti-Mouse IgG (H+L)	Jackson Immunoresearch	Cat#115-035-003; RRID: AB_10015289
Bacterial and Virus Strains		
Vesicular stomatitis virus (VSV), Indiana strain	Laboratory of Birgit Strobl	N/A
Chemicals, Peptides, and Recombinant Proteins		
Murine Interferon-gamma	eBioscience	Cat#14-8311
Human Interferon-gamma	Laboratory of James E. Darnell	N/A
Human Interferon-gamma	Fisher Scientific	Cat#PHC4031
Cortistatin A	Laboratory of Matthew Shair	N/A
NM-PP1	Calbiochem	Cat#529581
3MB-PP1	Cayman Chemical	Cat#56025-83-5
4-Hydroxytamoxifen	Sigma-Aldrich	Cat#T176-10MG
G418 (Geneticin)	Invivogen	Cat#ant-gn-1
Lipofectamine RNAiMAX Transfection Reagent	Invitrogen	Cat#13778-150
cOmplete Protease Inhibitor Cocktail	Roche	Cat#11836145001
SUPERase-In	Invitrogen	Cat#AM2694
RQ1 RNase-Free DNase	Promega	Cat#M6101
DNase I	Merck	Cat #4716728001
RevertAid Reverse Transcriptase	Thermo Scientific	Cat#EP0442
Q5® High-Fidelity DNA Polymerase	NEB	Cat#M0491L
XbaI	Thermo Scientific	Cat#FD0684
NotI	Thermo Scientific	Cat#FD0594
XhoI	Thermo Scientific	Cat#FD0694
T4 DNA Ligase	Thermo Scientific	Cat#EL001
Ampicillin	AppliChem	Cat#A0839
Fast Alkaline Phosohatase	Thermo Scientific	Cat#EF0651

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
NEBNext Ultra Directional RNA Library Prep Kit	NEB	Cat# E7420S
NEBNext Ultra II RNA Library Prep Kit	NEB	Cat# E7770S
NEBNext rRNA Depletion Kit	NEB	Cat# E6310S
NEBNext Poly(A) mRNA Magnetic Isolation Module	NEB	Cat# E7490S
RNase-Free DNase Set	QIAGEN	Cat#79254
RNeasy Mini Kit	QIAGEN	Cat#74104
QIAzol Lysis Reagent	QIAGEN	Cat#79306
TRIzol Lysis Reagent	Invitrogen	Cat#15596026
RNA 6000 Nano Assays	Agilent	Cat#5067-1511
Ambion Fragmentation Reagents	Ambion	Cat#AM8740
Illustra MicroSpin G-25 columns	GE Healthcare	Cat#27532501
Anti-BrU agarose beads	Santa Cruz Biotechnology	Cat#sc-32323AC
Monarch DNA gel extraction kit	NEB	Cat#T1020L
Plasmid mini prep kit	QIAGEN	Cat#2710
Gibson Assembly Master Mix	NEB	Cat#E2611L
Deposited Data		
Raw and analyzed data - RNA-Seq experiments, GRO-Seq experiment	This paper	SRA: PRJNA542065 https://www.ncbi.nlm.nih.gov/sra
Raw and analyzed data - PRO-Seq experiment	This paper	GEO: GSE129501 https://www.ncbi.nlm.nih.gov/geo/
Raw image files - western blots	This paper	https://doi.org/10.17632/crj8f3j63z.1
Mouse reference genome (GRCm38/mm10)	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/mouse
Human reference genome (GRCh37/hg19)	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/human
HOCOMOCO database	Kulakovskiy et al., 2013	http://hocomoco11.autosome.ru/
Experimental Models: Cell Lines		
WT MEFs	Bancerek et al., 2013	N/A
CDK8fl-MEFs	This study	N/A
CDK19 KO MEFs	This study	N/A
CDK19 KO MEFs expressing CDK19-WT	This study	N/A
CDK19 KO MEFs expressing CDK19-KDead	This study	N/A
HAP1 WT	Haplogen	N/A
HAP1 CDK8 KO (clone 325-1)	Haplogen	N/A
HAP1 CDK8as	This study	N/A
HCT116 WT	Galbraith et al., 2017	N/A
HCT116 CDK8as	Galbraith et al., 2017	N/A
Experimental Models: Organisms/Strains		
Cdk8tm1c$^{(EUCOMM)Hmgu}$	Yann Herault, IGMBC	N/A
Rosa26CreERT2	Hameyer et al., 2007	N/A
CreERT2-CDK8fl/fl	This study	N/A
Oligonucleotides		
ON-TARGET plus SMART pool siRNA CDK19	Dharmacon	Cat#L-059630-00-0010
ON-TARGET plus non-targeting siRNA pool	Dharmacon	Cat#D-001810-10-20
PCR primers and guide RNAs		
PCR primers and guide RNAs	This paper	see Table S9
Recombinant DNA		
pSpCas9(BB)-2A-Puro (PX459)	Addgene plasmid	Cat#48139
pOG44 Flp-Recombinase expression vector	Thermo Scientific	Cat#V600520

(*Continued on next page*)

CellPress

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PB-EF1α-MCS-IRES-Neo cDNA cloning and expression vector	SBI System Biosciences	Cat#PB533A-2
pB_EF1_CDK19_IRES_Neo	This study	N/A
piggyBac transposase plasmid	Cadiñanos and Bradley, 2007	N/A
pKozak plasmid	Knuesel et al., 2009	N/A
Software and Algorithms		
Image Lab version 5.2.1	Bio-Rad	N/A
Gen5 Microplate Reader Software	BioTek	N/A
AAT-Bioquest EC_{50}	AAT-Bioquest	https://www.aatbio.com/tools/ec50-calculator
GraphPad Prism 6	Graph Pad Software	https://www.graphpad.com
HTSeq	Anders et al., 2015	https://github.com/simon-anders/htseq
cutadapt		https://doi.org/10.14806/ej.17.1.200
FastQC version 0.11.5		http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FastQ Screen version 0.11.0		https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/
STAR version 2.5	Dobin et al., 2013	https://code.google.com/archive/p/rna-star/
R-project version 3.4.2 with RStudio IDE version 1.0.143	R Core Team, 2017	https://www.R-project.org/
GenePattern server	Reich et al., 2006	http://software.broadinstitute.org/cancer/software/genepattern/
DESeq2 version 1.18.1 and 1.22.1	Love et al., 2014	http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html
RSeQC		http://rseqc.sourceforge.net/
Mix2 RNA-Seq data analysis software	Lexogen	N/A
Broad Institute sgRNA designer		https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design
BBDUK from BBTools (v37.99)		https://sourceforge.net/projects/bbmap/
FASTQ-MCF from ea-utils (v1.05)	Kechin et al., 2017	https://expressionanalysis.github.io/ea-utils/
Hisat2 (v2.1.0)	Kim et al., 2015	https://ccb.jhu.edu/software/hisat2/index.shtml
Samtools (v1.5)	Li et al., 2015	https://sourceforge.net/projects/samtools/files/samtools/1.5/
featureCount (v1.6.2)	Liao et al., 2014	http://subread.sourceforge.net/
HOMER (v4.9.1)	Heinz et al., 2010	http://homer.ucsd.edu/homer/
Gviz package (v1.26.4)	Hahne and Ivanek, 2016	https://bioconductor.org/packages/release/bioc/html/Gviz.html
CRISPOR tool	Haeussler et al., 2016	http://crispor.tefor.net/
TIDE	Brinkman et al., 2014	https://tide.deskgen.com/
Broad Institute sgRNA designer		https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pavel Kovarik (pavel.kovarik@univie.ac.at).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

MEF (mouse embryonic fibroblasts) cell lines established in this study:

Conditional CDK8 knockout MEFs: MEFs allowing inducible deletion of CDK8 were derived from CreERT2-Cdk8fl/fl mice. Cdk8fl/fl mice (allele Cdk8tm1c$^{(EUCOMM)Hmgu}$, provided by Yann Herault, IGMBC, Illkirch, France) on C57BL/6 background were crossed with

Rosa26CreERT2 mice (Hameyer et al., 2007) to generate CreERT-Cdk8fl/fl mice. Primary MEFs from CreERT2-CDK8fl/fl mice were isolated on day 13.5 and immortalized via the 3T3 method (Todaro and Green, 1963). Activation of the CreERT2 recombinase led to the excision of the loxP-flanked exon 5 resulting in a frameshift and nonsense mediated decay.

CDK19 KO MEFs

CDK19 knockout cells were generated in CDK8fl-MEFs using CRISPR/Cas9 (described in detail in the Method Details section).

CDK19 KO MEFs expressing CDK19-WT or CDK19-KDead protein

CDK19 KO cells were reconstituted with either CDK19-WT or CDK19-KDead expression constructs as described in the Method details section.

METHOD DETAILS

Cell culture

Mouse embryonic fibroblasts expressing wild-type STAT1 (WT MEFs) were described previously (Bancerek et al., 2013). MEFs allowing inducible deletion of CDK8 were generated by immortalization of primary MEFs derived from CreERT2-Cdk8fl/fl mice. Briefly, Cdk8fl/fl mice (allele Cdk8tm1c$^{(EUCOMM)Hmgu}$, provided by Yann Herault, IGMBC, Illkirch, France) on C57BL/6 background were crossed with Rosa26CreERT2 mice (Hameyer et al., 2007) to generate CreERT-Cdk8fl/fl mice. Primary MEFs from CreERT2-CDK8fl/fl mice were isolated on day 13.5 and immortalized via the 3T3 method (Todaro and Green, 1963). Activation of the CreERT2 recombinase led to the excision of the loxP-flanked exon 5 resulting in a frameshift and nonsense mediated decay. MEFs were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS) and penicillin-streptomycin. Wild-type HAP1 cells (Carette et al., 2011) and CRISPR-Cas9 generated CDK8 knockout HAP1 cells (clone 325-1) were purchased from Haplogen (Austria). HAP1 cells were grown in Iscove's modified Dulbecco's medium (IMDM) supplemented with 10% FCS, 200 nM L-glutamine and penicillin-streptomycin. HCT116 cells were grown in McCoy's media (GIBCO, 16600082) with GIBCO 100x Antibiotic-Antimycotic (Fisher Scientific, 15240062) penicillin-streptomycin and 10% fetal bovine serum (FBS) supplementation.

Cytokines and inhibitors

Murine IFN-γ (eBioscience) was used at a concentration of 10 ng/ml for stimulation of MEFs. Human IFN-γ (kind gift from James E. Darnell, Rockefeller University, US) was used for stimulation of HAP1 cells at 10 ng/ml. HCT116 cells were treated with 10 ng/ml IFN-γ (Fisher Scientific, #PHC4031). Cortistatin A (kindly provided by Matthew Shair, Harvard University, Cambridge, USA) was applied at a concentration of 100 nM one hour before IFN-γ stimulation. NM-PP1 (Calbiochem, 529581) used for inhibition of analog-sensitive CDK8 mutant (CDK8as) was applied at a concentration of 10 μM 4 h before IFN-γ stimulation. The ATP analog 3MB-PP1 (Cayman Chemical, 56025-83-5) was applied at 10 μM for 45 minutes (simultaneously with IFN-γ).

RNA-Seq

The RNA isolation was done as described (Audetat et al., 2017). In brief cells were seeded with 70% confluency on 15 cm and 6 cm dishes, respectively. For total RNA isolation, 7 mL (15 cm dish) or 2 mL (6 cm dish) Qiazol Lysis Reagent (QIAGEN, 79306) were added. The samples were mixed thoroughly before taking 1 aliquot for Chloroform extraction. RNA was precipitated with Isopropanol and Sodium chloride, followed by DNase treatment using RNase-free DNase Set (QIAGEN, 79254) and clean up using RNeasy Mini Kit (QIAGEN, 74104). For library preparation the NEBNext rRNA Depletion Kit (NEB E6310S) or the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490S), together with the NEBNext Ultra II RNA Library Prep Kit from NEB (NEB E7770S) were used according to the manufacturer's protocol. The RNA quality was assessed using Agilent RNA 6000 Nano Assays (5067-1511) that were analyzed on an Agilent 2100 Bioanalyzer. The library quality check and Solexa sequencing was performed at the VBCF NGS Unit (https://www.viennabiocenter.org/facilities). Single-end fragment libraries (50 bp) were sequenced on the Illumina HiSeq 2500 platform. Processing of raw reads and mapping were done as described for GRO-seq experiment. Quantitation of RNA-Seq data for WT MEFs stimulated with IFN-γ for 6 h ± CA treatment (Figure 1) was carried out using HTSeq (Anders et al., 2015). Reads mapped to exons and exon-exon junctions were defined as mRNA reads, while reads mapped to introns and intron-exon junctions were defined as pre-mRNA reads. The raw as well as processed data are accessible via the NCBI's Sequence Read Archive (SRA) database (accession number PRJNA542065). Differential expression analysis was performed based on read counts using DESeq2 (Love et al., 2014). Principal component analysis (PCA) and normalization of read counts to library size and composition (using DESeq2) revealed that replicate 1 of sample 6 h IFN-γ without CA was an outlier. This replicate was removed from the subsequent analysis. IFN-γ-induced genes were defined by $\log_2$-fold-change (lfc) ≥ 1, padj < 0.05 and FPKM stimulated ≥ 1. Analysis of RNA-Seq data for siCtrl, CDK8-iKO and siCDK19 in the presence or absence of CA with or without 3 h IFN-γ stimulation (Figure 5 and 6) revealed more than 95% uniquely mapped reads in each sample. Sample integrity was analyzed by gene body coverage plots using RSeQC. Transcripts were quantified using Mix^2 RNA-Seq data analysis software (Lexogen). Differential expression analysis was performed using DESeq2 version 1.18.1. Exploratory data analysis and visualizations were performed in R-project version 3.4.2 (R Core Team, 2017) (Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/) with Rstudio IDE version 1.0.143, ggplot2 (2.2.1), dplyr (0.7.4), readr (1.1.1), gplots (3.0.1). GSEA was completed as described for GRO-Seq experiments. Genes were required to have FPKM > 0.5 in all replicates and conditions.

CellPress

GRO-Seq

Cells were seeded on three 15 cm dishes (4x10^6 per dish) for each time point or treatment 16 h prior to the experiments (~70% confluency at the time of experiment). Cells were stimulated or treated as desired, washed twice with cold PBS and detached using 2 mL Trypsin/EDTA per dish. Cells were re-suspended in 10 mL cold PBS and collected by centrifugation at 270 g for 5 minutes. For lysis, cells were incubated for 10 minutes in 10 mL lysis buffer (10 mM Tris-HCl, 5 mM $MgCl_2$, 10 mM NaCl, 0.5% NP-40, 1 mM DTT, 1 mM sodium metabisulfite, 1 mM benzamidine, 0.025 mM PMSF, 4 U/ml SUPERase-In) on ice. Nuclei were collected by centrifugation at 170 g for 10 minutes, washed once with 1 mL reaction buffer (20 mM Tris-HCl, 10 mM $MgCl_2$, 150 mM KCl, 20% Glycerol, 4 U/ml SUPERase-In), and re-suspended in 50 μl reaction buffer and counted. Total 5x10^6 nuclei in 100 μl reaction buffer were used per run-on reaction. The Run-on reaction and capturing of the labeled RNA was done as described previously (Allen et al., 2014). Briefly, 28.9 μl reaction buffer, 5 μl of rATP, rCTP, rGTP and 5-Bromo-UTP (10 mM each), 0.1 μl DTT, 1 μl RNase-In and 50 μl 2% sarkosyl were added per run-on reaction. The samples were incubated at 30°C for 5 minutes, followed by RNA isolation using TRIzol reagent (Invitrogen, 15596026). RNA precipitation was done using Isopropanol and GlycoBlue. RNA was fragmented by incubation with Ambion Fragmentation Reagents (AM8740) at 70°C for 10 minutes, run over an Illustra MicroSpin G-25 column (GE Healthcare, 27532501) according to the manufacturer's protocol and DNase treated (Promega, M6101) for 10 minutes at 37°C. To capture the labeled RNA, two rounds of bead binding were performed. Therefore the samples were incubated for 1 h with Anti-BrU agarose beads (Santa Cruz, sc-32323 AC), followed by extensive washing and elution. Eluted RNA was phenol/chloroform extracted and ethanol precipitated. After the second RNA precipitation samples were immediately used for library preparation. The sequencing library was prepared using the NEBNext Ultra Directional RNA Library Prep Kit (NEB, E7420S) following the manufacturer's protocol for highly degraded RNA with RNA integrity number (RIN) $\leq$ 2. The library quality check and sequencing was performed at the Genomics and Microarray Core Facility at the University of Colorado Anschutz Medical Campus and at the BioFrontiers Sequencing Facility at the University of Colorado Boulder. Single-end fragment libraries (50 bp) were sequenced on the Illumina HiSeq 4000 platform. Raw sequencing reads were demultiplexed, and after barcode, adaptor and quality trimming with cutadapt (https://doi.org/10.14806/ej.17.1.200), quality control was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The remaining reads were mapped to the GRCm38/mm10 mouse genome assembly using genomic short-read RNA-Seq aligner STAR version 2.5 (Dobin et al., 2013). We obtained at least 65% uniquely mapped reads in each sample. Data analysis and visualizations were performed in R-project version 3.4.2 with Rstudio IDE version 1.0.143. Reads were counted in 2 intervals per transcript (relative to transcription start site, interval 1: (−500;500), interval 2: (501:end)) using featureCounts (Liao et al., 2014). Pausing index was calculated as ratio of read counts in interval 1 to read counts in interval 2 normalized to the length of intervals. Inhibitor response was calculated as ratio of pausing index in the presence of inhibitor to pausing index in the control. Calculation of transcripts per million reads (log_2TPM) excluded the first 500 bases downstream the transcriptional start side to minimize effects of RNAPII pausing. PI was calculated both for individual replicates and pooled replicates (by summing up counts). For gene set enrichment analysis (GSEA) the GSEA preranked module on the GenePattern server (Reich et al., 2006) was used, with log_2-fold-change values for all detected genes for the indicated comparisons as the ranking metric. Genes were required to have TPM > 0.5 in all replicates and conditions.

The GRO-Seq data are accessible via the NCBI's Sequence Read Archive (SRA) database (accession number PRJNA542065).

PRO-Seq

Nuclei Preparation: HCT116 cells (WT or CDK8as) were seeded on three 15 cm dishes (1x10^7 cells/dish), 24 h prior to the experiments (~70% confluency at time of experiment). Cells were treated simultaneously with 10 ng/ml IFN-γ and/or 10 μM 3MB-PP1 for 45 min, washed 3x with ice cold PBS, and then treated with 10 mL (per 15 cm plate) ice-cold lysis buffer (10 mM Tris–HCl pH 7.4, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% NP-40, 10% glycerol, 1 mM DTT, 1x Protease Inhibitors (1 mM Benzamidine (Sigma B6506-100G), 1 mM Sodium Metabisulfite (Sigma 255556-100G), 0.25 mM Phenylmethylsulfonyl Fluoride (American Bioanalytical AB01620), and 4 U/ml SUPERase-In) and scraped from the plates. Cells were centrifuged 1000 g for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1.5 mL lysis buffer to a homogeneous mixture by pipetting 20-30X before adding another 8.5 mL lysis buffer. Suspension was centrifuged with a fixed-angle rotor at 1000 g for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1 mL of lysis buffer and transferred to a 1.7 mL pre-lubricated tube (Costar cat. No. 3207). Suspensions were then pelleted in a microcentrifuge at 1000 g for 5 min at 4°C. Next, supernatant was removed and pellets were resuspended in 500 μL of freezing buffer (50 mM Tris pH 8.3, 40% glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA, 4 U/ml SUPERase-In). Nuclei were centrifuged 2000 g for 2 min at 4°C. Pellets were resuspended in 100 μL freezing buffer. To determine concentration, nuclei were counted from 1 μL of suspension and freezing buffer was added to generate 100 μL aliquots of 10x10^6 nuclei. Aliquots were flash frozen in liquid nitrogen and stored at −80°C.

Nuclear run-on and RNA preparation

Nuclear run-on experiments were performed as described (Mahat et al., 2016) with the following modifications: the final concentration of non-biotinylated CTP was raised from 0.25 μM to 25 μM, and the final library clean-up and size selection was accomplished using 1X AMPure XP beads (Beckman).

Sequencing

Sequencing of PRO-Seq libraries was performed at the BioFrontiers Sequencing Facility (UC-Boulder). Single-end fragment libraries (75 bp) were sequenced on the Illumina NextSeq 500 platform (RTA version: 2.4.11, Instrument ID: NB501447), demultiplexed and converted BCL to fastq format using bcl2fastq (bcl2fastq v2.20.0.422); sequencing data quality was assessed using FASTQC

(v0.11.5) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FastQ Screen (v0.11.0, https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/). Trimming and filtering of low-quality reads was performed using BBDUK from BBTools (v37.99) and FASTQ-MCF from EAUtils (v1.05) (Kechin et al., 2017). Alignment to the human reference genome (GRCh37/hg19) was carried out using Hisat2 (v2.1.0) (Kim et al., 2015) in unpaired, no-spliced-alignment mode with a GRCh37/hg19 index, and alignments were sorted and filtered for mapping quality (MAPQ > 10) using Samtools (v1.5) (Li et al., 2015). Gene-level count data for transcription start site (TSS, −30 to +300) and gene body (+301 to end) regions were obtained using featureCounts from the Subread package (v1.6.2) (Liao et al., 2013) with custom annotation files for single unique TSS and gene body regions per gene. Custom annotation files with single unique TSS and gene body regions per gene were generated as follows: 1) hg19 RefSeqCurated transcript-level annotation was downloaded from the UCSC genome table browser (09-07-2018), transcripts shorter than 1500 bp and non-standard chromosome were removed, and only transcripts with unique start/stop coordinates per gene were retained; 2) Sense and anti-sense counts were tabulated and each candidate TSS region was ranked by sense and antisense reads to obtain a single 'most-active' TSS per gene; 3) Finally, per gene, the TSS was combined with the shortest gene body to avoid the influence of alternative transcription termination/polyadenylation sites. Differential expression analysis of gene body regions was assessed using the DESeq2 package (v1.22.1) (Love et al., 2014) with a custom R script (R v3.5.1 / RStudio v1.1.453 / Bioconductor v3.7) with cutoffs as described in text and figure legends. Analysis of RNAPII pausing was carried out using a custom R script (R v3.5.1 / RStudio v1.1.453) with the ggplot2 package (v3.1.0) used for visualizations. Gene level TSS and gene body counts were normalized by counts-per-million and by region length (cpm/bp), and Pausing Index (PI) calculated as the ratio of normalized reads in the TSS (cpm/bp) to normalized reads in the gene body (cpm/bp). Genes with < 0.5 cpm in all samples were excluded from analysis. Means of replicate values were used for plots and Wilcoxon/Mann-Whitney U tests. For genome browser snapshots, aligned reads were downsampled to the lower aligned read count per replicate using Samtools, to ensure equal contributions from each replicate, followed by merging of replicates and generation of coverage tracks in the bedgraph format using HOMER (v4.9.1) (Heinz et al., 2010) Genome browser snapshots were then generated from the bedgraph files using a custom R script (R v3.5.1 / RStudio v1.1.453 / Bioconductor v3.7) and the Gviz package (v1.26.4) (Hahne and Ivanek, 2016). PRO-Seq data are accessible via the NCBI's Gene Expression Omnibus (GEO) database (accession number GSE129501).

Modified motif displacement score (MD Score) analysis

We performed the motif displacement (MD) analysis as described (Azofeifa et al., 2018) with the top 20% differentially transcribed enhancer RNAs (eRNAs) quantified by DE-Seq2. This modification was made to improve the signal-to-noise ratio of the MD analysis. For the analysis of TF motifs associated with eRNA transcription we used the hand-curated database of TF binding motif models HOCOMOCO (Kulakovskiy et al., 2013). We note that within the HOCOMOCO database for mouse TF binding motifs, the principal binding model for STAT1 (motif ID: STAT1_MOUSE.H11MO.0.A) corresponds to IRF binding motifs whereas the alternate binding model (motif ID: STAT1_MOUSE.H11MO.1.A) corresponds to the canonical STAT1 motif (Decker et al., 1997; Mancino and Natoli, 2016). We have included this information in the table containing MD scores (Table S6) and used the correct designation in the MD score figure (Figures 3A and B).

Inducible CDK8 knockout

Inducible CDK8 knockout in CDK8fl-MEFs was induced by 4-hydroxytamoxifen (4OHT) treatment (3 h, 250 nM) in low fetal calf serum medium (2% FCS), followed by recovery (2 days). CDK8 knockout was validated by genotyping using the primers CDK8 Intron4/Exon5 fwd 5′-AATAGGTGTGTATCTTATGGCTTCC-3′ and CDK8 Intron4/Exon5 rev 5′-ATTTTTACTCTTCCTCGCTCAGGAC-3′ and by western blotting.

Knockdown of CDK19

Silencing was performed as described (Bancerek et al., 2013). In brief, approximately $7x10^4$ cells were seeded on a 6 cm dish and incubated 7 h followed by transfection with 100 pmol ON-TARGET plus™ SMART pool siRNA targeting CDK19 (Dharmacon, L-059630-00-0010) or non-targeting control (Dharmacon, D-001810-10-20) using Lipofectamine RNAiMAX Reagent (Invitrogen, 13778-150) in Opti-MEM I (GIBCO, 31985070) for 48 h.

Knockout of CDK19

CDK19 knockout was generated in CDK8fl-MEFs using CRISPR/Cas9. Guide RNA (gRNA) sequence (5′-GTACAGCAGTGATTTAACCATGG-3′) targeting exon 4 of CDK19 was designed using the CRISPOR tool (Haeussler et al., 2016). The gRNA together with the purified Cas9 protein was delivered into CDK8fl-MEFs by electroporation. Single cells were grown and screened for the successful knockout using TIDE (Brinkman et al., 2014), sequencing and western blotting.

HAP1 cells expressing analog-sensitive CDK8 (CDK8as)

CRISPR/Cas9-mediated homologous recombination was used to engineer an ATP analog-sensitive CDK8 mutant (CDK8as) in HAP1 cells. To introduce the F97G mutation in the exon 3 of CDK8 a guide RNA (gRNA; 5′-TGTTTCTGTCTCATGCTGAT-3′) was designed using the Broad Institute sgRNA designer (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). A homology-directed repair template (HDRT) harboring the exon 3 with phenylalanine 97 changed to glycine (F97G) (nt change CTG TTT to

CellPress

CTC GGG), a mutated PAM site (silent mutation), a PGK-neomycin phosphotransferase (Neo) gene flanked by flippase recognition target (FRT) sites and flanking homology arms up- and downstream of exon 3 was generated. The HDRT was cloned together with the gRNA into the pSpCas9(BB)-2A-Puro (PX459) plasmid (Addgene #48139, (Ran et al., 2013)). Transfected cells were selected using 1 μg/ml puromycin for 24 h followed by 3 mg/ml G418 (InvivoGen) treatment on day 6 post-transfection. Surviving clones were analyzed for successful knock-in by PCR-based genotyping and sequencing. In the obtained positive knock-in clones the neomycin selection cassette was removed by transient expression of flippase (pOG44 Flp-Recombinase expression vector, Thermo Scientific, V600520) followed by a negative selection treating the cells with 3 mg/ml G418 (InvivoGen).

CDK19-WT expression construct

Total RNA from WT MEF cells was isolated with the Trizol – Isopropanol method (QIAzol, #79306, QIAGEN), followed by DNA-digestion (DNase I #4716728001, Merck), acidic-phenol-chloroform clean-up, isopropanol RNA precipitation and reverse transcription with RevertAid Reverse Transcriptase (#EP0442, Thermo Scientific) and oligo dT_{18} primers (Eurofins Genomics) according to manufacturer's protocol. Nested PCR strategy was applied to amplify and modify the coding sequence of CDK19 for cloning. First, the cDNA was amplified with Q5® High-Fidelity DNA Polymerase (#M0491L, NEB) and target specific primers (Forward: 5′-GAGGA GGCGGGACTGTAGAT-3′, Reverse: 5′-TTTGCATGGTGTCAGTCTTCATTC-3′) followed by gel purification (Monarch DNA gel extraction kit #T1020L, NEB). Primers of the second PCR were designed in a way that they flank the CDS, remove the stop codon and introduce a 5′ XbaI and 3′ XhoI NotI cleavage site (Forward: 5′-CATTCTAGACCGAGGAGTCCCTTGCTGAA-3′, Reverse: 5′-TATGCGGCCGCTATCTCGAGTACCGGTGGGTCTGGTGAGAT-3′). After gel purification and sequence validation, the PCR product and the PiggyBac-EF1-MCS-IRES-Neo cDNA Cloning and Expression Vector (#PB533A-2, SBI System Biosciences) were double digested using XbaI (#FD0684, Thermo Scientific) and NotI (#FD0594, Thermo Scientific) followed by gel purification, T4 DNA Ligase (#EL0011, Thermo Scientific) mediated ligation, transformation into chemically competent DH10B and ampicillin selection (100 μg/ml, #A0839, AppliChem). Colonies were PCR-screened (Forward: 5′-CAATTGAACGGGTGCCTAGAG-3′, Reverse: 5′-CCTT GTTGAATACGCTTGAGGAGA-3′) and plasmids were isolated by using plasmid mini prep kit (QIAprep, #27106, QIAGEN). C-terminal triple-Flag-tag was introduced into double-digested (XhoI, #FD0694 Thermo Scientific and NotI in the presence of alkaline phosphatase, FastAP, #EF0651 Thermo Scientific) plasmid using double strand oligonucleotide (purchased from Integrated DNA Technologies as single strand oligonucleotides: +strand 5′ TCG Agagac tacaaagaccatgacggtgattataaagatcatgacatcgattacaaggatgacg atgacaagTAGC, -strand 5′ GGCCGCTActtgtcatcgtcatccttgtaatcgatgtcatgatctttataatcaccgtcatggtctttgtagtctC). Sequence of final plasmids (pB_EF1_CDK19_IRES_Neo) was validated with PCR and sequencing.

CDK19-KDead expression construct

Kinase-dead CDK19 (CDK19-KDead) was generated by mutation of aspartate 151 to alanine in the consensus active site. The D151A mutation was introduced in the parental plasmid (pB_EF1_CDK19_IRES_Neo) using Gibson Assembly site-directed mutagenesis. Briefly, PCR-amplified fragments (PCR_A: Forward: 5′-agctgtgaccggcgcctactctagaCTAGATGGGGGAAGCAGACAATGG-3′, Reverse: 5′-ctggtttcagTGCCCTGTGGAGCACCCA-3′; PCR_B: Forward: 5′-ccacagggcaCTGAAACCAGCAAATATCC-3′, Reverse: 5′-taggggggggggagggagaggggcgcggccgcGCTACTTGTCATCGTCATC-3′) were purified with column DNA Clean-up kit (Monarch® PCR & DNA Cleanup Kit, # T1030L, NEB) and mixed with XhoI/NotI double digested and gel purified plasmid in a molar ratio of insert to plasmid of 1:2 (0.02pmol for the 6,886bp parental plasmid, 0.04pmol of the 508bp PCR_A and of the 1,168bp PCR_B, respectively), followed by addition of 10 μL 2x Gibson Assembly Master Mix (E2611L, NEB) and incubation at 50°C for 90 min. 5 μl of the reaction mix were used for transformation of 100 μl chemically competent DH10B.

Rescue of CDK19-KO cells with CDK19-WT or CDK19-KDead

CDK19-KO cells ($8x10^5$) were electroporated with 1 μg piggyBac transposase plasmid (Cadiñanos and Bradley, 2007) and 10 ng CDK19-WT plasmid or 50 ng CDK19-KDead plasmid. Cells were selected for integration of the transposon by G418 treatment (400 μg/μl) for 10 days and subsequently analyzed for CDK19 expression by qRT-PCR and western blotting.

Generation of CDK8as mutant for *in vitro* kinase assays

N-terminally glu-tagged CDK8 in a pKozak plasmid (Knuesel et al., 2009) was used to generate the analog-sensitive CDK8 mutant (CDK8as) by mutation of the F97 codon (TTT) to a glycine (GGG) using site-directed mutagenesis. The mutated glu-CDK8as was transformed into DH5α *E. coli*, and individual clones were selected, grown, purified, and sequenced. Sequence-verified glu-CDK8as was cloned into the Baculovirus transfer vector pACEBac1. The resulting vector was transformed into DH5α *E. coli* and colonies were selected, purified, and sequenced. Sequence-verified pACEBac1 containing glu-CDK8as was used to transfer glu-CDK8as plus gentamycin resistance into the baculoviral genome using Tn7 transposition. Successful integration was assessed by blue/white screening plus gentamycin resistance. Bacmid DNA was prepared from selected clones and used to transfect insect cells for protein production.

In vitro kinase assays

Wild-type (WT CDK8) and analog-sensitive CDK8 (CDK8as) modules (containing CDK8, CCNC, MED12, and MED13) were purified and assembled as described (Knuesel et al., 2009). Reactions were performed at 30°C for 45 minutes in kinase buffer (25 mM Tris

pH 7.9, 100 mM KCl, 300 μM ATP, 10 mM MgCl2, and 2 mM DTT). WT CDK8 and CDK8as module, STAT1 substrate (2μL), 2.5 μCi [γ-^{32}P]ATP and increasing concentrations of NM-PP1 ATP analog (1 to 100 μM) were added. SDS-PAGE was used to separate proteins and the gels were subsequently stained with Coomassie, dried at 55°C for 60 min, exposed on a phosphor-imager screen for 72 hours, and imaged using a Typhoon 9400 scanner. Quantitation of auto-rad bands was performed using ImageJ.

Whole cell extracts and western blotting

The procedures for whole cell extracts and immunoblotting were described (Sadzak et al., 2008). In brief, whole cell extracts were prepared by lysing the cells for 5 minutes in Frackelton buffer (10mM Tris-HCl, 30 mM $Na_4P_2O_7$, 50mM NaCl, 50 mM NaF, 1% Triton X-100, 1 mM DTT, 1 mM vanadate and 1x protease inhibitor (Roche, 11836145001)). Lysates were cleared by centrifugation at 13200 rpm and 4°C. SDS loading buffer was added in a 2:1 ratio (lysate:loading buffer) and boiled for 5 minutes. Antibodies against pSer727 STAT1 (Kovarik et al., 1999; Cell signaling, 9177), pY701 STAT1 (Cell Signaling, 9167S and 7649), STAT1 (Santa Cruz, sc-346), CDK8 (Santa Cruz, sc-1521; Cell signaling, 4101S), CDK19 (Sigma Aldrich, HPA007053), IRF1 (Cell Signaling 8478T), tubulin (Cell signaling, 2144S; Sigma Aldrich, T9026), MED12 (Bethyl, A300-774A), MED13 (Santa Cruz, sc-515557) and CCNC (Bethyl, A301-989A) were used for western blotting.

Cytotoxicity assay

Cells were seeded on a 96well plate with 9000 cells per well. The next day the cells were treated with 100 nM CA for 2, 4, 6 and 8 hours or left untreated. The amount of living cells was assessed by crystal violet staining (0.1% crystal violet, 2% methanol in H_2O) for 1 hour in the dark. Cells were then washed twice with PBS, air-dried and subsequently incubated with 100 μL solubilization buffer (50:50 mixture of 0.1 M NaH_2PO_4, pH = 4.5 and 50% ethanol) per well. Crystal violet intensity, that was proportional to the number of alive cells, was determined at 595 nm using a microplate reader (BioTek, Synergy H1). n = 8 per condition, statistical testing was done using One-way ANOVA testing in Prism 6 (GraphPad Software).

Vesicular stomatitis virus (VSV) infection assay

VSV infection and survival assays was performed as described (Bancerek et al., 2013). Briefly, $3x10^4$ cells were seeded on 6-well plates. 4 hours after seeding the medium was exchanged to medium containing siRNA against non-targeting control or CDK19 (50 pmol). After 48 h cells were re-seeded on 96well plates, 3500 cells per well. Four hours after seeding the medium was replaced with new medium that was supplemented with fresh siRNA and IFN-γ in two-fold serial dilutions starting at 10 units. After 24 hours the medium was replaced with medium without siRNA and without IFN-γ, and VSV was added at a multiplicity of infection (MOI) of 0.1. In case of CDK19-KO, CDK19-WT and CDK19-KDead cells 4000 cells were seeded on 96well plates. Four hours after seeding medium was replaced with medium supplemented with IFN-γ in two-fold serial dilutions starting at 10 units. After 24 hours the medium was replaced with medium without IFN-γ, and VSV was added at a multiplicity of infection (MOI) of 0.1. After incubation of cells with VSV for 39 h, cells were washed twice with PBS and stained with crystal violet (0.1% crystal violet, 2% methanol in H_2O) for 1 h in the dark. After 2 additional washes, cells were incubated with 100 μL solubilization buffer (50:50 mixture of 0.1 M NaH_2PO_4, pH = 4.5 and 50% ethanol) per well. Crystal violet intensity, that was proportional to number of surviving cells, was determined at 595 nm using a microplate reader (BioTek, Synergy H1). EC_{50} calculations were calculated using the AAT-Bioquest EC_{50} (https://www.aatbio.com/tools/ec50-calculator), with minimum response set to zero.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-Seq experiments were carried out in triplicates; GRO-Seq and PRO-Seq experiments were performed in duplicates. Statistical analysis of RNA-Seq, GRO-Seq and PRO-Seq experiments is described in the corresponding parts of the section Method Details.

DATA AND CODE AVAILABILITY

The accession number for GRO_Seq and RNA_Seq data reported in this paper is SRA: PRJNA542065. The accession number for the PRO_Seq data reported in this paper is GEO: GSE129501.

Unprocessed western blot images of this study are available at Mendeley under: https://doi.org/10.17632/crj8f3j63z.1

Chapter 5

Mediator Kinase Activity in Response to Serum

5.1 Preamble

This chapter is a work in progress towards submission of a manuscript related to the effects of Mediator kinase inhibition during serum response. A preliminary title for this manuscript along with an author list is as follows:

Mediator kinases Target Signaling Pathways Involved in Cell Proliferation during Serum Reponse

Jonathan D. Rubin, Justin Moser, Sabrina L. Spencer, Robin D. Dowell, Dylan J. Taatjes

5.2 Introduction

Gene regulation is an essential process that is required for proper cellular function. This regulation occurs primarily at the transcriptional level and involves the interplay of many key factors. Among these factors is a large protein complex called Mediator. Mediator is necessary for regulated transcription of protein coding genes and acts as a molecular bridge between distal transcription factor (TF) binding events and the pre-initiation complex (PIC) containing RNA-polymerase II (RNAPII).

In addition to the scaffolding functions of the Mediator complex, Mediator can also reversibly associate with four-subunit Mediator kinase modules that contain enzymatic activity. In humans, there are two Mediator kinases - CDK8 and its paralog CDK19. Although not much is known about CDK19, CDK8 has been shown to be overexpressed in several cancers and plays roles in general

transcription as well as cellular signalling (Steinparzer 2018, others). While the mechanisms by which Mediator kinases exert their functions are unclear, even less is known about the kinase activity associated with them.

Studying Mediator kinase activity has received special attention recently because of its potential use as a therapeutic target in cancer (refs). CDK8 is often overexpressed in colorectal tumors and its overexpression correlates with poor patient outcomes (refs). While total protein knockdown/knockout is difficult to achieve in a clinical setting, inhibiting Mediator kinase activity with a small molecule is more tractable. Because of this, there has been a drive to screen small molecules for their potential to inhibit Mediator kinase activity (refs). One such molecule, called Cortistatin A (CA), was shown to be both extremely potent and specific for inhibiting Mediator kinase activity (Pelish).

Because of its potential roles in cancer, some attention has been given to how Mediator kinases could affect cellular proliferation. One such study determined that the CDK8 protein itself played important roles during serum response. This study however, performed whole protein knockdown and therefore could not decouple the roles of Mediator Kinase activity vs. Mediator kinase protein. Additionally, several studies have recently focused on the role of Mediator kinase activity within cellular signaling pathways. Studies have shown that Mediator kinases play roles in the interferon response (Steinparzer 2018), Glycolysis (espinosa ref), and p53 activation (Audetat ref). We therefore sought to characterize the effects of Mediator kinase inhibition within the context of serum response. This model system mimics cellular proliferation and known nutrient deprivation seen in tumors before angiogenesis.

In this study, we performed a variety of -omics experiments in the presence of CA during serum response at many early (15min, 30min, 45min) and late (3hrs, 6hrs, 18hrs) timepoints. We performed 1) nascent RNA sequencing to probe for immediate effects on gene transcription, 2) RNA-seq to quantify downstream gene expression, 3) phosphoproteomics to identify key factors responsible for observed changes, 4) metabolomics to measure changes in cellular metabolism, and 5) proliferation assays to determine cell outcomes. Compared to the normal response to

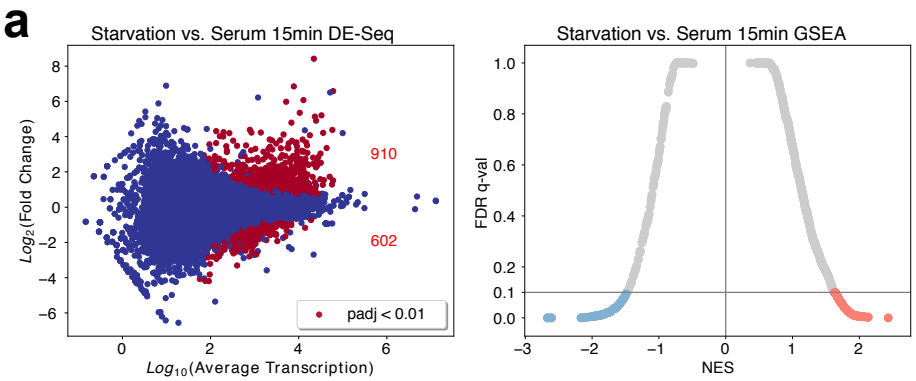
serum, cells treated with the Mediator kinase inhibitor CA exhibited defects in the timing of gene transcription and repression of protein phosphorylation involved in proliferative signaling. These phosphorylation events had an impact on cell proliferation and flux through the TCA cycle during serum response. Our results suggest that inhibiting Mediator kinase activity may affect colon cancer cell proliferation. In contrast to previous studies on normally(?) growing cells (ref liver metastasis paper), we show a decrease in cellular proliferation during serum response suggesting context specific cellular outcomes. Whether these different cellular contexts are more or less representative of cellular response *in vivo* is still an open question but we show here evidence for further study in animal models and a potential explanation for poor outcomes when transition from cell culture models to animal models.

5.3 Results

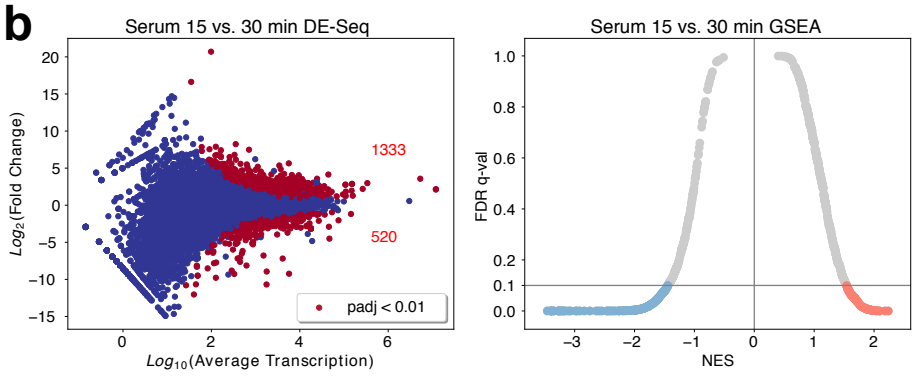
5.3.1 Serum induction results in temporally coordinated gene transcription

We performed nascent sequencing (GRO-Seq and PRO-Seq) at several time points on cells during serum response (0min, 15min, 30min, 45min) treated with the Mediator kinase inhibitor Cortistatin A (CA) or DMSO control. To first determine how cells temporally regulate cellular processes in response to serum, we performed adjacent timepoint comparisons (*i.e.* 0min vs. 15min, 15min vs. 30min, 30min vs. 45min) using DE-Seq and GSEA (Figure 5.1). As expected, the immediate serum response involved the differential transcription of hundreds of genes (910 up, 602 down) related to MAPK signaling (Figure 5.1a). We also observed a significant decrease in mitochondrial translation which is a typical response to cellular stress. A secondary response between 15 and 30min involved a relatively equal number of differentially transcribed genes (1333 up, 520 down) with an overall decrease in the magnitude of fold change of differentially transcribed genes. In this secondary response we observed the activation of cholesterol biosynthesis mediated by the TF SREBP Figure 5.1b. We also observed even greater decreases to translation pathways. Finally, at later timepoints, we observed a more modest change in gene transcription (226 up,

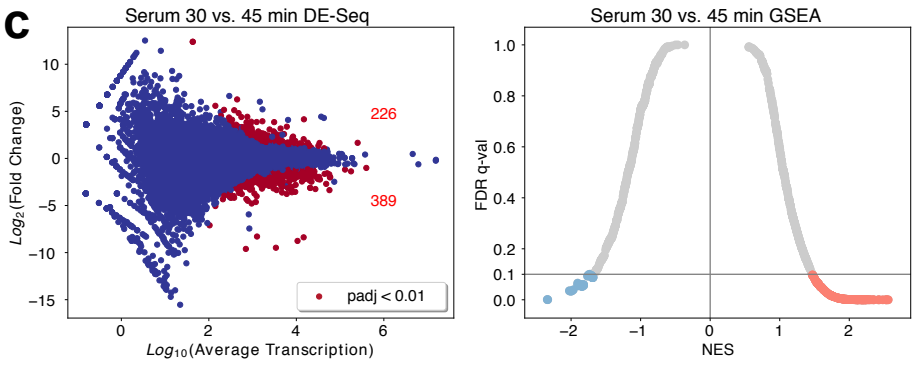
389 down) and the activation of mitochondrial translation Figure5.1c. Additionally, we observed decreases in SREBP mediated cholesterol biosynthesis and decreases in G1/S specific transcription suggesting cells might be escaping G0 arrest. These results point to a highly coordinated activation and inhibition of key pathways needed for cells to escape G0 and re-enter into the cell cycle.



Top 10 Reactome Pathways	NES	FDR
Circadian Clock	2.43	0
Negative regulation of MAPK pathway	2.11	0.00254
Interleukin-17 signaling	2.13	0.00263
MAP kinase activation	2.13	0.00282
Toll Like Receptor 10 (TLR10) Cascade	2.06	0.00436
Mitochondrial translation elongation	-2.67	0
Mitochondrial translation termination	-2.66	0
Mitochondrial translation	-2.62	0
Mitochondrial translation initiation	-2.61	0
Cleavage of the damaged pyrimidine	-2.17	0.000148



Top 10 Reactome Pathways	NES	FDR
Activation of gene expression by SREBF (SREBP)	2.24	0
Regulation of cholesterol biosynthesis by SREBP (SREBF)	2.18	0
Non-integrin membrane-ECM interactions	2.09	0
MET promotes cell motility	2.07	0.000205
Integrin cell surface interactions	2.03	0.000589
Selenocysteine synthesis	-3.46	0
Eukaryotic Translation Elongation	-3.4	0
Eukaryotic Translation Termination	-3.4	0
Viral mRNA Translation	-3.39	0
Peptide chain elongation	-3.39	0



Top 10 Reactome Pathways	NES	FDR
Mitochondrial translation termination	2.56	0
Mitochondrial translation elongation	2.54	0
Mitochondrial translation initiation	2.53	0
Mitochondrial translation	2.53	0
ER-Phagosome pathway	2.48	0
Regulation of cholesterol biosynthesis by SREBP (SREBF)	-2.34	0.000422
Activation of gene expression by SREBF (SREBP)	-2.34	0.000844
G1/S-Specific Transcription	-2.01	0.0334
Meiosis	-1.99	0.0345
Circadian Clock	-2.02	0.0353

Figure 5.1: **Serum Effects on Gene Transcription**

5.3.2 Serum results in the phosphorylation of thousands of sites within signaling networks related to cell proliferation

To probe for direct targets of Mediator kinases, we performed phosphoproteomics at 45min serum +/- CA. To determine the phosphorylation events that are important for the cellular response to serum, we first compared starved cells to cells induced with serum. As expected we observed differential phosphorylation of thousands of sites (3637 up, 2604 down; Figure 5.2a). To determine which pathways these sites were involved in, we used ingenuity pathway analysis (IPA). We found enrichment of many signaling pathways related to cell proliferation(Figure 5.2b). These pathways exhibited increased phosphorylation as evidenced by positive z-scores . We then used the molecule activity predictor function within IPA to visualize flux through the most significant pathway Rho GTPase signaling. We observed predicted activation of many components of this signaling pathway including the activation of the AP-1 complex containing Fos and JUN TFs.

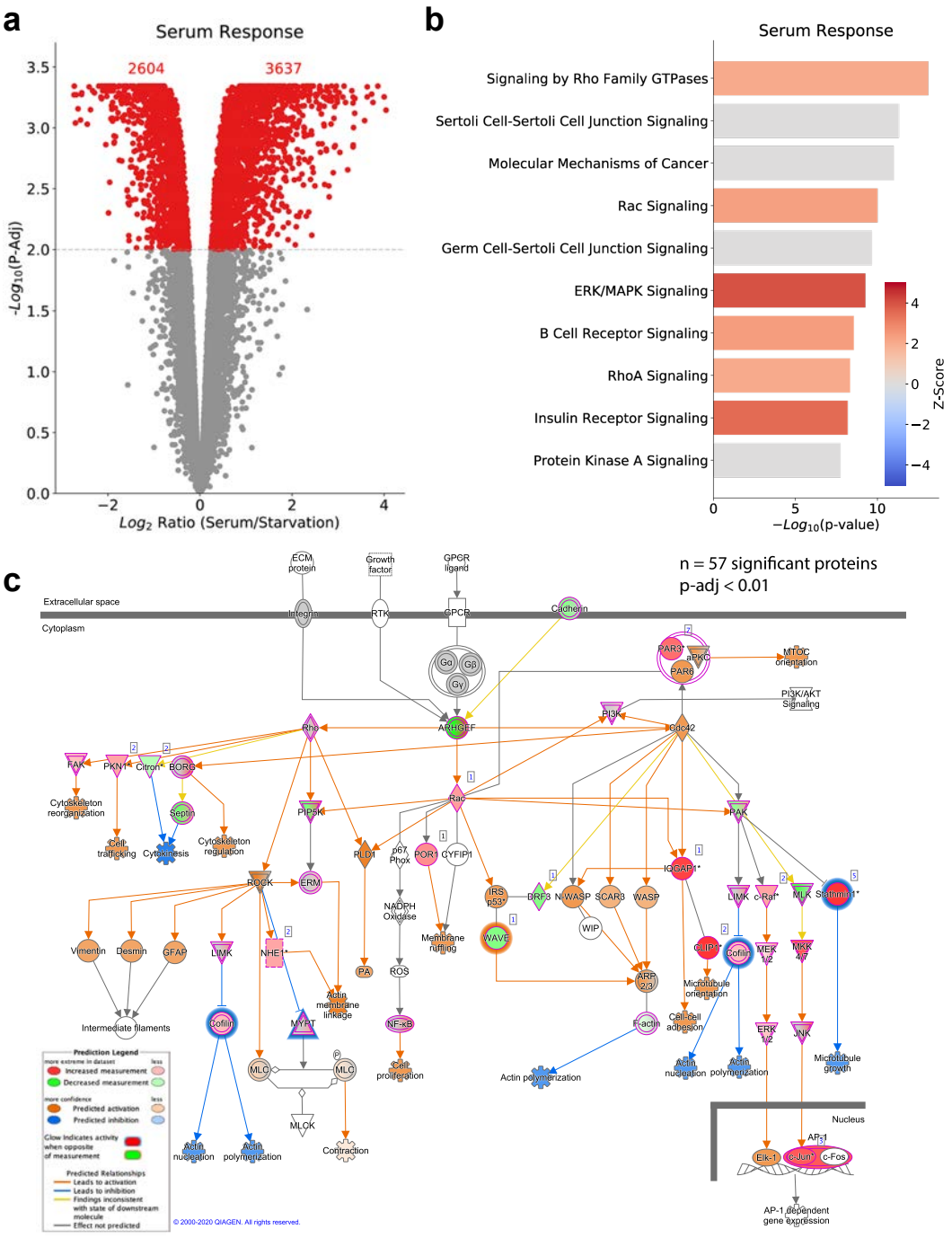


Figure 5.2: **Serum Effects on Protein Phosphorylation**

5.3.3 Serum results in increased abundance of hundreds of metabolites and flux through the TCA cycle

To probe the sustained effects of Mediator kinase inhibition, we performed metabolomics at later timepoints. To first determine the effects of serum, we compared starved cells with serum induced cells after 3hrs and 18hrs. We observed expected increases in hundreds of metabolites especially within the nucleotide, lipid, and amino acid superfamilies of molecules (Figure 5.3a). We also observed decreases in carbohydrate metabolite abundances consistent with cells increasing flux through glycolysis and the TCA cycle. Consistent with this observation, we see significant increases in key intermediates within the TCA cycle (Figure 5.3b).

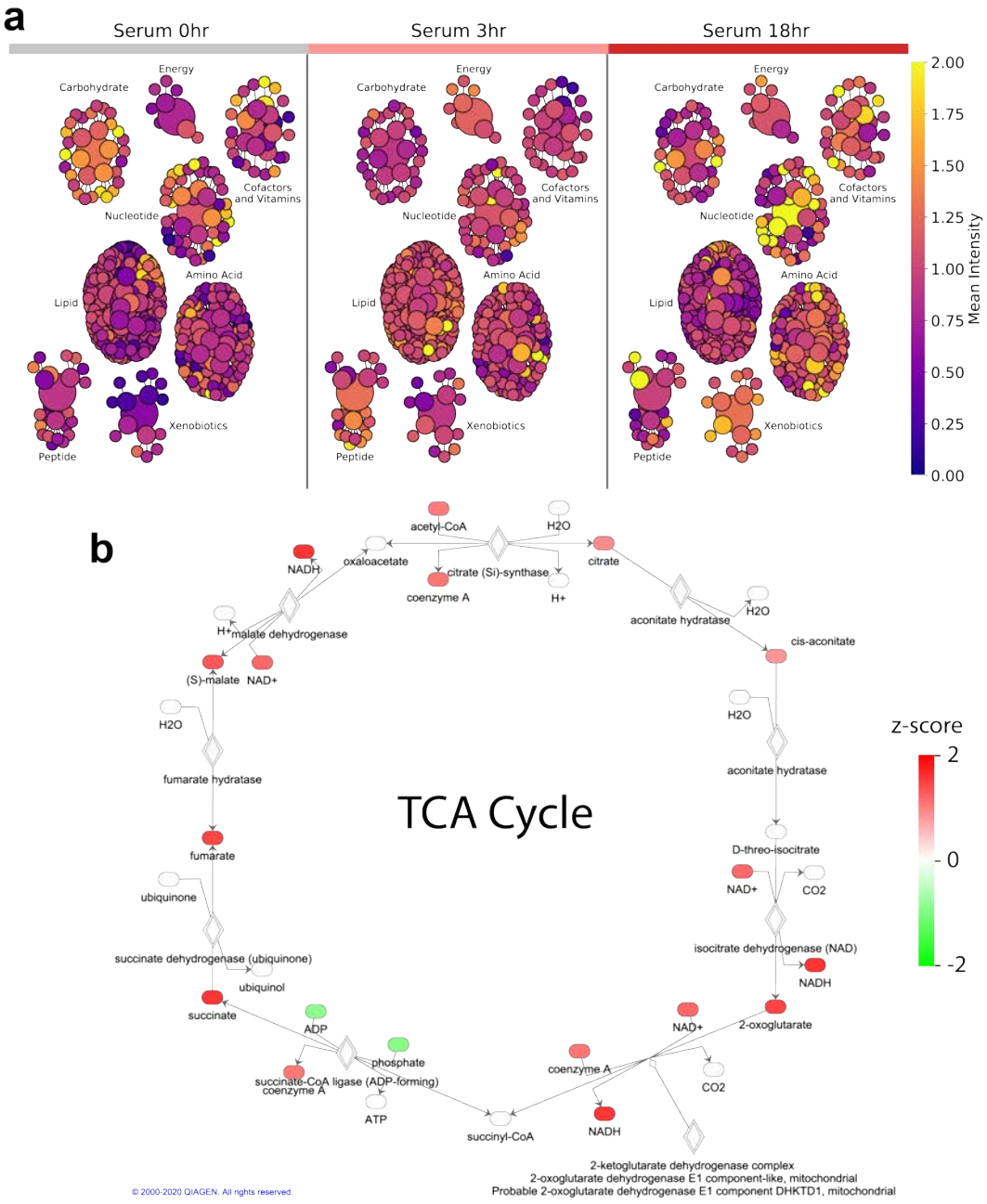
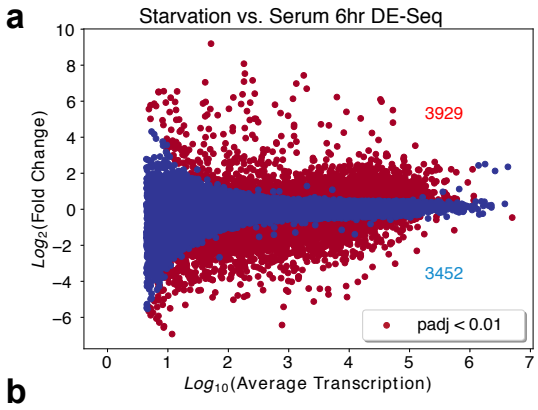


Figure 5.3: **Serum Effects on Cell Metabolism**

5.3.4 Late gene expression timepoints show cells maintain expression of genes related to RNA processing and DNA replication

To determine the gene expression patterns of cells after 6hrs of serum induction, we performed RNA-Seq experiments. Comparing the starvation to serum conditions, we found the differential expression of thousands of genes (3929 up, 3452 down) Figure 5.4a. Analyzing these gene sets with GO enrichment, We recover RNA processing and DNA replication pathways suggesting cells are beginning to transition into S-Phase (Figure 5.4b). Within the differentially decreased set of genes, we observe enrichment of RNAPII transcription pathways and cholesterol biosynthesis.



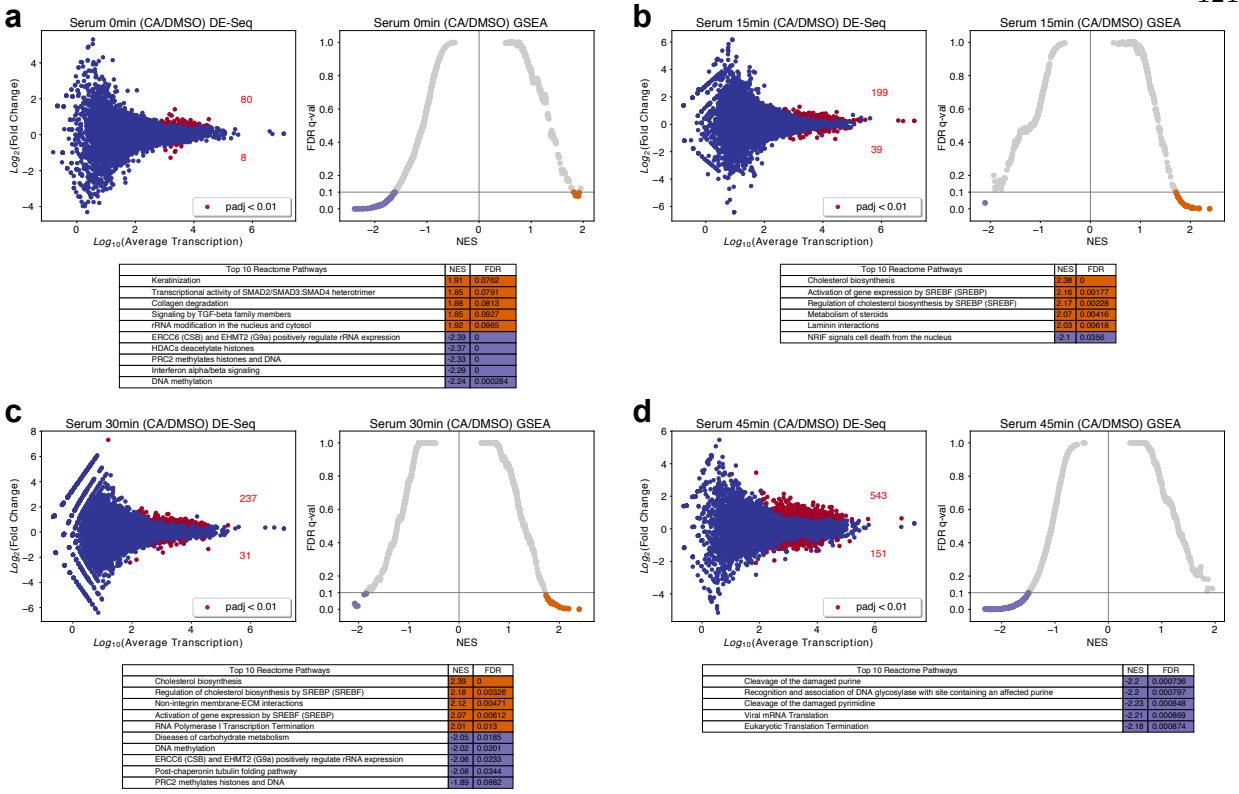
GO Pathway	Ratio	FDR
rRNA processing	(86, 3547)	4.11e-28
mRNA splicing, via spliceosome	(102, 3547)	5.45e-18
DNA replication	(62, 3547)	1.71e-12
mitochondrial translational elongation	(46, 3547)	7.39e-11
mRNA export from nucleus	(50, 3547)	8.25e-11
positive regulation of transcription by RNA polymerase II	(237, 2805)	1.7e-09
negative regulation of transcription by RNA polymerase II	(182, 2805)	2.86e-08
regulation of transcription by RNA polymerase II	(160, 2805)	9.19e-08
positive regulation of transcription, DNA-templated	(125, 2805)	2.37e-05
regulation of cholesterol biosynthetic process	(17, 2805)	0.000335

Figure 5.4: **Serum Effects on Gene Expression**

5.3.5 Inhibition of Mediator kinases during serum response results in temporal dysregulation of gene transcription

Upon full characterization of the cellular response to serum, we next sought to determine the impact of Mediator kinase inhibition within the serum response. We analyzed our nascent sequencing data across treatments at identical time points after serum induction (*i.e.* DMSO vs.

CA at t=0,15,30,45min). We observed differential transcription of several hundred genes using DE-Seq spanning across the 4 measured timepoints Figure 5.5. Performing GSEA on these timepoints revealed interesting temporal effects of Mediator kinase inhibition. At 0min serum (starvation) we observed decreases in pathways associated with histone modification Figure 5.5a. We also observed changes to pathways that are known to be affected by Mediator kinases such as interferon signaling and TGF-β signaling in addition to the downstream effector SMAD family of TFs. Upon 15min induction with serum, we saw the upregulation of SREBP-mediated cholesterol biosynthesis (Figure 5.5b) - a pathway upregulated at 30min when only observing serum effects (Figure 5.1b). SREBP has been shown to be regulated by Mediator kinases and these results suggest defects in the timing of its activation following serum induction and CA treatment. At the 30min timepoint, we also observed dysregulation of cholesterol biosynthesis (Figure 5.5c), however at 45min we see a shift in pathways to exclusively repressed pathways related to nucleotide metabolism and translation Figure 5.5d. These results suggest that Mediator kinases are key players in regulating the temporal cellular response to serum.



Top 10 Reactome Pathways	NES	FDR
Keratinization	1.91	0.0762
Transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer	1.85	0.0791
Collagen degradation	1.88	0.0813
Signaling by TGF-beta family members	1.85	0.0927
rRNA modification in the nucleus and cytosol	1.92	0.0965
ERCC6 (CSB) and EHMT2 (G9a) positively regulate rRNA expression	-2.39	0
HDACs deacetylate histones	-2.37	0
PRC2 methylates histones and DNA	-2.33	0
Interferon alpha/beta signaling	-2.29	0
DNA methylation	-2.24	0.000284

Top 10 Reactome Pathways	NES	FDR
Cholesterol biosynthesis	2.38	0
Activation of gene expression by SREBF (SREBP)	2.16	0.00177
Regulation of cholesterol biosynthesis by SREBP (SREBF)	2.17	0.00228
Metabolism of steroids	2.07	0.00416
Laminin interactions	2.03	0.00618
NRIF signals cell death from the nucleus	-2.1	0.0356

Top 10 Reactome Pathways	NES	FDR
Cholesterol biosynthesis	2.39	0
Regulation of cholesterol biosynthesis by SREBP (SREBF)	2.18	0.00326
Non-integrin membrane-ECM interactions	2.12	0.00471
Activation of gene expression by SREBF (SREBP)	2.07	0.00612
RNA Polymerase I Transcription Termination	2.01	0.013
Diseases of carbohydrate metabolism	-2.05	0.0185
DNA methylation	-2.02	0.0201
ERCC6 (CSB) and EHMT2 (G9a) positively regulate rRNA expression	-2.06	0.0233
Post-chaperonin tubulin folding pathway	-2.08	0.0344
PRC2 methylates histones and DNA	-1.89	0.0882

Top 10 Reactome Pathways	NES	FDR
Cleavage of the damaged purine	-2.2	0.000736
Recognition and association of DNA glycosylase with site containing an affected purine	-2.2	0.000797
Cleavage of the damaged pyrimidine	-2.23	0.000848
Viral mRNA Translation	-2.21	0.000869
Eukaryotic Translation Termination	-2.18	0.000874

Figure 5.5: **Mediator Kinase Inhibition Effects on Gene Transcription**

5.3.6 Mediator kinases phosphorylate proteins within cell proliferative signaling networks

To identify pohsphorylation sites that could explain the difference observed after Mediator kinase inbibition, we next analyzed the phosphoproteomics data across treatments (DMSO vs. CA) at 45min serum induction. We observed differential phosphorylation of thousands of sites (805 up, 1403 down). Because this dataset deals with inhibition of Mediator kinases, we focused on the set of sites that exhibited decreased phosphorylation as putative direct targets of Mediator kinases. When we analyzed this set using IPA, we found enrichment of the same proliferative signaling pathways that increased in phosphorylation with serum alone (Figure 5.6b). Performing molecule activity prediction within IPA, we found that signaling through Rho family GTPases was significantly reduced following Mediator kinase inhibition with members of the AP-1 complex and

several upstream regulators identified as potential direct targets of Mediator kinase activity (Figure 5.6c).

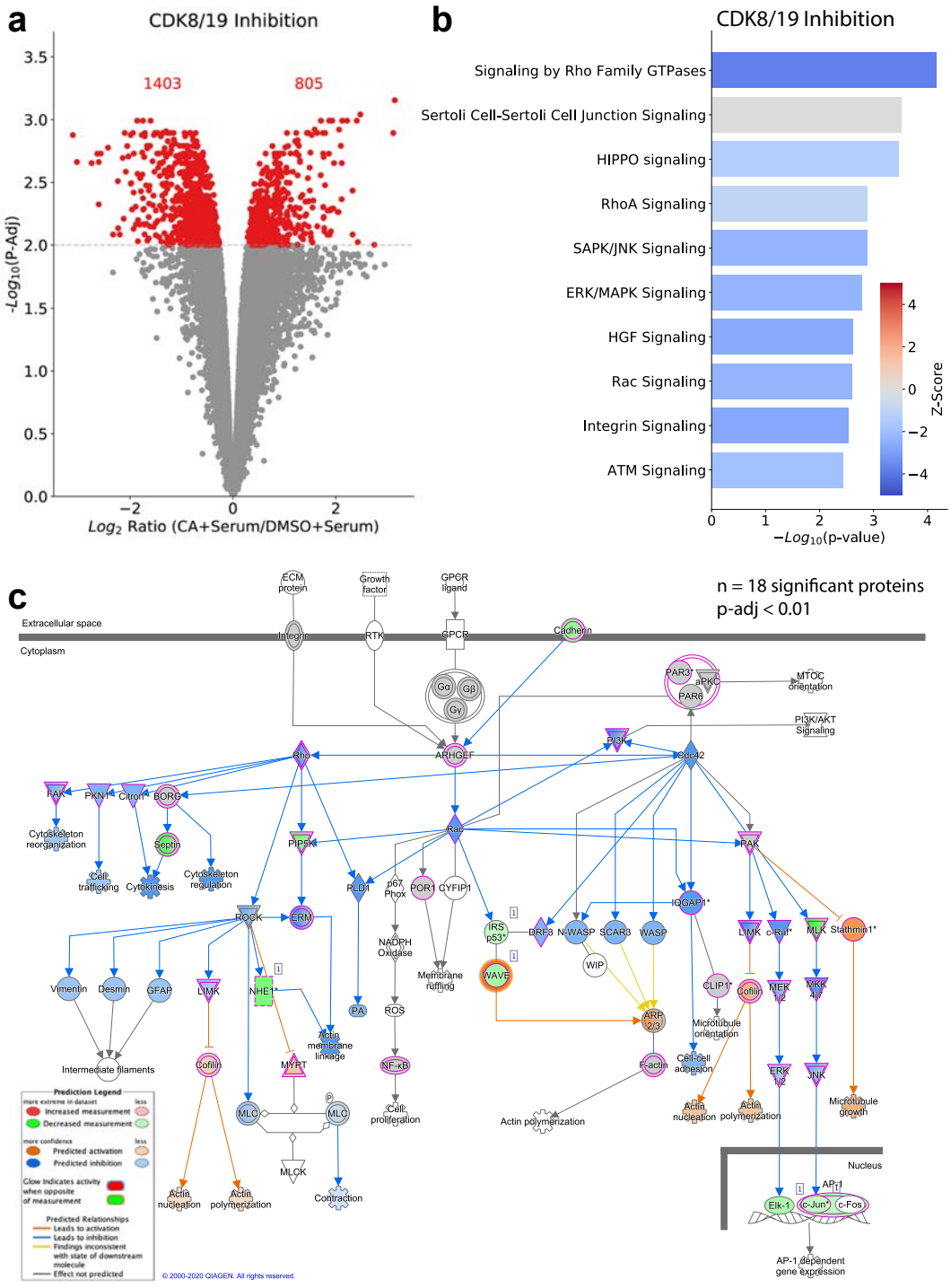


Figure 5.6: **Mediator Kinase Inhibition Effects on Protein Phosphorylation**

5.3.7 Mediator kinase inhibition results in defects in the temporal induction of JUN and FOSL2

Next, we analyzed the differential transcription of bidirectional transcripts (enhancer RNAs and promoters) within our nascent data. As expected we observed significant activation of the serum response factor (SRF) following serum induction. Additionally, we observed activation of AP-1 related factors JUN and FOSL2 (Figure 5.7a). When comparing the effects of Mediator kinase inhibition we observed a further activation of SRF and a change in the temporal dynamics of JUN and FOSL2 as evidenced by the switch in sign of E-Scores between 30min and 45min (Figure 5.7b). These results confirm the role of Mediator kinases in the temporal regulation of genes during serum response.

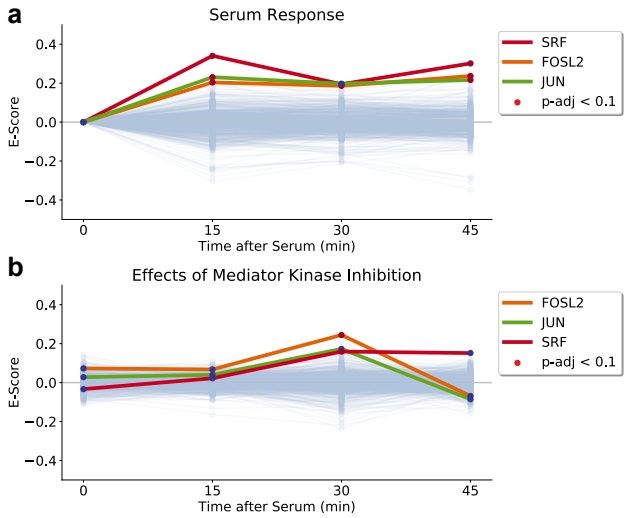


Figure 5.7: **Mediator Kinase Inhibition Effects on TF Activity**

5.3.8 At later timepoints, Mediator kinase inhibition results in decreased metabolism and flux through the TCA cycle

To determine what impacts Mediator kinase inhibition had on metabolism, we next analyzed our metabolomics dataset across treatments (DMSO vs. CA) at 3hrs and 18hrs after serum in-

duction (Figure 5.8a). We observed high dysregulation of lipid and amino acid metabolites during starvation consistent with previous studies tying Mediator kinase activity to the regulation of lipid metabolism via SREBP (ref). Upon serum induction, we observed global decreases in these same metabolites. To determine whether these cells exhibited defects in energy metabolism, we focused on the TCA cycle Figure 5.8b. We observed significant decreases in key intermediates within the TCA cycle.

5.3.9 Mediator kinase inhibition results in differentially decreased gene expression of cell division

Examining differences in gene expression +/- CA, we see dysregulation of around 1000 genes (553 up, 550 down; Figure 5.9a). Performing GO enrichment on both sets of genes we observe significant decrease in cell division and RNAPII transcription. These results suggest that cells treated with CA may have proliferation defects following serum induction.

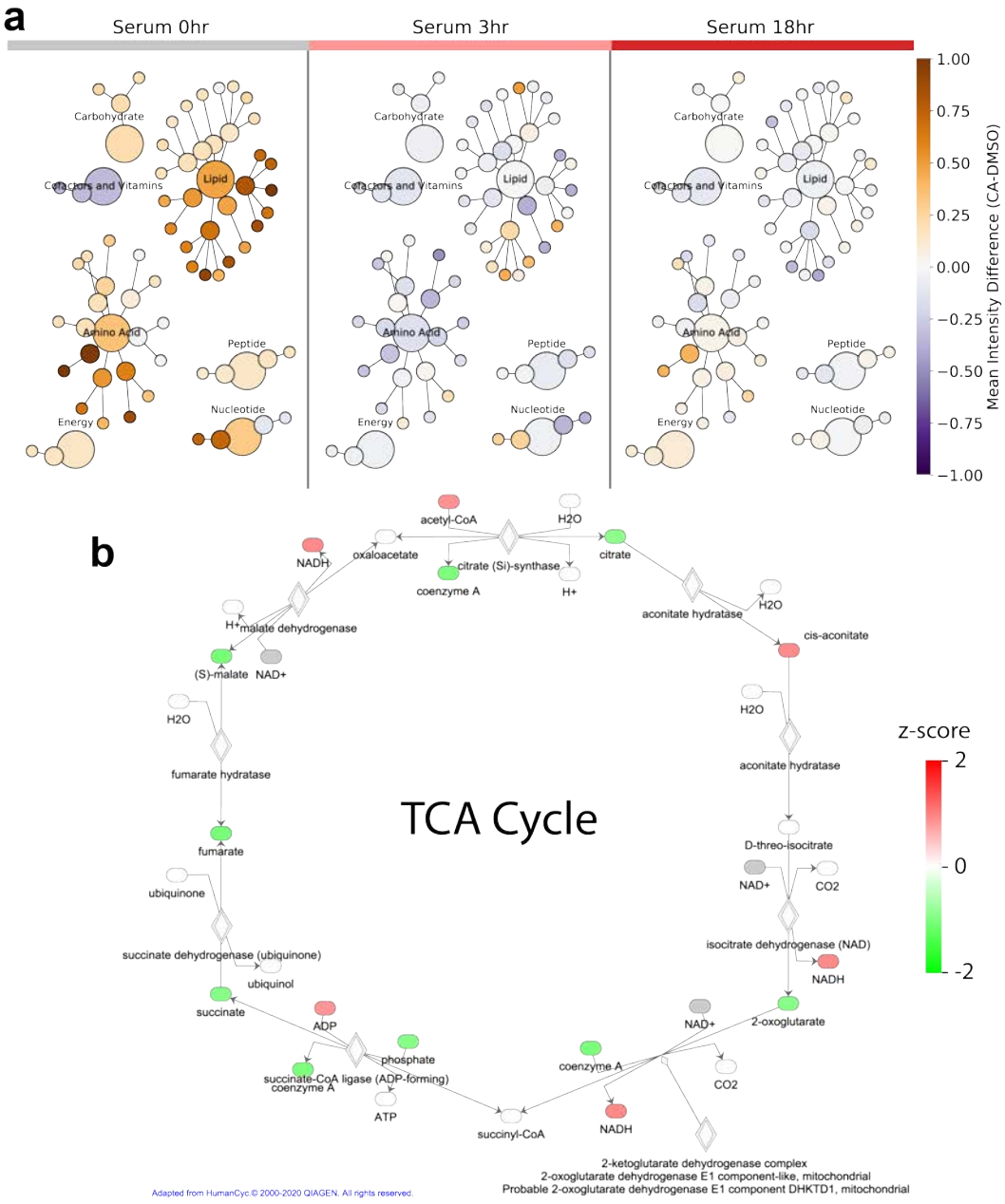
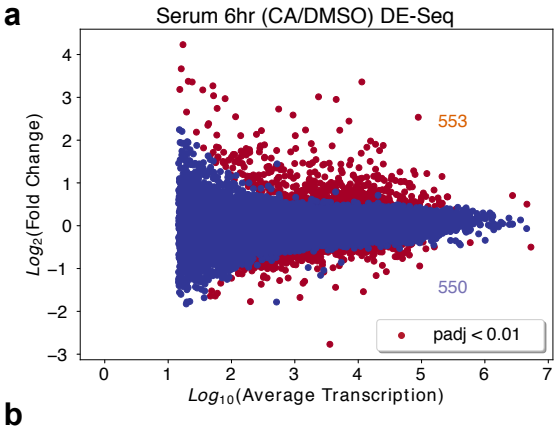


Figure 5.8: **Mediator Kinase Inhibition Effects on Cellular Metabolism**



b

GO Pathway	Ratio	FDR
cornification	(21, 494)	2.6e-09
extracellular matrix organization	(26, 494)	2.46e-07
positive regulation of gene expression	(31, 494)	2.11e-06
positive regulation of cell migration	(24, 494)	2.31e-06
epidermis development	(13, 494)	9.47e-05
positive regulation of transcription by RNA polymerase II	(64, 526)	1.04e-05
cell division	(28, 526)	0.000345
negative regulation of transcription by RNA polymerase II	(47, 526)	0.0013
sister chromatid cohesion	(6, 526)	0.0397
viral process	(24, 526)	0.0555

Figure 5.9: **Mediator Kinase Inhibition Effects on Gene Expression**

5.3.10 G1/S transition is temporally impaired in cells treated with CA during serum response

Since we observed a general repression of pathways related to cell proliferation, we sought to determine the impact of Mediator kinase inhibition on cell cycle progression following serum response. We first performed FACS using propidium iodide (PI) staining every 6hrs post-serum addition +/- CA (Fig. 5.10a). We observed that at the earliest timepoint of 6hrs there was no difference between DMSO and CA treated cells - a majority of both populations were still arrested in G1/G0. After 12 hours, we noted that both cell populations (treated and untreated) began transitioning into S-phase, however we began to see a lag with CA-treated cells as compared to DMSO cells. This lag carried over to the 12 and 18hr with the effects unclear at 24hrs since cells appeared to have entered a second round of division. To validate this temporal lag specifically in the transition from G1 to S, we performed EdU staining at the 12hr timepoint +/- CA (Fig. 5.10b). We observed a consistent decrease in DNA synthesis following CA treatment. Finally, to

determine whether these effects were truly a difference in timing or a shift in cellular populations, we performed single cell live imaging over a 24hr time course of serum response +/- CA. As a marker for cell cycle progression, we measured the activity of CDK2 as described previously (Spencer et al.). In general, we observed a decrease of CDK2 activity following CA treatment up to 24 hours after serum induction (Fig. 5.10c). Plotting average CDK2 activity over time, we see CA differences starting at 10 hrs after serum (Fig. 5.10d). Next, we clustered individual cell traces into three clusters of cell behaviors - cycling, non-arrested, and quiescent and observed decreases in cycling cells and increases in quiescent cells after CA treatment (Fig. 5.10e). Using cell populations from different wells as technical replicates, we then calculated the relative proportion of each cell population for each treatment (Fig. 5.10f). We observed that in agreement with our other data, CA-treated cells show a larger population of non-cycling cells and a decrease in proliferative cells. Overall, the data agree that cell cycle progression during serum response is impaired in cells treated with CA.

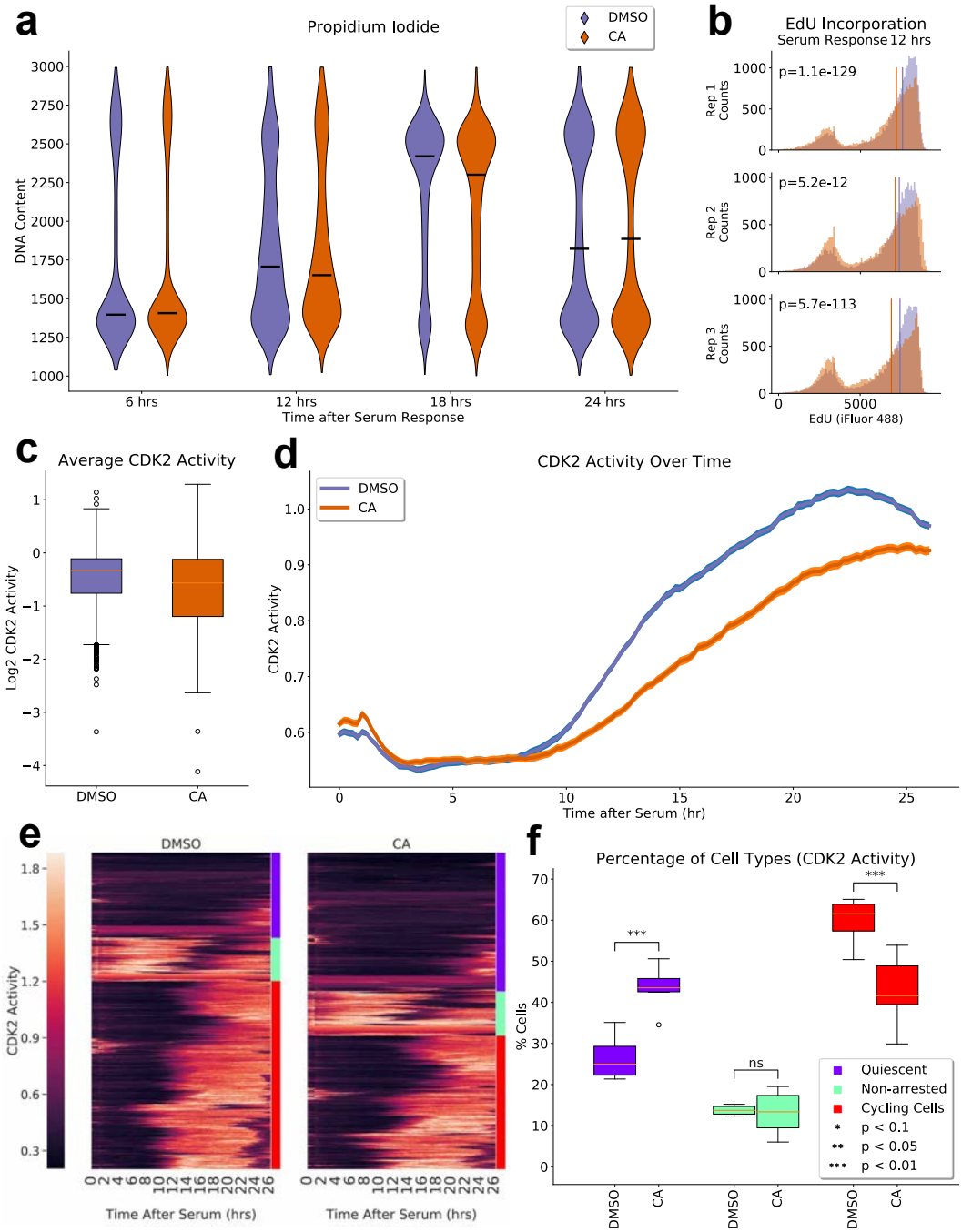


Figure 5.10: **Mediator Kinase Inhibition Effects on Cell Proliferation**

5.4 Discussion

We show here that Mediator kinase inhibition during serum response results in a lag in cell cycle progression. We observe defects in the temporal response to serum and decreased phosphorylation of signaling pathways related to cell proliferation. At later timepoints, we see decreases in genes related to cell division and a decrease in TCA cycle intermediates. Quantifying this proliferation defect in three separate assays, we see that during serum response, Mediator kinase inhibition results in cell proliferation defects.

Although we make significant strides in understanding the molecular mechanisms underlying the role of Mediator kinases in cancer cell proliferation, extensive animal studies will need to be conducted before potential therapies using CA can be administered. Even though animal studies using Senexin B, a different Mediator kinase inhibitor, did not decrease primary tumor size there were significant decreases in liver metastasis of colorectal cancer in mouse models. This type of metastasis is a known cause of mortality among cancer patients. It is unclear whether the results of this study are due to the difference in binding affinities of Senexin B and Cortsiatin A (1000-fold stronger binding), or whether this result will remain consistent in studies with CA. If these results remain the same with CA treatment in mouse models, there could be several explanations. It is possible that livers experience a higher effective dose of drug as compared to the colon which would cause proliferation defects to be more pronounced in the liver. It is also possible that metastasized cells undergo additional transcriptional reprogramming and thus are more susceptible to cellular stress and therefore are increasingly affected by Mediator kinase inhibition. These mechanism of cell proliferation defects need to be explored more in mouse models.

Although we and others have shown the effectiveness of Mediator kinase inhibition in colon cancer cell proliferation, it is still unknown whether the effects of Mediator kinase inhibition are generalizable across cancer types. Recent studies on AML have shown that Mediator kinase inhibition has an even more pronounced effect on cell proliferation. Other cell types did not exhibit cell proliferation defects, however we show here that this is potentially due to differences in cellular

context. Furthermore, preventing cancer metastasis would significantly improve patient outcomes and it is unclear if Mediator kinase inhibition generally prevents metastasis or if these effects are cancer type specific.

We present here a data driven mechanism into the role of Mediator kinases in cell cycle progression during serum response. In contrast to cell studies that inhibit Mediator kinase activity and observe no changes to cell proliferation, we demonstrate here that within the context of serum response, HCT116 exhibit defects in cell cycle progression. Recently, a study found that inhibiting Mediator kinases through a different compound, Senexin B, resulted in decreased tumor metastases of HCT116 cells in mice. Taken together, these studies point to the potential effectiveness of Mediator kinase inhibition in the treatment of colon cancer. We note that this was not being explored due to the lack of an observed effect of Mediator kinase inhibition in cell culture models. However, the addition of cellular perturbations such as serum response potentially are better mimics of the cellular environment within an organism. Therefore, adding cellular perturbations to small molecule screens may increase the success of small molecule discovery that lead to positive outcomes outside of the cell culture model.

5.5 Methods

5.5.1 Cell Culture and Treatment

HCT116 Cells STR profiling was performed on HCT116 cells obtained from the Espinosa Lab. Cells were cultured in DMEM with 1x Penn/Strep and 10% FBS. Cells were split at least 1:2 every day to prevent overconfluence. For all experiments, cells were kept to passages below 20.

MCF10A MCF10A expressing DHB-mCherry, H2B-mTurquoise, and mCitrine-p21 were generated and maintained as described previously[234]. Briefly, cells were grown at 37°C, 5% CO2, in DMEM/F12 media supplemented with 5% horse serum, 100ng/mL cholera toxin, 20ng/mL EGF, 10μg/mL insulin, 0.5μg/mL hydrocortisone, and 100μg/mL of penicillin and streptomycin. For serum starvation media, 0.3% BSA was added to DMEM/F12 instead of serum and growth

factors.

Cell Treatments For all treatments, cells received 100nM of cortistatin A (CA) from a 10,000X stock solution. Control cells received the same volume of DMSO per volume of media (1uL DMSO/10mL media). Treatments during starvation (pre-incubation) were administered by directly pipetting into the cell media. Treatments at the time of serum response were administered by first adding CA to pre-warmed serum containing media followed by the addition of this media to cells.

Serum Starvation Before starvation, cells were washed three times with 1x PBS. Total starvation time for HCT116 cells was 40hrs, 48hrs for MCF10A cells. Up

5.5.2 Nascent Sequencing

Nuclei Isolation Nuclei isolation was performed as described in [18] **GRO-Seq run on, enrichment, and library preparation** Nuclei were incubated at 37°C with BrUTP, CTP, ATP, and GTP for 5min. RNA was then isolated and nascent RNA enriched using anti-BrdU beads from Santa Cruz () two times. From purified RNA, libraries were prepared using NEB Next Ultradirectional kit ().

PRO-Seq run on, enrichment, and library preparation Nuclei were incubated at 37°C with Biotin-CTP, ATP, GTP, UTP for 3 min. RNA was then isolated and nascent RNA enriched using magnetic streptavidin beads from XXX three times. Purified nascent RNA was then prepared into a sequencing library using NEB Next Ultradirectional II kit.

Read QC and Mapping Read QC and mapping were performed using the Nascent-flow pipeline v1.1 (https://github.com/Dowell-Lab/Nascent-Flow). Bam files from different sequencing runs were first checked for quality using read duplication statistics and complexity from rseqc and preseq and visualization in IGV. Reads from samples from different lanes and sequencing runs that were of good quality were then concatenated and the Nascent-Flow pipeline v1.1 re-run.

Computational Analysis Differential gene transcription was performed using the DE-Seq2 package with batch correction provided as a separate column in the design matrix. GSEA

was performed using the PreRanked option with genes ranked on fold change. TFEA (v1.1.4) was run on samples specifying the same batches and default parameters.

5.5.3 RNA-Seq

Sample Preparation 10 million cells were seeded on 15cm plates the day before the experiment. Cells were then starved for a total of 40hrs with a pre-incubation with CA or DMSO at 39hrs. Three replicates were collected at 40hrs starvation for cells pre-incubated with DMSO using trizol extraction. Cells were then treated with pre-warmed serum-containing media with an additional dose of either CA or DMSO. After 6hrs of serum response, cells were collected using a trizol extraction. RNA from trizol extracts was purified using a chloroform extraction step followed by DNAse treatment and a P30 column to remove free nucleotides and buffer exchange. RNA was then run on a tapestation to verify RIN numbers and measure concentration. As input into library preparation, we normalized all RNA to contain 1ug of input. We next added 2uL of 1:100 diluted ERCC spike-in as per manufacturer's protocol (). Libraries were then prepared from samples following the Poly-A enrichment kit from NEB () and the NEB Ultradirectional II kit with some modifications to select for larger RNA fragments.

Read QC and mapping Raw fastq files were processed using the nf-core rnaseq package v1.4.2 (https://github.com/nf-core/rnaseq) **Computational Analysis** Gene expression analysis was performed with DE-Seq2 as described above. GO enrichment analysis was performed in Python 3 using the Goatools package v.1.0.3.

5.5.4 Live Cell Imaging

Time-lapse Imaging For imaging, cells were plated onto a collagen-coated 96-well glass-bottom plate (Cellvis Cat. No. P96-1.5H-N). At least 24h after plating, cells were washed twice with serum starvation media, and then incubated in serum starvation media. Cells were then maintained in this media for 39h, at which point the media was supplemented with either DMSO or Cortistatin A at a final concentration of 100nM for a 1h incubation. Cells were then released

into phenol-red free full growth media supplemented with DMSO or 100nM Cortistatin A and imaged for at least 24h on a Nikon TiE in a humidified, temperature controlled incubator (37°C, 5% CO2). Total exposure times for all channels (CFP, YFP, mCherry) were limited to 350ms to limit phototoxicity.

Computational Analysis Time-lapse microscopy images were processed as described previously (doi: 10.1016/j.cell.2016.05.077). The code used for tracking can be found at: https://github.com/scappell/ Custom Python3 scripts were used to generate boxplots, line plots, and heatmaps. Cell traces were clustered based on CDK2 activity using Ward method and a maximum cluster size of 3. Differences in the percentage of cells within each cluster were calculated using a t-test assuming independent samples from the individual wells present in each experiment.

5.5.5 Flow Cytometry

Propidium Iodide 1 million cells were plated in 10 cm plates the day before starting the experiment. Cells were starved and pre-incubated as previously described. At 6, 12, 18, and 24hrs, cells were harvested using trypsing and fixed in 60% ethanol. Fixed cells were then stained with propidium iodide following manufacturer's protocol using the XXXXX kit (). Cell sorting was then performed on a BD Biosciences Celeste machine using the BD software.

EdU Staining Cells were treated as described in the previous section. Three replicates were collected at 12hrs after serum addition. EdU staining was performed according to manufacturer's protocol using the abcam kit (iFluor 488). Cells were sorted using the same instrument as the section above.

Computational Analysis Raw .fcs files were imported using the python package FlowCytometryTools v0.5.0. Raw PI data was transformed using hlog with parameters b=100000. Gating for the PI experiment was done by eye, choosing cutoffs that matched two peaks (G1 and G2, 1000 and 3000 respectively). Raw data from the EdU experiment was transformed using tlog with default parameters and gating was performed by plotting a scatter using FSC-A and FSC-H channels to select for single cells.

5.5.6 Metabolomics

The following methods were obtained through Metabolon. **Sample Accessioning** Following receipt, samples were inventoried and immediately stored at -80oC. Each sample received was accessioned into the Metabolon LIMS system and was assigned by the LIMS a unique identifier that was associated with the original source identifier only. This identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80oC until processed.

Sample Preparation Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

QA/QC Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chro-

matographic alignment. Tables 1 and 2 describe these QC samples and standards. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections, as outlined in Figure 1.

Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS) All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 μm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MSn scans using dynamic

exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

Bioinformatics The informatics system consisted of four major components, the Laboratory Information Management System (LIMS), the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by data analysts. The hardware and software foundations for these informatics components were the LAN backbone, and a database server running Oracle 10.2.0.1 Enterprise Edition. **LIMS** The purpose of the Metabolon LIMS system was to enable fully auditable laboratory automation through a secure, easy to use, and highly specialized system. The scope of the Metabolon LIMS system encompasses sample accessioning, sample preparation and instrumental analysis and reporting and advanced data analysis. All of the subsequent software systems are grounded in the LIMS data structures. It has been modified to leverage and interface with the in-house information extraction and data visualization systems, as well as third party instrumentation and data analysis software.

Data Extraction and Compound Identification Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. These systems are built on a web-service platform utilizing Microsoft's .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors,

the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

Curation A variety of curation procedures were carried out to ensure that a high quality data set was made available for statistical analysis and data interpretation. The QC and curation processes were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, mis-assignments, and background noise. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

Metabolite Quantification and Data Normalization Peaks were quantified using area-under-the-curve. For studies spanning multiple days, a data normalization step was performed to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected in run-day blocks by registering the medians to equal one (1.00) and normalizing each data point proportionately (termed the "block correction"; Figure 2). For studies that did not require more than one day of analysis, no normalization is necessary, other than for purposes of data visualization. In certain instances, biochemical data may have been normalized to an additional factor (e.g., cell counts, total protein as determined by Bradford assay, osmolality, etc.) to account for differences in metabolite levels due to differences in the amount of material present in each sample.

Computational Analysis Additional computational analysis was performed in Python3 using networkx v2.4 in order to visualize the abundances or differences in abundances of metabolites. Ingenuity pathway analysis (IPA) analysis was performed manually using version 51963813.

5.5.7 Phosphoproteomics

Computational Analysis Post-processing of data was performed in R using the limma package lmFit and eBayes methods with Benjamini Hochberg adjusted p-value calculation.

5.6 Future Work

This manuscript describes findings related to the effects of Mediator kinase activity on cell proliferation. One major aspect that is not discussed in this work is the role of Mediator kinase activity in general transcription during serum response. While significant analyses has been accomplished towards answering this question, these results are still in a preliminary stage and not ready for publication. Preliminary analysis suggest that there is little to no effect on pausing via pause index calculation and elongation rate using groHMM. Additionally, splicing analysis on the RNA-Seq dataset has been performed but these results are also in a preliminary stage. These figures are not included in this thesis because of remaining questions about the validity of these analyses.

In addition to a thorough probe into the effects of Mediator kinase activity on transcription, this study lacks mechanistic validation through experimental approaches. Which of the identified phosphorylation sites are most important for the observed effect of Mediator kinases? Questions like this, while would be invaluable to address likely do not have straightforward answers. Validation of one or several sites would add a tremendous amount to this work yet it is entirely possible and more likely that a combination of phosphorylation events lead to the observed effects of Mediator kinase inhibition on cell proliferation.

Finally, as the goal of these basic research projects are to address issues in human health, further exploration into whether CA (or Mediator kinase inhibition generally) are good therapeutic strategies to address cancer cell proliferation or metastasis. This study provides good evidence to support studies in animal models. While normally growing HCT116 cells do not respond to CA treatment, within the context of serum response, we show that they do. Serum response is

possible more representative of a growing tumor microenvironment as these often become deprived of nutrients as they rapidly grow and need to generate blood vessels via angiogenesis to extract more nutrients from the body.

Chapter 6

Conclusions

6.1 Overview

Throughout this thesis, I have presented the results of several projects investigating transcription factor (TF) activity and Mediator kinase activity. In the first two chapters, computational methods were designed to quantify TF activity in a high-throughput manner in sets of perturbation data. These methods were then applied to subsequent studies presented in the next two chapters. Overall, this thesis dealt with investigating the mechanisms of gene regulation at the transcriptional stage.

6.2 MD-Score

The motif displacement score (MD-Score) is a metric for quantifying the activity of TFs within nascent sequencing data. The method works by first identifying sites of bidirectional transcription which are putative sites of RNA Polymerase II (RNAPII) loading and initiation. Therefore these are sites of active RNAPII recruitment presumably accomplished by functional binding of a TF. By measuring the co-localization of bidirectional sites with TF motifs, we infer the association of that TF with functional binding events. Comparing this MD-Score across perturbations, we obtain the differential activation of TFs following treatment.

The caveats with this technique are that it relies on the binary presence or absence of bidirectional sites and cannot take into account bidirectional sites that change in transcription levels. Additionally, the MD-Score approach has no inherent way to handle multiple replicates. Finally,

the MD-Score approach uses a fixed cutoff to determine whether a TF motif is close enough to a bidirectional site that it is considered co-localized. These caveats are addressed somewhat in the TFEA implementation of this approach.

In addition to the development of the MD-Score, this project also involved validation of bidirectional functionality. Several key analyses were performed that increase the validity of using bidirectionals sites as readouts of functional TF binding. One such analysis determined that the enhancer mark H3K27ac was more associated with bidirectionals than other histone modifications. More importantly, we found that the transcriptional level of eRNAs correlated positively with the transcription of the nearest gene. Although it is not always the case that the nearest gene to an enhancer is it's target, the assumption is that on average this is the case. This analysis was done by comparing the transcription of eRNA sites and target gene transcription across cell types. [look at paper again to verify how this was done].

6.3 TFEA

Transcription factor enrichment analysis (TFEA) was meant to improve upon the foundation that was built by the MD-Score approach. TFEA sought to address the main concerns regarding the MD-Score method including the detection of differences in bidirectional transcription levels and the incorporation of replicates into the analysis pipeline. Additionally, TFEA removed the need for fixed threshold cutoffs to determine motif proximity by incorporating an exponential decay function scaled by the background distribution of motifs.

TFEA works on two datasets - a treatment and a control. First, TFEA generates a list of consensus regions of interest (ROI) if these are directly measured from the input datasets. Instead of performing the classic region merge or intersect techniques, we developed a probabilistic merging strategy called *muMerge* that assumes the underlying distribution of reads over these regions is normal. Once a consensus list of ROI are generated, TFEA then ranks regions based on the p-value of differential read coverage generated by the DE-Seq R package. Next, for each TF motif, TFEA first scans ROIs for motif hits, then calculates an enrichment curve that traverses the ranked

list of ROIs increasing in value depending on the distance of the motif hit to the center of the ROI. The motif enrichment is then quantified as the area between the calculated enrichment curve and the expected background curve assuming a uniform distribution of motif hits (the diagonal). Finally, to calculate statistical significance, the ROI rank are shuffled and enrichment recalculated 1000 times to obtain an empirical background distribution of enrichment scores. The observed enrichment score is then compared to this background distribution to obtain a p-value.

We rigorously benchmarked the performance of TFEA against comparable methods including the MD-Score approach, the MDD-Score (a variant of the MD-Score approach), and AME. We found that TFEA was better at separating true TF activation signal from background noise in nascent datasets. Using simulated data, we showed that TFEA can utilize motif positional information to improve its sensitivity, however in the absence of that signal, TFEA performs on par with non-positional enrichment methods (AME).

Once we determined that TFEA performance rivaled or improved upon existing techniques, we then sought to test it on a variety of datasets. The MD-Score approach relied heavily on the accuracy obtained by detecting bidirectionals in nascent data however TFEA was able to detect activation of known TFs following perturbation in CAGE-Seq, EP300 ChIP-Seq, H3K27ac ChIP-Seq, and DNA accessibility assays (DNAse-Seq and ATAC-Seq). Further, we tested TFEA on time series datasets and were able to temporally unravel complex regulatory networks in response to LPS and dexamethasone treatment. We observed rapid (5min) activation of the glucocorticoid receptor after treatment with dexamethasone, the fastest reported activation of a TF following a perturbation [is this true?].

TFEA was designed to be a convenient toolbox to perform TF motif enrichment analysis. Within the software package are implementations to run the MD-Score and MDD-Score alongside TFEA. To facilitate qualitative assessment of the data, interactive HTML results can be generated for each TF motif analyzed. TFEA in itself is a hypothesis-generating tool and can assist researchers in identifying potential targets to pursue within given cellular perturbations or model systems.

6.4 Mediator kinases within infection signaling

The development of computational approaches for inferring TF activity is directly applicable to the study of Mediator kinases (CDK8/CDK19). Mediator is a large protein complex that is essential for regulated gene transcription by RNAPII and works by integrating distal TF binding events to the general transcription machinery. Mediator kinases can reversibly associate with Mediator playing roles in general transcription but have also been shown to be key regulators of TFs in several cellular contexts.

The IFN-γ response is one such context with the main downstream effector, STAT1, being one of the well validated targets of Mediator kinase activity both *in vitro* and in cells. Using small molecule inhibitors of Mediator kinase activity coupled with genetic techniques introducing analog sensitive mutants into cells, this study showed that the IFN-γ response depended on CDK8 activity and CDK19 protein.

IFN-γ treatment for 6hrs results in the activation of around 221 genes in mouse embryonic fibroblasts (MEFs). Concurrent treatment of cells with IFN-γ and CA resulted in decreased activation of 38 of these IFN genes (with 7 more strongly induced). Although these effects are in part due to S727 phosphorylation of STAT1, S727A mutants showed similar effects of CA treatment during the IFN-γ response. Since Mediator kinases are also known to target general transcription factors, nascent sequencing techniques were employed to determine whether Mediator kinase inhibition had effects on any particular stage of gene transcription during the IFN-γ response. After 30min of IFN-γ treatment, genes that exhibited increase in transcription showed increased pausing levels in cells treated with CA. To determine whether this was due to the activity of CDK8, and whether these effects would translate to humans, these experiments were replicated in human HCT116 cells with an analog sensitive CDK8 (asCDK8). Similar effects on pausing was observed in this cell line when treating cells with the analog 3MB to inhibit asCDK8 activity.

The nascent data was also used to probe the effects of Mediator kinase activity on bidirectional transcription through the use of the MD-Score with some modifications (see methods for details).

While the cannonical effectors of the IFN-γ response (STAT and Irf family of TFs) were reliably induced following IFN-γ treatment, addition of CA (in MEFs) or 3MB (in asCDK8 HCT116 cells) resulted in decreased activation of these same TFs. This demonstrated that Mediator kinase activity (specifically CDK8 in HCT116 cells) are responsible for the activation of upstream regulators involved in the transcriptional response to IFN-γ.

To parse out the relative structural and enzymatic contributions of CDK8 and CDK19 in this response, an array of RNA-Seq experiments were performed with inducible knockouts (CDK8) and/or knockdowns (CDK19) and/or CA treatment in addition to a 3hr IFN-γ treatment. As further proof of its specificity, CA had no effect in the absence of both CDK8 and CDK19. The CA-specific effects observed in the presence of CDK8 and CDK19 however, could be reproduced with knockdown of CDK19 indicating that the activity of CDK8 was responsible for these effects. These results were further replicated (to a lesser extent) with CDK8 inducible knockout. Surprisingly, effects of CDK19 knockdown were observed that were different than the effects observed by CDK8 inhibition or knockdown. To determine whether the CDK19 played an important role in the anti-viral response, infection assays were performed in which cells are first pre-treated with IFN-γ then exposed to viral infection. Upon knockdown of CDK19, significantly less cells survived infection.

This study made significant contributions to the Mediator kinase field showing that CDK8 inhibition directly affected the activity of IFN regulators and that CDK19 also plays important roles in this cellular response. Although mechanistically, it is still unclear how Mediator kinases exert their regulatory roles, this study showed that these roles have a direct impact on gene transcription and cell fate. Furthermore, it rigorously tested the enzymatic and structural roles of Mediator kinases showing that each could affect gene regulation in different ways. Ultimately, this study is further evidence that Mediator kinases are important in gene regulation and that the kinase activity plays important roles in this process.

6.5 Mediator kinases within cell proliferative signaling

Since Mediator kinases were shown to be involved in cellular signaling and overexpressed in some cancers, we wondered whether Mediator kinase activity was involved in cell proliferative signaling. To test this, we performed a multitude of -omics experiments looking at the effects of Mediator kinase inhibition within the cellular response to serum. Serum response is a model system to study cellular proliferation and involves a highly complex cellular response as cells deal with the stress of going from a starved state to one of rapid proliferation. Mediator kinases were shown to be involved in the serum response in one study, however this study used CDK8 knockdown which, apart from the many caveats of knockdowns, did not assay kinase activity *per se*.

Using nascent sequencing, we probed for the changes to gene transcription over time after serum and during Mediator kinase inhibition. We found that the response to serum was a highly coordinated event resulting in the temporal activation and inhibition of pathways involved in transcription, translation, and metabolism. When treated with the Mediator kinase inhibitor CA, cells activated several of these pathways at earlier timepoints relative to control cells. These results suggested that Mediator kinases were in part responsible for the appropriate temporal regulation of gene transcription. To confirm this result, TFEA performed on this dataset showed premature activation of JUN and FOSL2 - components of the AP-1 complex of TFs involved in cell proliferation. Phosphoproteomics analysis confirmed that Mediator kinases were directly involved in signaling by Rho GTPases and MAPK. JUN was also detected as a potential direct target of Mediator kinases.

At later timepoints, we measured gene expression changes and cellular metabolism. We found that in response to serum alone, cells had sustained expression of genes involved in RNA process and DNA replication suggesting they might be preparing for transition into S-phase. Cells treated with CA on the other hand showed differentially lower expression of genes related to cell division. Cellular metabolism revealed that cells treated with CA also exhibited lower flux through the TCA cycle, a key pathway involved in energy production and cell growth.

To test the impact of Mediator kinase inhibition on cellular proliferation during serum re-

sponse, we performed propidium iodide (PI) staining, EdU incorporation assays, and live cell imaging. Across all experiments (and two cell types), we found that cells treated with CA showed a lag in cell cycle progression upon serum induction. We observed these effects up to 24hrs after serum response, however long term proliferation assays are still required. Overall, this study confirmed that Mediator kinase activity is directly involved in cell signaling related to proliferation which has consequences in both gene transcription/expression and cellular metabolism.

6.6 Concluding Remarks

The work presented in this thesis makes significant strides towards understand TF activity and Mediator kinase activity. Both of these proteins are directly involved in transcription which is a main regulatory point of gene expression. I first developed computational methods to detect TF activity. As a consequence, I helped validate that enhancer RNAs in the form of bidirectional transcripts are functional readouts of TF activity. Furthermore, TF activity signal is present in histone marks, co-factor binding such as EP300, and chromatin accessibility. I developed a reliable method that can be widely used by the scientific community.

In addition to my work in inferring the activity of sequence-specific TFs, the investigation of Mediator kinases represents work on general TFs. In the last two chapters of this thesis, I presented work detailing the involvement of Mediator kinases in the interferon response and cell proliferative signaling. These are two key pathways involved in a myriad of cellular processes. Ultimately, these results suggest that Mediator kinases are essential in cell growth and survival and targeting their kinase activity could be a valid therapeutic strategy for cancer treatment.

Bibliography

[1] 1.4. Support Vector Machines — scikit-learn 0.17.1 documentation.

[2] Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

[3] BBDuk Guide.

[4] BBMap.

[5] CortistatinA is a High-Affinity Ligand of Protein Kinases ROCK, CDK8, and CDK11 - Cee - 2009 - Angewandte Chemie International Edition - Wiley Online Library.

[6] An introduction to seaborn — seaborn 0.9.0 documentation.

[7] limma powers differential expression analyses for RNA-sequencing and microarray studies | Nucleic Acids Research | Oxford Academic.

[8] Tarek Abbas and Anindya Dutta. p21 in cancer: intricate networks and multiple activities. Nature reviews. Cancer, 9(6):400–414, June 2009.

[9] Ahmed Abdelbaset-Ismail, Daniel Pedziwiatr, Ewa Suszyńska, Sylwia Sluczanowska-Glabowska, Gabriela Schneider, Sham S. Kakar, and Mariusz Z. Ratajczak. Vitamin D3 stimulates embryonic stem cells but inhibits migration and growth of ovarian cancer and teratocarcinoma cell lines. Journal of Ovarian Research, 9:26, 2016.

[10] Richard A Adams, Angela M Meade, Matthew T Seymour, Richard H Wilson, Ayman Madi, David Fisher, Sarah L Kenny, Edward Kay, Elizabeth Hodgkinson, Malcolm Pope, Penny Rogers, Harpreet Wasan, Stephen Falk, Simon Gollins, Tamas Hickish, Eric M Bessell, David Propper, M John Kennedy, Richard Kaplan, and Timothy S Maughan. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. The Lancet Oncology, 12(7):642–653, June 2011.

[11] Karen Adelman and John T. Lis. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews. Genetics, 13(10):720–731, October 2012.

[12] Adam S. Adler, Mark L. McCleland, Tom Truong, Shari Lau, Zora Modrusan, Tim M. Soukup, Merone Roose-Girma, Elizabeth M. Blackwood, and Ron Firestein. CDK8 Maintains Tumor Dedifferentiation and Embryonic Stem Cell Pluripotency. Cancer Research, 72(8):2129–2139, April 2012.

[13] Sasha Akoulitchev, Sergei Chuikov, and Danny Reinberg. TFIIH is negatively regulated by cdk8-containing mediator complexes. Nature, 407(6800):102–106, September 2000.

[14] Claudio Alarcón, Alexia-Ileana Zaromytidou, Qiaoran Xi, Sheng Gao, Jianzhong Yu, Sho Fujisawa, Afsar Barlas, Alexandria N. Miller, Katia Manova-Todorova, Maria J. Macias, Gopal Sapkota, Duojia Pan, and Joan Massagué. CDK8/9 drive Smad transcriptional action, turnover and YAP interactions in BMP and TGF pathways. Cell, 139(4):757–769, November 2009.

[15] B. L. Allen, K. Quach, C. B. Levandowski, J. D. Rubin, T. Read, R. D. Dowell, A. Schepartz, and D. J. Taatjes. Suppression of p53 response by targeting p53-Mediator binding with a stapled peptide. bioRxiv, page 757401, September 2019.

[16] Benjamin L. Allen and Dylan J. Taatjes. The Mediator complex: a central integrator of transcription. Nature Reviews Molecular Cell Biology, 16(3):155–166, March 2015. Number: 3 Publisher: Nature Publishing Group.

[17] Mary A. Allen and Robin Dowell. Improved Methods for GRO-Seq. In Preparation.

[18] Mary Ann Allen, Zdenek Andrysik, Veronica L Dengler, Hestia S Mellert, Anna Guarnieri, Justin A Freeman, Kelly D Sullivan, Matthew D Galbraith, Xin Luo, W Lee Kraus, Robin D Dowell, and Joaquin M Espinosa. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. eLife, 3:e02200, May 2014.

[19] Mary Ann Allen, Zdenek Andrysik, Veronica L. Dengler, Hestia S. Mellert, Anna Guarnieri, Justin A. Freeman, Kelly D. Sullivan, Matthew D. Galbraith, Xin Luo, W. Lee Kraus, Robin D. Dowell, and Joaquin M. Espinosa. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. eLife, 3:e02200, May 2014.

[20] Mehdi Amirhosseini, Magnus Bernhardsson, Pernilla Lång, Göran Andersson, Johan Flygare, and Anna Fahlgren. Cyclin-dependent kinase 8/19 inhibition suppresses osteoclastogenesis by downregulating RANK and promotes osteoblast mineralization and cancellous bone healing. Journal of Cellular Physiology, 0(0), February 2019.

[21] Ido Amit, Ami Citri, Tal Shay, Yiling Lu, Menachem Katz, Fan Zhang, Gabi Tarcic, Doris Siwak, John Lahad, Jasmine Jacob-Hirsch, Ninette Amariglio, Nora Vaisman, Eran Segal, Gideon Rechavi, Uri Alon, Gordon B. Mills, Eytan Domany, and Yosef Yarden. A module of negative feedback regulators defines growth factor signaling. Nature Genetics, 39(4):503–512, April 2007.

[22] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. Genome Biology, 11(10):R106, October 2010.

[23] Jean-Christophe Andrau, Loes van de Pasch, Philip Lijnzaad, Theo Bijma, Marian Groot Koerkamp, Jeroen van de Peppel, Michel Werner, and Frank C. P. Holstege. Genome-Wide Location of the Coactivator Mediator: Binding without Activation and Transient Cdk8 Interaction on DNA. Molecular Cell, 22(2):179–192, April 2006.

[24] Zdenek Andrysik, Matthew D. Galbraith, Anna L. Guarnieri, Sara Zaccara, Kelly D. Sullivan, Ahwan Pandey, Morgan MacBeth, Alberto Inga, and Joaquín M. Espinosa. Identification

of a core TP53 transcriptional program with highly distributed tumor suppressive activity. Genome Research, September 2017.

[25] P. Angel and M. Karin. The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. Biochimica Et Biophysica Acta, 1072(2-3):129–157, December 1991.

[26] Erik Aronesty. Comparison of Sequencing Utility Programs. The Open Bioinformatics Journal, 7(1), January 2013.

[27] K. Audrey Audetat, Matthew D. Galbraith, Aaron T. Odell, Thomas Lee, Ahwan Pandey, Joaquin M. Espinosa, Robin D. Dowell, and Dylan J. Taatjes. A Kinase-Independent Role for Cyclin-Dependent Kinase 19 in p53 Response. Molecular and Cellular Biology, 37(13), July 2017.

[28] Joseph Azofeifa, Mary A. Allen, Josephina Hendrix, Timothy J. Read, Jonathan D. Rubin, and Robin D. Dowell. Enhancer RNA Expression Predicts Transcription Factor Activity. In revision. Genome Research.

[29] Joseph G. Azofeifa, Mary A. Allen, Josephina R. Hendrix, Timothy Read, Jonathan D. Rubin, and Robin D. Dowell. Enhancer RNA profiling predicts transcription factor activity. Genome Research, February 2018.

[30] Joseph G. Azofeifa, Mary A. Allen, Manuel E. Lladser, and Robin D. Dowell. An Annotation Agnostic Algorithm for Detecting Nascent RNA Transcripts in GRO-Seq. IEEE/ACM transactions on computational biology and bioinformatics, 14(5):1070–1081, October 2017.

[31] Joseph G. Azofeifa and Robin D. Dowell. A generative model for the behavior of RNA polymerase. Bioinformatics (Oxford, England), 33(2):227–234, January 2017.

[32] Timothy L. Bailey, Mikael Boden, Fabian A. Buske, Martin Frith, Charles E. Grant, Luca Clementi, Jingyuan Ren, Wilfred W. Li, and William S. Noble. MEME Suite: tools for motif discovery and searching. Nucleic Acids Research, 37(suppl 2):W202–W208, July 2009.

[33] Timothy L. Bailey and Philip Machanick. Inferring direct DNA binding from ChIP-seq. Nucleic Acids Research, 40(17):e128–e128, September 2012.

[34] J. Kenneth Baillie, Erik Arner, Carsten Daub, Michiel De Hoon, Masayoshi Itoh, Hideya Kawaji, Timo Lassmann, Piero Carninci, Alistair R. R. Forrest, Yoshihide Hayashizaki, Geoffrey J. Faulkner, Christine A. Wells, Michael Rehli, Paul Pavli, Kim M. Summers, and David A. Hume. Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. PLoS Genetics, 13(3), March 2017.

[35] Nathalie Q. Balaban, Jack Merrin, Remy Chait, Lukasz Kowalik, and Stanislas Leibler. Bacterial Persistence as a Phenotypic Switch. Science, 305(5690):1622–1625, September 2004.

[36] Piotr J. Balwierz, Mikhail Pachkov, Phil Arnold, Andreas J. Gruber, Mihaela Zavolan, and Erik van Nimwegen. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. Genome Research, 24(5):869–884, May 2014.

[37] Laura A. Banaszynski, Ling-Chun Chen, Lystranne A. Maynard-Smith, A. G. Lisa Ooi, and Thomas J. Wandless. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell, 126(5):995–1004, September 2006.

[38] Joanna Bancerek, Zachary C. Poss, Iris Steinparzer, Vitaly Sedlyarov, Thaddäus Pfaffenwimmer, Ivana Mikulic, Lars Dölken, Birgit Strobl, Mathias Müller, Dylan J. Taatjes, and Pavel Kovarik. CDK8 Kinase Phosphorylates Transcription Factor STAT1 to Selectively Regulate the Interferon Response. Immunity, 38(2):250–262, February 2013.

[39] Andrew J. Bannister and Tony Kouzarides. Regulation of chromatin by histone modifications. Cell Research, 21(3):381–395, March 2011.

[40] Corine M. Beaufort, Jean C. A. Helmijr, Anna M. Piskorz, Marlous Hoogstraat, Kirsten Ruigrok-Ritstier, Nicolle Besselink, Muhammed Murtaza, Wilfred F. J. van IJcken, Anouk A. J. Heine, Marcel Smid, Marco J. Koudijs, James D. Brenton, Els M. J. J. Berns, and Jozien Helleman. Ovarian Cancer Cell Line Panel (OCCP): Clinical Importance of In Vitro Morphological Subtypes. PLoS ONE, 9(9), September 2014.

[41] Tim Beck, Robert K. Hastings, Sirisha Gollapudi, Robert C. Free, and Anthony J. Brookes. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. European Journal of Human Genetics, 22(7):949–952, July 2014.

[42] Finn Becker, Vincent Joerg, Marie C. Hupe, Doris Roth, Rosemarie Krupar, Verena Lubczyk, Rainer Kuefer, Verena Sailer, Stefan Duensing, Jutta Kirfel, Axel S. Merseburger, Johannes Brägelmann, Sven Perner, and Anne Offermann. Increased Mediator complex subunit CDK19 expression associates with aggressive prostate cancer. International Journal of Cancer, July 2019.

[43] L. Becker, T. Mito, S. Takashima, and K. Onodera. Growth and development of the brain in Down syndrome. Progress in Clinical and Biological Research, 373:133–152, 1991.

[44] Afshin Beheshti, Egle Cekanaviciute, David J. Smith, and Sylvain V. Costes. Global transcriptomic analysis suggests carbon dioxide as an environmental stressor in spaceflight: A systems biology GeneLab case study. Scientific Reports, 8(1):4191, March 2018.

[45] Ivan Berest, Christian Arnold, Armando Reyes-Palomares, Giovanni Palla, Kasper Dindler Rasmussen, Kristian Helin, and Judith B. Zaugg. Quantification of differential transcription factor activity and multiomics-based classification into activators and repressors: diffTF. bioRxiv, page 368498, December 2018.

[46] Helen K Berrieman, Michael J Lind, and Lynn Cawkwell. Do -tubulin mutations have a role in resistance to chemotherapy? The Lancet Oncology, 5(3):158–164, March 2004.

[47] Anand S. Bhagwat, Jae-Seok Roe, Beverly Y. L. Mok, Anja F. Hohmann, Junwei Shi, and Christopher R. Vakoc. BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. Cell Reports, 15(3):519–530, April 2016.

[48] JosephW. Bigger. TREATMENT OF STAPHYLOCOCCAL INFECTIONS WITH PENICILLIN BY INTERMITTENT STERILISATION. The Lancet, 244(6320):497–500, October 1944.

[49] Andrea Bild and Phillip George Febbo. Application of a priori established gene sets to discover biologically important differential expression in microarray data. Proceedings of the National Academy of Sciences, 102(43):15278–15279, October 2005.

[50] Charles Blatti and Saurabh Sinha. Motif Enrichment Tool. Nucleic Acids Research, 42(Web Server issue):W20–W25, July 2014.

[51] Sheree L. Boulet, Noelle-Angelique Molinari, Scott D. Grosse, Margaret A. Honein, and Adolfo Correa-Villaseñor. Health care expenditures for infants and young children with Down syndrome in a privately insured population. The Journal of Pediatrics, 153(2):241–246, August 2008.

[52] T. Boulikas. Phosphorylation of transcription factors and control of the cell cycle. Critical Reviews in Eukaryotic Gene Expression, 5(1):1–77, 1995.

[53] Frédérique Braun, Joséphine Bertin-Ciftci, Anne-Sophie Gallouet, Julie Millour, and Philippe Juin. Serum-Nutrient Starvation Induces Cell Death Mediated by Bax and Puma That Is Counteracted by p21 and Unmasked by Bcl-xL Inhibition. PLOS ONE, 6(8):e23577, August 2011.

[54] Marilyn J. Bull and Committee on Genetics. Health supervision for children with Down syndrome. Pediatrics, 128(2):393–406, August 2011.

[55] Brian Bushnell. BBMap: A Fast, Accurate, Splice-Aware Aligner. Technical Report LBNL-7065E, Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), March 2014.

[56] Eliezer Calo and Joanna Wysocka. Modification of enhancer chromatin: what, how and why? Molecular cell, 49(5), March 2013.

[57] Jiwei Cao, Gang Xu, Jing Lan, Qingqing Huang, Zuxiong Tang, and Liping Tian. High expression of piwi-like RNA-mediated gene silencing 1 is associated with poor prognosis via regulating transforming growth factor- receptors and cyclin-dependent kinases in breast cancer. Molecular Medicine Reports, February 2016.

[58] S. Cara and I. F. Tannock. Retreatment of patients with the same chemotherapy: Implications for clinical mechanisms of drug resistance. Annals of Oncology, 12(1):23–27, January 2001.

[59] S. Cara and I. F. Tannock. Retreatment of patients with the same chemotherapy: Implications for clinical mechanisms of drug resistance. Annals of Oncology, 12(1):23–27, January 2001.

[60] Joseph F. Cardiello, James A. Goodrich, and Jennifer F. Kugel. Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. Molecular and Cellular Biology, 38(18), September 2018. Publisher: American Society for Microbiology Journals Section: Research Article.

[61] Marie Cargnello and Philippe P. Roux. Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. Microbiology and Molecular Biology Reviews : MMBR, 75(1):50–83, March 2011.

[62] VictorJ. Cee, DavidY.-K. Chen, MatthewR. Lee, and K. C. Nicolaou. CortistatinA is a High-Affinity Ligand of Protein Kinases ROCK, CDK8, and CDK11. Angewandte Chemie International Edition, 48(47):8952–8957, November 2009.

[63] Center for History and New Media. Zotero Quick Start Guide.

[64] J. Chai and A. S. Tarnawski. Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. Journal of Physiology and Pharmacology: An Official Journal of the Polish Physiological Society, 53(2):147–157, June 2002.

[65] Ganta V. Chaitanya, Jonathan S. Alexander, and Phanithi P. Babu. PARP-1 cleavage fragments: signatures of cell-death proteases in neurodegeneration. Cell Communication and Signaling, 8(1):31, December 2010.

[66] Andrew L. Chang, Yiming Kang, and Tamara L. Doering. Cdk8 and Ssn801 Regulate Oxidative Stress Resistance and Virulence in Cryptococcus neoformans. mBio, 10(1):e02818–18, February 2019.

[67] Mengqian Chen, Jiaxin Liang, Hao Ji, Zhengguan Yang, Serena Altilia, Bing Hu, Adam Schronce, Martina S. J. McDermott, Gary P. Schools, Chang-uk Lim, David Oliver, Michael S. Shtutman, Tao Lu, George R. Stark, Donald C. Porter, Eugenia V. Broude, and Igor B. Roninson. CDK8/19 Mediator kinases potentiate induction of transcription by NFB. Proceedings of the National Academy of Sciences, page 201710467, August 2017.

[68] Bo Cheng, Tiandao Li, Peter B. Rahl, Todd E. Adamson, Nicholas B. Loudas, Jiannan Guo, Katayoun Varzavand, Jeffrey J. Cooper, Xiaopeng Hu, Averell Gnatt, Richard A. Young, and David H. Price. Functional association of Gdown1 with RNA polymerase II poised on human genes. Molecular Cell, 45(1):38–50, January 2012.

[69] Y. Chi, M. J. Huddleston, X. Zhang, R. A. Young, R. S. Annan, S. A. Carr, and R. J. Deshaies. Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. Genes & Development, 15(9):1078–1092, May 2001.

[70] Hyung-lok Chung, Xiao Mao, Hua Wang, Ye-Jin Park, Paul C. Marcogliese, Jill A. Rosenfeld, Lindsay C. Burrage, Pengfei Liu, David R. Murdock, Shinya Yamamoto, Michael F. Wangler, Hsiao-Tuan Chao, Hongyu Long, Li Feng, Carlos A. Bacino, Hugo J. Bellen, and Bo Xiao. De Novo Variants in CDK19 Are Associated with a Syndrome Involving Intellectual Disability and Epileptic Encephalopathy. The American Journal of Human Genetics, April 2020.

[71] Hyung-Lok Chung, Xiao Mao, Hua Wang, Ye-Jin Park, Paul C. Marcogliese, Jill A. Rosenfeld, Lindsay C. Burrage, Pengfei Liu, David R. Murdock, Shinya Yamamoto, Michael F. Wangler, Undiagnosed Diseases Network, Hsiao-Tuan Chao, Hongyu Long, Li Feng, Carlos A. Bacino, Hugo J. Bellen, and Bo Xiao. De Novo Variants in CDK19 Are Associated with a Syndrome Involving Intellectual Disability and Epileptic Encephalopathy. American Journal of Human Genetics, April 2020.

[72] Dean Clift, William A. McEwan, Larisa I. Labzin, Vera Konieczny, Binyam Mogessie, Leo C. James, and Melina Schuh. A Method for the Acute and Rapid Degradation of Endogenous Proteins. Cell, November 2017.

[73] S. Connelly and J. L. Manley. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes & Development, 2(4):440–452, April 1988. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[74] The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414):57–74, September 2012.

[75] Leighton J. Core, Joshua J. Waterfall, and John T. Lis. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (New York, N.Y.), 322(5909):1845–1848, December 2008.

[76] N. Costlow and J. T. Lis. High-resolution mapping of DNase I-hypersensitive sites of Drosophila heat shock genes in Drosophila melanogaster and Saccharomyces cerevisiae. Molecular and Cellular Biology, 4(9):1853–1863, September 1984. Publisher: American Society for Microbiology Journals Section: Research Article.

[77] Alessia Curina, Alberto Termanini, Iros Barozzi, Elena Prosperini, Marta Simonatto, Sara Polletti, Alessio Silvola, Monica Soldi, Liv Austenaa, Tiziana Bonaldi, Serena Ghisletti, and Gioacchino Natoli. High constitutive activity of a broad panel of housekeeping and tissue-specific cis-regulatory elements depends on a subset of ETS proteins. Genes & Development, 31(4):399–412, February 2017.

[78] Timothy Daley and Andrew D. Smith. Predicting the molecular complexity of sequencing libraries. Nature Methods, 10(4):325–327, April 2013.

[79] Chi V. Dang. MYC on the Path to Cancer. Cell, 149(1):22–35, March 2012.

[80] Charles G. Danko, Nasun Hah, Xin Luo, André L. Martins, Leighton Core, John T. Lis, Adam Siepel, and W. Lee Kraus. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Molecular Cell, 50(2):212–222, April 2013.

[81] Charles G. Danko, Stephanie L. Hyland, Leighton J. Core, Andre L. Martins, Colin T Waters, Hyung Won Lee, Vivian G. Cheung, W. Lee Kraus, John T. Lis, and Adam Siepel. Identification of active transcriptional regulatory elements with GRO-seq. Nature methods, 12(5):433–438, May 2015.

[82] Marius Volker Dannappel, Dhanya Sooraj, Jia Jian Loh, and Ron Firestein. Molecular and in vivo Functions of the CDK8 and CDK19 Kinase Modules. Frontiers in Cell and Developmental Biology, 6, January 2019.

[83] Carrie A. Davis, Benjamin C. Hitz, Cricket A. Sloan, Esther T. Chan, Jean M. Davidson, Idan Gabdank, Jason A. Hilton, Kriti Jain, Ulugbek K. Baymuradov, Aditi K. Narayanan, Kathrina C. Onate, Keenan Graham, Stuart R. Miyasato, Timothy R. Dreszer, J. Seth Strattan, Otto Jolanki, Forrest Y. Tanaka, and J. Michael Cherry. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Research, 46(D1):D794–D801, 2018.

[84] Michael A. Davis, Elizabeth A. Larimore, Brian M. Fissel, Jherek Swanger, Dylan J. Taatjes, and Bruce E. Clurman. The SCF–Fbw7 ubiquitin ligase degrades MED13 and MED13L and regulates CDK8 module association with Mediator. Genes & Development, 27(2):151–156, January 2013.

[85] F. De Graeve, A. Bahr, K. T. Sabapathy, C. Hauss, E. F. Wagner, C. Kedinger, and B. Chatton. Role of the ATFa/JNK2 complex in Jun activation. Oncogene, 18(23):3491–3500, June 1999.

[86] Manuel de la Mata, Claudio R. Alonso, Sebastián Kadener, Juan P. Fededa, Matías Blaustein, Federico Pelisch, Paula Cramer, David Bentley, and Alberto R. Kornblihtt. A slow RNA polymerase II affects alternative splicing in vivo. Molecular Cell, 12(2):525–532, August 2003.

[87] Francesca De Santa, Iros Barozzi, Flore Mietton, Serena Ghisletti, Sara Polletti, Betsabeh Khoramian Tusi, Heiko Muller, Jiannis Ragoussis, Chia-Lin Wei, and Gioacchino Natoli. A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. PLoS Biology, 8(5), May 2010.

[88] Marion Dejosez, Stuart S. Levine, Garrett M. Frampton, Warren A. Whyte, Sabrina A. Stratton, Michelle C. Barton, Preethi H. Gunaratne, Richard A. Young, and Thomas P. Zwaka. Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. Genes & Development, 24(14):1479–1484, July 2010. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[89] Russell C. DeKelver, Vivian M. Choi, Erica A. Moehle, David E. Paschon, Dirk Hockemeyer, Sebastiaan H. Meijsing, Yasemin Sancak, Xiaoxia Cui, Eveline J. Steine, Jeffrey C. Miller, Phillip Tam, Victor V. Bartsevich, Xiangdong Meng, Igor Rupniewski, Sunita M. Gopalan, Helena C. Sun, Kathleen J. Pitz, Jeremy M. Rock, Lei Zhang, Gregory D. Davis, Edward J. Rebar, Iain M. Cheeseman, Keith R. Yamamoto, David M. Sabatini, Rudolf Jaenisch, Philip D. Gregory, and Fyodor D. Urnov. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Research, 20(8):1133–1142, August 2010.

[90] Claire Delahunty and John R. Yates III. Protein identification using 2D-LC-MS/MS. Methods, 35(3):248–255, March 2005.

[91] Shala Dezfouli, Antony Bakke, Jie Huang, Anthony Wynshaw-Boris, and Peter J. Hurlin. Inflammatory Disease and Lymphomagenesis Caused by Deletion of the Myc Antagonist Mnt in T Cells. Molecular and Cellular Biology, 26(6):2080–2092, March 2006.

[92] Paolo Di Tommaso, Maria Chatzou, Evan W. Floden, Pablo Prieto Barja, Emilio Palumbo, and Cedric Notredame. Nextflow enables reproducible computational workflows. Nature Biotechnology, 35(4):316–319, April 2017.

[93] Aaron J. Donner, Christopher C. Ebmeier, Dylan J. Taatjes, and Joaquín M. Espinosa. CDK8 is a positive regulator of transcriptional elongation within the serum response network. Nature structural & molecular biology, 17(2):194–201, February 2010.

[94] Aaron Joseph Donner, Stephanie Szostek, Jennifer Michelle Hoover, and Joaquin Maximiliano Espinosa. CDK8 is a stimulus-specific positive coregulator of p53 target genes. Molecular Cell, 27(1):121–133, July 2007.

[95] Gwendal Dujardin, Celina Lafaille, Manuel de la Mata, Luciano E. Marasco, Manuel J. Muñoz, Catherine Le Jossic-Corcos, Laurent Corcos, and Alberto R. Kornblihtt. How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. Molecular Cell, 54(4):683–690, May 2014.

[96] Christopher Dye. After 2015: infectious diseases in a new era of health and development. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1645), June 2014.

[97] Nicholas J. Dyson. RB1: a prototype tumor suppressor and an enigma. Genes & Development, 30(13):1492–1502, July 2016.

[98] Katrin Düvel, Jessica L. Yecies, Suchithra Menon, Pichai Raman, Alex I. Lipovsky, Amanda L. Souza, Ellen Triantafellow, Qicheng Ma, Regina Gorski, Stephen Cleaver, Matthew G. Vander Heiden, Jeffrey P. MacKeigan, Peter M. Finan, Clary B. Clish, Leon O. Murphy, and Brendan D. Manning. Activation of a metabolic gene regulatory network downstream of mTOR complex 1. Molecular Cell, 39(2):171–183, July 2010.

[99] Christopher C. Ebmeier and Dylan J. Taatjes. Activator-Mediator binding regulates Mediator-cofactor interactions. Proceedings of the National Academy of Sciences, 107(25):11283–11288, June 2010.

[100] Hans Elmlund, Vera Baraznenok, Martin Lindahl, Camilla O. Samuelsen, Philip J. B. Koeck, Steen Holmberg, Hans Hebert, and Claes M. Gustafsson. The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America, 103(43):15788–15793, October 2006.

[101] Philip Ewels, Måns Magnusson, Sverker Lundin, and Max Käller. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19):3047–3048, October 2016.

[102] Philip A. Ewels, Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso, and Sven Nahnsen. The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology, 38(3):276–278, March 2020.

[103] Charli Fant, Cecilia Levandowski, K. Gupta, Zachary Maas, John Moir, J. D. Rubin, Andrew Sawyer, Meagan Esbin, Jenna Rimel, Michael Marr, Imre Berger, Robin Dowell, and Dylan J. Taatjes. TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. SSRN Scholarly Paper ID 3429345, Social Science Research Network, Rochester, NY, July 2019.

[104] Charli B. Fant, Cecilia B. Levandowski, Kapil Gupta, Zachary L. Maas, John Moir, Jonathan D. Rubin, Andrew Sawyer, Meagan N. Esbin, Jenna K. Rimel, Olivia Luyties, Michael T. Marr, Imre Berger, Robin D. Dowell, and Dylan J. Taatjes. TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. Molecular Cell, March 2020.

[105] Matthew J. Farrar, Jonathan D. Rubin, Darcy M. Diago, and Chris B. Schaffer. Characterization of blood flow in the mouse dorsal spinal venous system before and after dorsal spinal vein occlusion. Journal of Cerebral Blood Flow & Metabolism, 35(4):667–675, April 2015.

[106] Matthew J. Farrar, Jonathan D. Rubin, Darcy M. Diago, and Chris B. Schaffer. Characterization of blood flow in the mouse dorsal spinal venous system before and after dorsal spinal vein occlusion. Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism, 35(4):667–675, 2015.

[107] Daorong Feng, Dou Yeon Youn, Xiaoping Zhao, Yanguang Gao, William J. Quinn 3rd, Alus M. Xiaoli, Yan Sun, Morris J. Birnbaum, Jeffrey E. Pessin, and Fajun Yang. mTORC1 Down-Regulates Cyclin-Dependent Kinase 8 (CDK8) and Cyclin C (CycC). PLOS ONE, 10(6):e0126240, June 2015.

[108] Maria Feoktistova, Peter Geserick, and Martin Leverkus. Crystal Violet Assay for Determining Viability of Cultured Cells. Cold Spring Harbor Protocols, 2016(4):pdb.prot087379, April 2016.

[109] C. Ferlini, R. Biselli, G. Scambia, and A. Fattorossi. Probing chromatin structure in the early phases of apoptosis. Cell Proliferation, 29(7):427–436, July 1996.

[110] P. Ferré and F. Foufelle. SREBP-1c Transcription Factor and Lipid Homeostasis: Clinical Perspective. Hormone Research in Paediatrics, 68(2):72–82, March 2007.

[111] Ron Firestein, Adam J. Bass, So Young Kim, Ian F. Dunn, Serena J. Silver, Isil Guney, Ellen Freed, Azra H. Ligon, Natalie Vena, Shuji Ogino, Milan G. Chheda, Pablo Tamayo, Stephen Finn, Yashaswi Shrestha, Jesse S. Boehm, Supriya Jain, Emeric Bojarski, Craig Mermel, Jordi Barretina, Jennifer A. Chan, Jose Baselga, Josep Tabernero, David E. Root, Charles S. Fuchs, Massimo Loda, Ramesh A. Shivdasani, Matthew Meyerson, and William C. Hahn. CDK8 is a colorectal cancer oncogene that regulates -catenin activity. Nature, 455(7212):547–551, September 2008.

[112] Ron Firestein, Kaori Shima, Katsuhiko Nosho, Natsumi Irahara, Yoshifumi Baba, Emeric Bojarski, Edward L. Giovannucci, William C. Hahn, Charles S. Fuchs, and Shuji Ogino. CDK8 Expression in 470 Colorectal Cancers in Relation to -Catenin Activation, Other Molecular Alterations and Patient Survival. International journal of cancer. Journal international du cancer, 126(12):2863–2873, June 2010.

[113] Ryan A. Flynn, Albert E. Almada, Jesse R. Zamudio, and Phillip A. Sharp. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proceedings of the National Academy of Sciences, 108(26):10460–10465, June 2011.

[114] Nova Fong, Hyunmin Kim, Yu Zhou, Xiong Ji, Jinsong Qiu, Tassa Saldi, Katrina Diener, Ken Jones, Xiang-Dong Fu, and David L. Bentley. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes & Development, 28(23):2663–2676, December 2014.

[115] Oriol Fornes, Jaime A. Castro-Mondragon, Aziz Khan, Robin van der Lee, Xi Zhang, Phillip A. Richmond, Bhavi P. Modi, Solenne Correard, Marius Gheorghe, Damir Baranašić, Walter Santana-Garcia, Ge Tan, Jeanne Chèneby, Benoit Ballester, François Parcy, Albin Sandelin, Boris Lenhard, Wyeth W. Wasserman, and Anthony Mathelier. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Research, 48(D1):D87–D92, January 2020. Publisher: Oxford Academic.

[116] Alistair R. R. Forrest, Hideya Kawaji, Michael Rehli, J. Kenneth Baillie, Michiel J. L. de Hoon, Vanja Haberle, Timo Lassmann, Ivan V. Kulakovskiy, Marina Lizio, Masayoshi Itoh, Robin Andersson, Christopher J. Mungall, Terrence F. Meehan, Sebastian Schmeier, Nicolas Bertin, Mette Jørgensen, Emmanuel Dimont, Erik Arner, Christian Schmidl, Ulf Schaefer, Yulia A. Medvedeva, Charles Plessy, Morana Vitezic, Jessica Severin, Colin A. Semple, Yuri Ishizu, Robert S. Young, Margherita Francescatto, Intikhab Alam, Davide Albanese, Gabriel M. Altschuler, Takahiro Arakawa, John A. C. Archer, Peter Arner, Magda Babina, Sarah Rennie, Piotr J. Balwierz, Anthony G. Beckhouse, Swati Pradhan-Bhatt, Judith A. Blake, Antje Blumenthal, Beatrice Bodega, Alessandro Bonetti, James Briggs, Frank Brombacher, A. Maxwell Burroughs, Andrea Califano, Carlo V. Cannistraci, Daniel Carbajo, Yun Chen, Marco Chierici, Yari Ciani, Hans C. Clevers, Emiliano Dalla, Carrie A. Davis, Michael Detmar, Alexander D. Diehl, Taeko Dohi, Finn Drabløs, Albert S. B. Edge, Matthias Edinger, Karl Ekwall, Mitsuhiro Endoh, Hideki Enomoto, Michela Fagiolini, Lynsey Fairbairn, Hai Fang, Mary C. Farach-Carson, Geoffrey J. Faulkner, Alexander V. Favorov, Malcolm E. Fisher, Martin C. Frith, Rie Fujita, Shiro Fukuda, Cesare Furlanello, Masaaki Furuno, Jun-ichi Furusawa, Teunis B. Geijtenbeek, Andrew P. Gibson, Thomas Gingeras, Daniel Goldowitz, Julian Gough, Sven Guhl, Reto Guler, Stefano Gustincich, Thomas J. Ha, Masahide Hamaguchi, Mitsuko Hara, Matthias Harbers, Jayson Harshbarger, Akira Hasegawa, Yuki Hasegawa, Takehiro Hashimoto, Meenhard Herlyn, Kelly J. Hitchens, Shannan J. Ho Sui, Oliver M. Hofmann, Ilka Hoof, Fumi Hori, Lukasz Huminiecki, Kei Iida, Tomokatsu Ikawa, Boris R. Jankovic, Hui Jia, Anagha Joshi, Giuseppe Jurman, Bogumil Kaczkowski, Chieko Kai, Kaoru Kaida, Ai Kaiho, Kazuhiro Kajiyama, Mutsumi Kanamori-Katayama, Artem S. Kasianov, Takeya Kasukawa, Shintaro Katayama, Sachi Kato, Shuji Kawaguchi, Hiroshi Kawamoto, Yuki I. Kawamura, Tsugumi Kawashima, Judith S. Kempfle, Tony J. Kenna, Juha Kere, Levon M. Khachigian, Toshio Kitamura, S. Peter Klinken, Alan J. Knox, Miki Kojima, Soichi Kojima, Naoto Kondo, Haruhiko Koseki, Shigeo Koyasu, Sarah Krampitz, Atsutaka Kubosaki, Andrew T. Kwon, Jeroen F. J. Laros, Weonju Lee, Andreas Lennartsson, Kang Li, Berit Lilje, Leonard Lipovich, Alan Mackay-sim, Ri-ichiroh Manabe, Jessica C. Mar, Benoit Marchand, Anthony Mathelier, Niklas Mejhert, Alison Meynert, Yosuke Mizuno, David A. de Lima Morais, Hiromasa Morikawa, Mitsuru Morimoto, Kazuyo Moro, Efthymios Motakis, Hozumi Motohashi, Christine L. Mummery, Mitsuyoshi Murata, Sayaka Nagao-Sato, Yutaka Nakachi, Fumio Nakahara, Toshiyuki Nakamura, Yukio Nakamura, Kenichi Nakazato, Erik van Nimwegen, Noriko Ninomiya, Hiromi Nishiyori, Shohei Noma, Tadasuke Nozaki, Soichi Ogishima, Naganari Ohkura, Hiroko Ohmiya, Hiroshi Ohno, Mitsuhiro Ohshima, Mariko Okada-Hatakeyama, Yasushi Okazaki, Valerio Orlando, Dmitry A. Ovchinnikov, Arnab Pain, Robert Passier, Margaret Patrikakis, Helena Persson, Silvano Piazza, James G. D. Prendergast, Owen J. L. Rackham, Jordan A. Ramilowski, Mamoon Rashid, Timothy Ravasi, Patrizia Rizzu, Marco Roncador, Sugata Roy, Morten B. Rye, Eri Saijyo, Antti Sajantila, Akiko Saka, Shimon Sakaguchi, Mizuho Sakai, Hiroki Sato, Hironori Satoh, Suzana Savvi, Alka Saxena, Claudio Schneider, Erik A. Schultes, Gundula G. Schulze-Tanzil, Anita Schwegmann, Thierry Sengstag, Guojun Sheng, Hisashi Shimoji, Yishai Shimoni, Jay W. Shin, Christophe Simon, Daisuke Sugiyama, Takaaki Sugiyama, Masanori Suzuki, Naoko Suzuki, Rolf K. Swoboda, Peter A. C. 't Hoen, Michihira Tagami, Naoko Takahashi, Jun Takai, Hiroshi Tanaka, Hideki Tatsukawa, Zuotian Tatum, Mark Thompson, Hiroo Toyoda, Tetsuro Toyoda, Eivind Valen, Marc van de Wetering, Linda M. van den Berg, Roberto Verardo, Dipti Vijayan, Ilya E. Vorontsov, Wyeth W. Wasserman, Shoko Watanabe, Christine A. Wells, Louise N. Winteringham, Ernst Wolvetang, Emily J. Wood,

Yoko Yamaguchi, Masayuki Yamamoto, Misako Yoneda, Yohei Yonekura, Shigehiro Yoshida, Susan E. Zabierowski, Peter G. Zhang, Xiaobei Zhao, Silvia Zucchelli, Kim M. Summers, Harukazu Suzuki, Carsten O. Daub, Jun Kawai, Peter Heutink, Winston Hide, Tom C. Freeman, Boris Lenhard, Vladimir B. Bajic, Martin S. Taylor, Vsevolod J. Makeev, Albin Sandelin, David A. Hume, Piero Carninci, Yoshihide Hayashizaki, and The FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. Nature, 507(7493):462–470, March 2014.

[117] Hector L. Franco, Anusha Nagari, and W. Lee Kraus. TNF signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. Molecular Cell, 58(1):21–34, April 2015.

[118] Christy J. Fryer, J. Brandon White, and Katherine A. Jones. Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. Molecular Cell, 16(4):509–520, November 2004.

[119] Christy J. Fryer, J. Brandon White, and Katherine A. Jones. Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. Molecular Cell, 16(4):509–520, November 2004.

[120] Jun Fujimoto, Takaharu Hirayama, Yasuhiro Hirata, Yukiko Hikichi, Saomi Murai, Maki Hasegawa, Yuka Hasegawa, Kazuko Yonemori, Akito Hata, Kazunobu Aoyama, and Douglas R. Cary. Studies of CDK 8/19 inhibitors: Discovery of novel and selective CDK8/19 dual inhibitors and elimination of their CYP3A4 time-dependent inhibition potential. Bioorganic & Medicinal Chemistry, March 2017.

[121] Meital Gabay, Yulin Li, and Dean W. Felsher. MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. Cold Spring Harbor Perspectives in Medicine, 4(6):a014241, June 2014.

[122] Meital Gabay, Yulin Li, and Dean W. Felsher. MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. Cold Spring Harbor Perspectives in Medicine, 4(6):a014241, June 2014.

[123] Matthew D. Galbraith, Mary A. Allen, Claire L. Bensard, Xiaoxing Wang, Marie K. Schwinn, Bo Qin, Henry W. Long, Danette L. Daniels, William C. Hahn, Robin D. Dowell, and Joaquín M. Espinosa. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell, 153(6):1327–1339, June 2013.

[124] Matthew D Galbraith, Aaron J Donner, and Joaquin M Espinosa. CDK8. Transcription, 1(1):4–12, 2010.

[125] Daniel A. Gilchrist, Gilberto Dos Santos, David C. Fargo, Bin Xie, Yuan Gao, Leping Li, and Karen Adelman. Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. Cell, 143(4):540–551, November 2010.

[126] Aaron Goldman, Biswanath Majumder, Andrew Dhawan, Sudharshan Ravi, David Goldman, Mohammad Kohandel, Pradip K. Majumder, and Shiladitya Sengupta. Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition. Nature Communications, 6, February 2015.

[127] D. W. Goodrich. The retinoblastoma tumor-suppressor gene, the exception that proves the rule. Oncogene, 25(38):5233, August 2006.

[128] Charles E. Grant, Timothy L. Bailey, and William Stafford Noble. FIMO: scanning for occurrences of a given motif. Bioinformatics, 27(7):1017–1018, April 2011.

[129] Saskia Gressel, Björn Schwalb, Tim Michael Decker, Weihua Qin, Heinrich Leonhardt, Dirk Eick, and Patrick Cramer. CDK9-dependent RNA polymerase II pausing controls transcription initiation. eLife, 6:e29736, October 2017. Publisher: eLife Sciences Publications, Ltd.

[130] Kapil Gupta, Duygu Sari-Ak, Matthias Haffke, Simon Trowitzsch, and Imre Berger. Zooming in on Transcription Preinitiation. Journal of Molecular Biology, 428(12):2581–2591, June 2016.

[131] S. Gupta, D. Campbell, B. Derijard, and R. J. Davis. Transcription factor ATF2 regulation by the JNK signal transduction pathway. Science, 267(5196):389–393, January 1995.

[132] Michael J. Haas. CDK8 inhibitor: Senex's best thing. SciBX: Science-Business eXchange, 5(33), August 2012.

[133] Nasun Hah, Shino Murakami, Anusha Nagari, Charles G. Danko, and W. Lee Kraus. Enhancer transcripts mark active estrogen receptor binding sites. Genome Research, 23(8):1210–1223, August 2013.

[134] Xingchun Han, Min Jiang, Chengang Zhou, Zheng Zhou, Zhiheng Xu, Lisha Wang, Alexander V. Mayweg, Rui Niu, Tai-Guang Jin, and Song Yang. Discovery of potent and selective CDK8 inhibitors through FBDD approach. Bioorganic & Medicinal Chemistry Letters, August 2017.

[135] Douglas Hanahan and Robert A. Weinberg. Hallmarks of cancer: the next generation. Cell, 144(5):646–674, March 2011.

[136] Alexandria M. Haslehurst, Madhuri Koti, Moyez Dharsee, Paulo Nuin, Ken Evans, Joseph Geraci, Timothy Childs, Jian Chen, Jieran Li, Johanne Weberpals, Scott Davey, Jeremy Squire, Paul C. Park, and Harriet Feilotter. EMT transcription factors snail and slug directly contribute to cisplatin resistance in ovarian cancer. BMC Cancer, 12:91, 2012.

[137] M. Hatakeyama and R. A. Weinberg. The role of RB in cell cycle control. Progress in Cell Cycle Research, 1:9–19, 1995.

[138] Katerina Hatzi, Yanwen Jiang, Chuanxin Huang, Francine Garrett-Bakelman, Micah D. Gearhart, Eugenia G. Giannopoulou, Paul Zumbo, Kevin Kirouac, Srividya Bhaskara, Jose M. Polo, Matthias Kormaksson, Alexander D. MacKerell, Fengtian Xue, Christopher E. Mason, Scott W. Hiebert, Gilbert G. Prive, Leandro Cerchietti, Vivian J. Bardwell, Olivier Elemento, and Ari Melnick. A hybrid mechanism of action for BCL6 in B-cells defined by formation of functionally distinct complexes at enhancers and promoters. Cell reports, 4(3), August 2013.

[139] D. K. Hawley and R. G. Roeder. Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. The Journal of Biological Chemistry, 260(13):8163–8172, July 1985.

[140] D. K. Hawley and R. G. Roeder. Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. The Journal of Biological Chemistry, 262(8):3452–3461, March 1987.

[141] Andreas M. F. Heilmann and Nicholas J. Dyson. Phosphorylation puts the pRb tumor suppressor into shape. Genes & Development, 26(11):1128–1130, June 2012.

[142] Sven Heinz, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular Cell, 38(4):576–589, May 2010.

[143] Harvey R. Herschman. Primary Response Genes Induced by Growth Factors and Tumor Promoters. Annual Review of Biochemistry, 60(1):281–319, 1991.

[144] K. Hirayoshi and J. T. Lis. Nuclear run-on assays: assessing transcription by measuring density of engaged RNA polymerases. Methods in Enzymology, 304:351–362, 1999.

[145] Denes Hnisz, Krishna Shrinivas, Richard A. Young, Arup K. Chakraborty, and Phillip A. Sharp. A phase separation model predicts key features of transcriptional control. Cell, 169(1):13–23, March 2017.

[146] Caitriona Holohan, Sandra Van Schaeybroeck, Daniel B. Longley, and Patrick G. Johnston. Cancer drug resistance: an evolving paradigm. Nature Reviews Cancer, 13(10):714–726, October 2013.

[147] Hana Holubec, Claire M. Payne, Harris Bernstein, Katerina Dvorakova, Carol Bernstein, Caroline N. Waltmire, James A. Warneke, and Harinder Garewal. Assessment of apoptosis by immunohistochemical markers compared to cellular morphology in ex vivo-stressed colonic mucosa. The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society, 53(2):229–235, February 2005.

[148] C. William Hooker and Peter J. Hurlin. Of Myc and Mnt. Journal of Cell Science, 119(2):208–216, January 2006.

[149] Genevieve Housman, Shannon Byler, Sarah Heerboth, Karolina Lapinska, Mckenna Longacre, Nicole Snyder, and Sibaji Sarkar. Drug Resistance in Cancer: An Overview. Cancers, 6(3):1769–1792, September 2014.

[150] Chen-Lin Hsieh, Teng Fei, Yiwen Chen, Tiantian Li, Yanfei Gao, Xiaodong Wang, Tong Sun, Christopher J. Sweeney, Gwo-Shu Mary Lee, Shaoyong Chen, Steven P. Balk, Xiaole Shirley Liu, Myles Brown, and Philip W. Kantoff. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proceedings of the National Academy of Sciences of the United States of America, 111(20):7319–7324, May 2014.

[151] T. Hunter and M. Karin. The regulation of transcription by phosphorylation. Cell, 70(3):375–387, August 1992.

[152] Paola Indovina, Francesca Pentimalli, Nadia Casini, Immacolata Vocca, and Antonio Giordano. RB1 dual role in proliferation and apoptosis: Cell fate control and implications for cancer therapy. Oncotarget, 6(20):17873–17890, June 2015.

[153] V. R. Iyer, M. B. Eisen, D. T. Ross, G. Schuler, T. Moore, J. C. Lee, J. M. Trent, L. M. Staudt, J. Hudson, M. S. Boguski, D. Lashkari, D. Shalon, D. Botstein, and P. O. Brown. The transcriptional program in the response of human fibroblasts to serum. Science (New York, N.Y.), 283(5398):83–87, January 1999.

[154] Melba C. Jaramillo and Donna D. Zhang. The emerging role of the Nrf2–Keap1 signaling pathway in cancer. Genes & Development, 27(20):2179–2191, October 2013.

[155] Célia Jeronimo, Marie-France Langelier, Alain R. Bataille, John M. Pascal, B. Franklin Pugh, and François Robert. Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. Molecular Cell, 64(3):455–466, November 2016.

[156] Lichun Jiang, Felix Schlesinger, Carrie A. Davis, Yu Zhang, Renhua Li, Marc Salit, Thomas R. Gingeras, and Brian Oliver. Synthetic spike-in standards for RNA-seq experiments. Genome Research, 21(9):1543–1551, September 2011.

[157] Qihuang Jin, Li-Rong Yu, Lifeng Wang, Zhijing Zhang, Lawryn H Kasper, Ji-Eun Lee, Chaochen Wang, Paul K Brindle, Sharon Y R Dent, and Kai Ge. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. The EMBO Journal, 30(2):249–262, January 2011.

[158] Liv Johannessen, Thomas B. Sundberg, Daniel J. O'Connell, Raivo Kolde, James Berstler, Katelyn J. Billings, Bernard Khor, Brinton Seashore-Ludlow, Anne Fassl, Caitlin N. Russell, Isabel J. Latorre, Baishan Jiang, Daniel B. Graham, Jose R. Perez, Piotr Sicinski, Andrew J. Phillips, Stuart L. Schreiber, Nathanael S. Gray, Alykhan F. Shamji, and Ramnik J. Xavier. Small-molecule studies identify CDK8 as a regulator of IL-10 in myeloid cells. Nature Chemical Biology, 13(10):1102–1108, October 2017.

[159] Iris Jonkers and John T. Lis. Getting up to speed with transcription elongation by RNA polymerase II. Nature Reviews. Molecular Cell Biology, 16(3):167–177, March 2015.

[160] Myungsoo Joo, Jeffrey G. Wright, Ning Ning Hu, Ruxana T. Sadikot, Gye Young Park, Timothy S. Blackwell, and John W. Christman. Yin Yang 1 enhances cyclooxygenase-2 gene expression in macrophages. American Journal of Physiology. Lung Cellular and Molecular Physiology, 292(5):L1219–1226, May 2007.

[161] Hyunchul Jung, Donghoon Lee, Jongkeun Lee, Donghyun Park, Yeon Jeong Kim, Woong-Yang Park, Dongwan Hong, Peter J. Park, and Eunjung Lee. Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nature Genetics, 47(11):1242–1248, November 2015.

[162] Stefan Jüttner, Thorsten Cramer, Silja Wessler, Anna Walduck, Feng Gao, Frank Schmitz, Christian Wunder, Matthias Weber, Susan M. Fischer, Wolfgang E. Schmidt, Bertram Wiedenmann, Thomas F. Meyer, Michael Naumann, and Michael Höcker. Helicobacter pylori stimulates host cyclooxygenase-2 gene transcription: critical importance of MEK/ERK-dependent activation of USF1/-2 and CREB transcription factors. Cellular Microbiology, 5(11):821–834, 2003.

[163] Michael H. Kagey, Jamie J. Newman, Steve Bilodeau, Ye Zhan, David A. Orlando, Nynke L. van Berkum, Christopher C. Ebmeier, Jesse Goossens, Peter B. Rahl, Stuart S. Levine,

Dylan J. Taatjes, Job Dekker, and Richard A. Young. Mediator and Cohesin Connect Gene Expression and Chromatin Architecture. Nature, 467(7314):430–435, September 2010.

[164] Yeon-Joo Kang, Byron A. Wingerd, Toshi Arakawa, and William L. Smith. Cyclooxygenase-2 Gene Transcription in a Macrophage Model of Inflammation. The Journal of Immunology, 177(11):8111–8122, December 2006.

[165] Avnish Kapoor, Matthew S. Goldberg, Lara K. Cumberland, Kajan Ratnakumar, Miguel F. Segura, Patrick O. Emanuel, Silvia Menendez, Chiara Vardabasso, Gary LeRoy, Claudia I. Vidal, David Polsky, Iman Osman, Benjamin A. Garcia, Eva Hernando, and Emily Bernstein. The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. Nature, 468(7327):1105–1109, December 2010.

[166] Hina N. Khan, Desiree Perlee, Lieke Schoenmaker, Anne-Jan van der Meer, Marek Franitza, Mohammad Reza Toliat, Peter Nürnberg, Aeilko H. Zwinderman, Tom van der Poll, and Brendon P. Scicluna. Leukocyte transcriptional signatures dependent on LPS dosage in human endotoxemia. Journal of Leukocyte Biology, 106(5):1153–1160, November 2019.

[167] Daehwan Kim, Ben Langmead, and Steven L. Salzberg. HISAT: a fast spliced aligner with low memory requirements. Nature Methods, 12(4):357–360, April 2015.

[168] Tae-Kyung Kim, Martin Hemberg, Jesse M. Gray, Allen M. Costa, Daniel M. Bear, Jing Wu, David A. Harmin, Mike Laptewicz, Kellie Barbara-Haley, Scott Kuersten, Eirene Markenscoff-Papadimitriou, Dietmar Kuhl, Haruhiko Bito, Paul F. Worley, Gabriel Kreiman, and Michael E. Greenberg. Widespread transcription at neuronal activity-regulated enhancers. Nature, 465(7295):182–187, May 2010.

[169] Tae-Kyung Kim and Ramin Shiekhattar. Architectural and Functional Commonalities between Enhancers and Promoters. Cell, 162(5):948–959, August 2015.

[170] Killeen S. Kirkconnell, Michelle T. Paulsen, Brian Magnuson, Karan Bedi, and Mats Ljungman. Capturing the dynamic nascent transcriptome during acute cellular responses: The serum response. Biology Open, 5(6):837–847, May 2016.

[171] Felix Klatt, Alexander Leitner, Iana V. Kim, Hung Ho-Xuan, Elisabeth V. Schneider, Franziska Langhammer, Robin Weinmann, Melanie R. Müller, Robert Huber, Gunter Meister, and Claus-D. Kuhn. A precisely positioned MED12 activation helix stimulates CDK8 kinase activity. Proceedings of the National Academy of Sciences, January 2020.

[172] Sandy L. Klemm, Zohar Shipony, and William J. Greenleaf. Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics, 20(4):207–220, April 2019. Number: 4 Publisher: Nature Publishing Group.

[173] Matthew T. Knuesel, Krista D. Meyer, Carrie Bernecky, and Dylan J. Taatjes. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes & Development, 23(4):439–451, February 2009.

[174] Michael F. T. Koehler, Philippe Bergeron, Elizabeth M. Blackwood, Krista Bowman, Kevin R. Clark, Ron Firestein, James R. Kiefer, Klaus Maskos, Mark L. McCleland, Linda

Orren, Laurent Salphati, Steve Schmidt, Elisabeth V. Schneider, Jiansheng Wu, and Maureen H. Beresini. Development of a Potent, Specific CDK8 Kinase Inhibitor Which Phenocopies CDK8/19 Knockout Cells. ACS Medicinal Chemistry Letters, 7(3):223–228, March 2016.

[175] Ivan V. Kulakovskiy, Yulia A. Medvedeva, Ulf Schaefer, Artem S. Kasianov, Ilya E. Vorontsov, Vladimir B. Bajic, and Vsevolod J. Makeev. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Research, 41(D1):D195–D202, January 2013.

[176] Ivan V. Kulakovskiy, Ilya E. Vorontsov, Ivan S. Yevshin, Ruslan N. Sharipov, Alla D. Fedorova, Eugene I. Rumynskiy, Yulia A. Medvedeva, Arturo Magana-Mora, Vladimir B. Bajic, Dmitry A. Papatsenko, Fedor A. Kolpakov, and Vsevolod J. Makeev. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Research, 46(D1):D252–D259, January 2018. Publisher: Oxford Academic.

[177] Ivan V. Kulakovskiy, Ilya E. Vorontsov, Ivan S. Yevshin, Ruslan N. Sharipov, Alla D. Fedorova, Eugene I. Rumynskiy, Yulia A. Medvedeva, Arturo Magana-Mora, Vladimir B. Bajic, Dmitry A. Papatsenko, Fedor A. Kolpakov, and Vsevolod J. Makeev. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Research, 46(D1):D252–D259, January 2018.

[178] Ivan V. Kulakovskiy, Ilya E. Vorontsov, Ivan S. Yevshin, Anastasiia V. Soboleva, Artem S. Kasianov, Haitham Ashoor, Wail Ba-alawi, Vladimir B. Bajic, Yulia A. Medvedeva, Fedor A. Kolpakov, and Vsevolod J. Makeev. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Research, 44(D1):D116–D125, January 2016.

[179] Malika Kumarasiri, Theodosia Teo, Mingfeng Yu, Stephen Philip, Sunita K. C. Basnet, Hugo Albrecht, Matthew J. Sykes, Peng Wang, and Shudong Wang. In Search of Novel CDK8 Inhibitors by Virtual Screening. Journal of Chemical Information and Modeling, February 2017.

[180] Edo Kussell and Stanislas Leibler. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. Science, 309(5743):2075–2078, September 2005.

[181] Thomas Köcher, Peter Pichler, Remco Swart, and Karl Mechtler. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nature Protocols, 7(5):882–890, May 2012.

[182] Kerstin Köhler, Luis Sanchez-Pulido, Verena Höfer, Anika Marko, Chris P. Ponting, Ambrosius P. Snijders, Regina Feederle, Aloys Schepers, and Dominik Boos. The Cdk8/19-cyclin C transcription regulator functions in genome replication through metazoan Sld7. PLOS Biology, 17(1):e2006767, January 2019.

[183] R. Labianca, A. Sobrero, L. Isa, E. Cortesi, S. Barni, D. Nicolella, M. Aglietta, S. Lonardi, D. Corsi, D. Turci, G. D. Beretta, G. Fornarini, E. Dapretto, I. Floriani, and A. Zaniboni. Intermittent versus continuous chemotherapy in advanced colorectal cancer: a randomised 'GISCAD' trial. Annals of Oncology, 22(5):1236–1242, May 2011.

[184] Fan Lai, Ulf A Orom, Matteo Cesaroni, Malte Beringer, Dylan J Taatjes, Gerd A. Blobel, and Ramin Shiekhattar. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature, 494(7438):497–501, February 2013.

[185] Michael T. Lam, Wenbo Li, Michael G. Rosenfeld, and Christopher K. Glass. Enhancer RNAs and regulated transcriptional programs. Trends in biochemical sciences, 39(4):170–182, April 2014.

[186] Michael T.Y. Lam, Han Cho, Hanna P. Lesch, David Gosselin, Sven Heinz, Yumiko Tanaka-Oishi, Christopher Benner, Minna U. Kaikkonen, Aneeza S. Kim, Mika Kosaka, Cindy Y. Lee, Andy Watt, Tamar R. Grossman, Michael G. Rosenfeld, Ronald M. Evans, and Christopher K. Glass. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature, 498(7455):511–515, June 2013.

[187] Samuel A. Lambert, Arttu Jolma, Laura F. Campitelli, Pratyush K. Das, Yimeng Yin, Mihai Albu, Xiaoting Chen, Jussi Taipale, Timothy R. Hughes, and Matthew T. Weirauch. The Human Transcription Factors. Cell, 172(4):650–665, February 2018.

[188] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4):357–359, April 2012.

[189] Tong Ihn Lee and Richard A. Young. Transcriptional Regulation and Its Misregulation in Disease. Cell, 152(6):1237–1251, March 2013.

[190] Richard M. Leggett, Ricardo H. Ramirez-Gonzalez, Bernardo J. Clavijo, Darren Waite, and Robert P. Davey. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. Frontiers in Genetics, 4, December 2013.

[191] G. Leone, J. DeGregori, R. Sears, L. Jakoi, and J. R. Nevins. Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F. Nature, 387(6631):422–426, May 1997.

[192] Tom Lesluyes, James Johnson, Philip Machanick, and Timothy L Bailey. Differential motif enrichment analysis of paired ChIP-seq experiments. BMC Genomics, 15(1), September 2014.

[193] Chengping Li, Samil Jung, Soonduck Lee, Dongjun Jeong, Young Yang, Keun-Il Kim, Jong-Seok Lim, Chung-Il Cheon, Changjin Kim, Young-Sook Kang, and Myeong-Sok Lee. Nutrient/serum starvation derived TRIP-Br3 down-regulation accelerates apoptosis by destabilizing XIAP. Oncotarget, 6(10):7522–7535, April 2015.

[194] Jia Li, Xiaoyu Li, Xiangjie Kong, Qifeng Luo, Junfeng Zhang, and Lin Fang. MiRNA-26b inhibits cellular proliferation by targeting CDK8 in breast cancer. International Journal of Clinical and Experimental Medicine, 7(3):558–565, March 2014.

[195] MingHua Li, XiaoDan Zhao, Ying Liu, Jun An, Hui Xiao, and Chao Wang. Aberrant expression of CDK8 regulates the malignant phenotype and associated with poor prognosis in human laryngeal squamous cell carcinoma. European archives of oto-rhino-laryngology: official journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS): affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery, February 2017.

[196] Wenbo Li, Dimple Notani, Qi Ma, Bogdan Tanasa, Esperanza Nunez, Aaron Yun Chen, Daria Merkurjev, Jie Zhang, Kenneth Ohgi, Xiaoyuan Song, Soohwan Oh, Hong-Sook Kim, Christopher K. Glass, and Michael G. Rosenfeld. Functional Importance of eRNAs for Estrogen-dependent Transcriptional Activation Events. Nature, 498(7455):516–520, June 2013.

[197] Wenbo Li, Dimple Notani, and Michael G. Rosenfeld. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nature Reviews Genetics, 17(4):207–223, April 2016.

[198] Jiaxin Liang, Mengqian Chen, Eugenia V. Broude, and Igor B. Roninson. Role of transcription-regulating kinase CDK8 in colon cancer metastasis. Oncotarget, 10(6):622–623, January 2019.

[199] Jiaxin Liang, Mengqian Chen, Daniel Hughes, Alexander A. Chumanevich, Serena Altilia, Vimala Kaza, Chang-Uk Lim, Hippokratis Kiaris, Karthikeyan Mythreye, Maria Marjorette Pena, Eugenia V. Broude, and Igor B. Roninson. CDK8 Selectively Promotes the Growth of Colon Cancer Metastases in the Liver by Regulating Gene Expression of TIMP3 and Matrix Metalloproteinases. Cancer Research, 78(23):6594–6606, December 2018.

[200] Lim. Role of CDK8 and -catenin in colorectal adenocarcinoma. Oncology Reports, 24(1), May 2010.

[201] Chengqi Lin, Edwin R. Smith, Hidehisa Takahashi, Ka Chun Lai, Skylar Martin-Brown, Laurence Florens, Michael P. Washburn, Joan W. Conaway, Ronald C. Conaway, and Ali Shilatifard. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Molecular Cell, 37(3):429–437, February 2010.

[202] Ye Lin, Zhixiang Jian, Haosheng Jin, Xiangling Wei, Xiongfeng Zou, Renguo Guan, and Jianfeng Huang. Long non-coding RNA DLGAP1-AS1 facilitates tumorigenesis and epithelial–mesenchymal transition in hepatocellular carcinoma via the feedback loop of miR-26a/b-5p/IL-6/JAK2/STAT3 and Wnt/-catenin pathway. Cell Death & Disease, 11(1):1–17, January 2020.

[203] Jason M. Link, Sara Ota, Zi-Qiang Zhou, Colin J. Daniel, Rosalie C. Sears, and Peter J. Hurlin. A critical role for Mnt in Myc-driven T-cell proliferation and oncogenesis. Proceedings of the National Academy of Sciences, 109(48):19685–19690, November 2012.

[204] John T. Lis. A 50 year history of technologies that drove discovery in eukaryotic transcription regulation. Nature Structural & Molecular Biology, 26(9):777–782, September 2019. Number: 9 Publisher: Nature Publishing Group.

[205] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 2014.

[206] Xin Luo, Minho Chae, Raga Krishnakumar, Charles G. Danko, and W. Lee Kraus. Dynamic reorganization of the AC16 cardiomyocyte transcriptome in response to TNF signaling revealed by integrated genomic analyses. BMC Genomics, 15(1):155, February 2014.

[207] Zhuojuan Luo, Chengqi Lin, and Ali Shilatifard. The super elongation complex (SEC) family in transcriptional control. Nature Reviews Molecular Cell Biology, 13(9):543–547, September 2012.

[208] H. D. Madhani and G. R. Fink. Combinatorial control required for the specificity of yeast MAPK signaling. Science (New York, N.Y.), 275(5304):1314–1317, February 1997.

[209] Dig B. Mahat, H. Hans Salamanca, Fabiana M. Duarte, Charles G. Danko, and John T. Lis. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell, 62(1):63–78, April 2016.

[210] Dig B. Mahat, H. Hans Salamanca, Fabiana M. Duarte, Charles G. Danko, and John T. Lis. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular Cell, 62(1):63–78, April 2016.

[211] Dig Bijay Mahat, Hojoong Kwak, Gregory T. Booth, Iris H. Jonkers, Charles G. Danko, Ravi K. Patel, Colin T. Waters, Katie Munson, Leighton J. Core, and John T. Lis. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nature Protocols, 11(8):1455–1476, August 2016. Number: 8 Publisher: Nature Publishing Group.

[212] Sohail Malik and Robert G. Roeder. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nature Reviews Genetics, 11(11):761–772, November 2010.

[213] Sohail Malik and Robert G. Roeder. Mediator: A Drawbridge across the Enhancer-Promoter Divide. Molecular Cell, 64(3):433–434, November 2016.

[214] Aurélie Mallinger, Kai Schiemann, Christian Rink, Frank Stieber, Michel Calderini, Simon Crumpler, Mark Stubbs, Olajumoke Adeniji-Popoola, Oliver Poeschke, Michael Busch, Paul Czodrowski, Djordje Musil, Daniel Schwarz, Maria-Jesus Ortiz-Ruiz, Richard Schneider, Ching Thai, Melanie Valenti, Alexis de Haven Brandon, Rosemary Burke, Paul Workman, Trevor Dale, Dirk Wienke, Paul A. Clarke, Christina Esdar, Florence I. Raynaud, Suzanne A. Eccles, Felix Rohdich, and Julian Blagg. Discovery of Potent, Selective, and Orally Bioavailable Small-Molecule Modulators of the Mediator Complex-Associated Kinases CDK8 and CDK19. Journal of Medicinal Chemistry, 59(3):1078–1101, February 2016.

[215] Xiangying Mao, Jeong Im Kim, Mitchell T. Wheeler, Anne K. Heintzelman, Vikki M. Weake, and Clint Chapple. Mutation of Mediator subunit CDK8 counteracts the stunted growth and salicylic acid hyper-accumulation phenotypes of an Arabidopsis MED5 mutant. New Phytologist, 0(ja).

[216] Edyta Marcon, Harshika Jain, Anandi Bhattacharya, Hongbo Guo, Sadhna Phanse, Shuye Pu, Gregory Byram, Ben C. Collins, Evan Dowdell, Maria Fenner, Xinghua Guo, Ashley Hutchinson, Jacob J. Kennedy, Bryan Krastins, Brett Larsen, Zhen-Yuan Lin, Mary F. Lopez, Peter Loppnau, Shane Miersch, Tin Nguyen, Jonathan B. Olsen, Marcin Paduch, Mani Ravichandran, Alma Seitova, Gouri Vadali, Maryann S. Vogelsang, Jeffrey R. Whiteaker, Guoqing Zhong, Nan Zhong, Lei Zhao, Ruedi Aebersold, Cheryl H. Arrowsmith, Andrew Emili, Lori Frappier, Anne-Claude Gingras, Matthias Gstaiger, Amanda G. Paulovich, Shohei Koide, Anthony A. Kossiakoff, Sachdev S. Sidhu, Shoshana J. Wodak, Susanne Gräslund, Jack F. Greenblatt, and Aled M. Edwards. Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. Nature Methods, 12(8):725–731, August 2015. Number: 8 Publisher: Nature Publishing Group.

[217] Manuel de la Mata, Claudio R. Alonso, Sebastián Kadener, Juan P. Fededa, Matıas Blaustein, Federico Pelisch, Paula Cramer, David Bentley, and Alberto R. Kornblihtt. A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Molecular Cell, 12(2):525–532, August 2003.

[218] Matthew T. Maurano, Richard Humbert, Eric Rynes, Robert E. Thurman, Eric Haugen, Hao Wang, Alex P. Reynolds, Richard Sandstrom, Hongzhu Qu, Jennifer Brody, Anthony Shafer, Fidencio Neri, Kristen Lee, Tanya Kutyavin, Sandra Stehling-Sun, Audra K. Johnson, Theresa K. Canfield, Erika Giste, Morgan Diegel, Daniel Bates, R. Scott Hansen, Shane Neph, Peter J. Sabo, Shelly Heimfeld, Antony Raubitschek, Steven Ziegler, Chris Cotsapas, Nona Sotoodehnia, Ian Glass, Shamil R. Sunyaev, Rajinder Kaul, and John A. Stamatoyannopoulos. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (New York, N.Y.), 337(6099):1190–1195, September 2012.

[219] Martina S. J. McDermott, Alexander A. Chumanevich, Chang-uk Lim, Jiaxin Liang, Mengqian Chen, Serena Altilia, David Oliver, James M. Rae, Michael Shtutman, Hippokratis Kiaris, Balázs Győrffy, Igor B. Roninson, Eugenia V. Broude, Martina S. J. McDermott, Alexander A. Chumanevich, Chang-uk Lim, Jiaxin Liang, Mengqian Chen, Serena Altilia, David Oliver, James M. Rae, Michael Shtutman, Hippokratis Kiaris, Balázs Győrffy, Igor B. Roninson, and Eugenia V. Broude. Inhibition of CDK8 Mediator kinase suppresses estrogen dependent transcription and the growth of estrogen receptor positive breast cancer. Oncotarget, 5(0), January 2017.

[220] Ian C. McDowell, Alejandro Barrera, Anthony M. D'Ippolito, Christopher M. Vockley, Linda K. Hong, Sarah M. Leichter, Luke C. Bartelt, William H. Majoros, Lingyun Song, Alexias Safi, D. Dewran Koçak, Charles A. Gersbach, Alexander J. Hartemink, Gregory E. Crawford, Barbara E. Engelhardt, and Timothy E. Reddy. Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. Genome Research, August 2018.

[221] Robert C. McLeay and Timothy L. Bailey. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics, 11(1):165, April 2010.

[222] Jan Paul Medema. Cancer stem cells: The challenges ahead. Nature Cell Biology, 15(4):338–344, April 2013.

[223] Ingeborg Menzl, Agnieszka Witalisz-Siepracka, and Veronika Sexl. CDK8-Novel Therapeutic Opportunities. Pharmaceuticals, 12(2), June 2019.

[224] Krista D. Meyer, Shih-Chieh Lin, Carrie Bernecky, Yuefeng Gao, and Dylan J. Taatjes. p53 activates transcription by directing structural shifts in Mediator. Nature Structural & Molecular Biology, 17(6):753–760, June 2010.

[225] Joëlle Michaud, Viviane Praz, Nicole James Faresse, Courtney K. JnBaptiste, Shweta Tyagi, Frédéric Schütz, and Winship Herr. HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. Genome Research, 23(6):907–916, June 2013. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[226] Irene M. Min, Joshua J. Waterfall, Leighton J. Core, Robert J. Munroe, John Schimenti, and John T. Lis. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes & Development, 25(7):742–754, April 2011.

[227] Nacho Molina, David M. Suter, Rosamaria Cannavo, Benjamin Zoller, Ivana Gotic, and Félix Naef. Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. Proceedings of the National Academy of Sciences of the United States of America, 110(51):20563–20568, December 2013.

[228] Céline Molle, Michel Goldman, and Stanislas Goriely. Critical Role of the IFN-Stimulated Gene Factor 3 Complex in TLR-Mediated IL-27p28 Gene Expression Revealing a Two-Step Activation Process. The Journal of Immunology, 184(4):1784–1792, February 2010.

[229] Marta Montes, Alexandre Cloutier, Noemí Sánchez-Hernández, Laetitia Michelle, Bruno Lemieux, Marco Blanchette, Cristina Hernández-Munain, Benoit Chabot, and Carlos Suñé. TCERG1 regulates alternative splicing of the Bcl-x gene by modulating the rate of RNA polymerase II transcription. Molecular and Cellular Biology, 32(4):751–762, February 2012.

[230] Chris B. Moore, Elizabeth H. Guthrie, Max Tze-Han Huang, and Debra J. Taxman. Short Hairpin RNA (shRNA): Design, Delivery, and Assessment of Gene Knockdown. Methods in molecular biology (Clifton, N.J.), 629:141–158, 2010.

[231] Magdalena Morawska and Helle D Ulrich. An expanded tool kit for the auxin-inducible degron system in budding yeast. Yeast (Chichester, England), 30(9):341–351, September 2013.

[232] Kevin V. Morris, Sharon Santoso, Anne-Marie Turner, Chiara Pastori, and Peter G. Hawkins. Bidirectional Transcription Directs Both Transcriptional Gene Activation and Suppression in Human Cells. PLoS Genetics, 4(11), November 2008.

[233] Jorge Moscat, Adam Richardson, and Maria T. Diaz-Meco. Nutrient stress revamps cancer cell metabolism. Cell Research, 25(5):537–538, May 2015.

[234] Justin Moser, Iain Miller, Dylan Carter, and Sabrina L. Spencer. Control of the Restriction Point by Rb and p21. Proceedings of the National Academy of Sciences, 115(35):E8219–E8227, August 2018.

[235] David M. Mosser and Justin P. Edwards. Exploring the full spectrum of macrophage activation. Nature Reviews Immunology, 8(12):958–969, December 2008.

[236] Arijit Mukhopadhyay, Jamie M. Kramer, Gerard Merkx, Dorien Lugtenberg, Dominique F. Smeets, Merel A. W. Oortveld, Ellen A. W. Blokland, Jyoti Agrawal, Annette Schenck, Hans van Bokhoven, Erik Huys, Eric F. Schoenmakers, Ad Geurts van Kessel, C. Erik van Nouhuys, and Frans P. M. Cremers. CDK19 is disrupted in a female patient with bilateral congenital retinal folds, microcephaly and mild mental retardation. Human Genetics, 128(3):281–291, September 2010.

[237] Patricia A. J. Muller and Karen H. Vousden. p53 mutations in cancer. Nature Cell Biology, 15(1):2, January 2013.

[238] Patricia A. J. Muller and Karen H. Vousden. Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. Cancer Cell, 25(3):304–317, March 2014.

[239] Nadja Muncke, Christine Jung, Heinz Rüdiger, Herbert Ulmer, Ralph Roeth, Annette Hubert, Elizabeth Goldmuntz, Deborah Driscoll, Judith Goodship, Karin Schön, and Gudrun Rappold. Missense Mutations and Gene Interruption in PROSIT240, a Novel TRAP240-Like Gene, in Patients With Congenital Heart Defect (Transposition of the Great Arteries). Circulation, 108(23):2843–2850, December 2003.

[240] Ginger W. Muse, Daniel A. Gilchrist, Sergei Nechaev, Ruchir Shah, Joel S. Parker, Sherry F. Grissom, Julia Zeitlinger, and Karen Adelman. RNA polymerase is poised for activation across the genome. Nature Genetics, 39(12):1507–1511, December 2007. Number: 12 Publisher: Nature Publishing Group.

[241] Akihiko Muto, Shingo Ikeda, Martha E. Lopez-Burks, Yutaka Kikuchi, Anne L. Calof, Arthur D. Lander, and Thomas F. Schilling. Nipbl and Mediator Cooperatively Regulate Gene Expression to Control Limb Development. PLOS Genetics, 10(9):e1004671, September 2014.

[242] Manuel J. Muñoz, M. Soledad Pérez Santangelo, Maria P. Paronetto, Manuel de la Mata, Federico Pelisch, Stéphanie Boireau, Kira Glover-Cutter, Claudia Ben-Dov, Matías Blaustein, Juan J. Lozano, Gregory Bird, David Bentley, Edouard Bertrand, and Alberto R. Kornblihtt. DNA Damage Regulates Alternative Splicing through Inhibition of RNA Polymerase II Elongation. Cell, 137(4):708–720, May 2009.

[243] Jing Nan, Yuxin Wang, Jinbo Yang, and George R. Stark. IRF9 and unphosphorylated STAT2 cooperate with NF-B to drive IL6 expression. Proceedings of the National Academy of Sciences, 115(15):3906–3911, April 2018.

[244] National Cancer Institute. Drugs Approved for Different Types of Cancer.

[245] Toyoaki Natsume, Tomomi Kiyomitsu, Yumiko Saga, and Masato T. Kanemaki. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. Cell Reports, 15(1):210–218, April 2016.

[246] Chris Nelson, Susan Goto, Karen Lund, Wesley Hung, and Ivan Sadowski. Srb10/Cdk8 regulates yeast filamentous growth by phosphorylating the transcription factor Ste12. Nature, 421(6919):187–190, January 2003.

[247] Jonas A. Nilsson and John L. Cleveland. Myc pathways provoking cell suicide and cancer. Oncogene, 22(56):9007, December 2003.

[248] Kohei Nishimura, Tatsuo Fukagawa, Haruhiko Takisawa, Tatsuo Kakimoto, and Masato Kanemaki. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nature Methods, 6(12):917–922, December 2009.

[249] Edgar Nägele, Martin Vollmer, Patric Hörth, and Cornelia Vad. 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. Expert Review of Proteomics, 1(1):37–46, June 2004.

[250] Peter Oettgen. The Role of Ets Factors in Tumor Angiogenesis. Journal of Oncology, 2010:e767384, May 2010.

[251] Koji Ono, Hiroshi Banno, Masanori Okaniwa, Takaharu Hirayama, Naoki Iwamura, Yukiko Hikichi, Saomi Murai, Maki Hasegawa, Yuka Hasegawa, Kazuko Yonemori, Akito Hata, Kazunobu Aoyama, and Douglas R. Cary. Design and synthesis of selective CDK8/19 dual inhibitors: Discovery of 4,5-dihydrothieno[3,4:3,4]benzo[1,2-d]isothiazole derivatives. Bioorganic & Medicinal Chemistry.

[252] Toshinori Ozaki and Akira Nakagawara. Role of p53 in Cell Death and Human Cancers. Cancers, 3(1):994–1013, March 2011.

[253] Daniel J. O'Connell, Raivo Kolde, Matthew Sooknah, Daniel B. Graham, Thomas B. Sundberg, Isabel J. Latorre, Tarjei S. Mikkelsen, and Ramnik J. Xavier. Simultaneous Pathway Activity Inference and Gene Expression Analysis Using RNA Sequencing. Cell Systems, 2(5):323–334, May 2016.

[254] Robert-Jan Palstra and Frank Grosveld. Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. Frontiers in Genetics, 3, September 2012.

[255] J. Brandon Parker, Santanu Palchaudhuri, Hanwei Yin, Jianjun Wei, and Debabrata Chakravarti. A Transcriptional Regulatory Role of the THAP11–HCF-1 Complex in Colon Cancer Cell Function. Molecular and Cellular Biology, 32(9):1654–1670, May 2012.

[256] Samantha E. Parker, Cara T. Mai, Mark A. Canfield, Russel Rickard, Ying Wang, Robert E. Meyer, Patrick Anderson, Craig A. Mason, Julianne S. Collins, Russell S. Kirby, Adolfo Correa, and National Birth Defects Prevention Network. Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006. Birth Defects Research. Part A, Clinical and Molecular Teratology, 88(12):1008–1016, December 2010.

[257] Michelle T. Paulsen, Artur Veloso, Jayendra Prasad, Karan Bedi, Emily A. Ljungman, Ya-Chun Tsan, Ching-Wei Chang, Brendan Tarrier, Joseph G. Washburn, Robert Lyons, Daniel R. Robinson, Chandan Kumar-Sinha, Thomas E. Wilson, and Mats Ljungman. Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. Proceedings of the National Academy of Sciences, 110(6):2240–2245, February 2013. Publisher: National Academy of Sciences Section: Biological Sciences.

[258] Natalya N. Pavlova and Craig B. Thompson. The Emerging Hallmarks of Cancer Metabolism. Cell Metabolism, 23(1):27–47, January 2016.

[259] Henry E. Pelish, Brian B. Liau, Ioana I. Nitulescu, Anupong Tangpeerachaikul, Zachary C. Poss, Diogo H. Da Silva, Brittany T. Caruso, Alexander Arefolov, Olugbeminiyi Fadeyi, Amanda L. Christie, Karrie Du, Deepti Banka, Elisabeth V. Schneider, Anja Jestel, Ge Zou, Chong Si, Christopher C. Ebmeier, Roderick T. Bronson, Andrei V. Krivtsov, Andrew G. Myers, Nancy E. Kohl, Andrew L. Kung, Scott A. Armstrong, Madeleine E. Lemieux, Dylan J. Taatjes, and Matthew D. Shair. Mediator kinase inhibition further activates super-enhancer-associated genes in AML. Nature, 526(7572):273–276, October 2015. Number: 7572 Publisher: Nature Publishing Group.

[260] R. Pergolizzi, V. Appierto, M. Crosti, E. Cavadini, L. Cleris, A. Guffanti, and F. Formelli. Role of retinoic acid receptor overexpression in sensitivity to fenretinide and tumorigenicity of human ovarian carcinoma cells. International Journal of Cancer, 81(5):829–834, May 1999.

[261] Natalia Petrenko, Yi Jin, Koon Ho Wong, and Kevin Struhl. Evidence that Mediator is essential for Pol II transcription, but is not a required component of the preinitiation complex in vivo. eLife, 6:e28447, July 2017.

[262] Jennifer E. Phillips-Cremins, Michael E. G. Sauria, Amartya Sanyal, Tatiana I. Gerasimova, Bryan R. Lajoie, Joshua S. K. Bell, Chin-Tong Ong, Tracy A. Hookway, Changying Guo, Yuhua Sun, Michael J. Bland, William Wagstaff, Stephen Dalton, Todd C. McDevitt, Ranjan Sen, Job Dekker, James Taylor, and Victor G. Corces. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell, 153(6):1281–1295, June 2013.

[263] Genaro Pimienta, Raghothama Chaerkady, and Akhilesh Pandey. SILAC for Global Phosphoproteomic Analysis. In Phospho-Proteomics, Methods in Molecular Biology^{TM}, pages 107–116. Humana Press, 2009.

[264] Jennifer L. Plank and Ann Dean. Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. Molecular Cell, 55(1):5–14, July 2014.

[265] Clemens Plaschka, Kayo Nozawa, and Patrick Cramer. Mediator architecture and RNA polymerase II interaction. Journal of Molecular Biology.

[266] Donald C. Porter, Elena Farmaki, Serena Altilia, Gary P. Schools, Deborah K. West, Mengqian Chen, Bey-Dih Chang, Anatoliy T. Puzyrev, Chang-uk Lim, Rebecca Rokow-Kittell, Lawrence T. Friedhoff, Athanasios G. Papavassiliou, Swathi Kalurupalle, Gregory Hurteau, Jun Shi, Phil S. Baran, Balazs Gyorffy, Mark P. Wentland, Eugenia V. Broude, Hippokratis Kiaris, and Igor B. Roninson. Cyclin-dependent kinase 8 mediates chemotherapy-induced tumor-promoting paracrine activities. Proceedings of the National Academy of Sciences, 109(34):13799–13804, August 2012.

[267] Zachary C. Poss, Christopher C. Ebmeier, Aaron T. Odell, Anupong Tangpeerachaikul, Thomas Lee, Henry E. Pelish, Matthew D. Shair, Robin D. Dowell, William M. Old, and Dylan J. Taatjes. Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. Cell Reports, 15(2):436–450, April 2016.

[268] Zachary C. Poss, Christopher C. Ebmeier, and Dylan J. Taatjes. The Mediator complex and transcription regulation. Critical Reviews in Biochemistry and Molecular Biology, 48(6):575–608, December 2013.

[269] Zachary C. Poss, Christopher C. Ebmeier, Aaron T. Odell, Anupong Tangpeerachaikul, Thomas Lee, Henry E. Pelish, Matthew D. Shair, Robin D. Dowell, William M. Old, and Dylan J. Taatjes. Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. Cell Reports, 15(2):436–450, April 2016.

[270] Angela P. Presson, Ginger Partyka, Kristin M. Jensen, Owen J. Devine, Sonja A. Rasmussen, Linda L. McCabe, and Edward R. B. McCabe. Current estimate of Down Syndrome population prevalence in the United States. The Journal of Pediatrics, 163(4):1163–1168, October 2013.

[271] N. J. Proudfoot. How RNA polymerase II terminates transcription in higher eukaryotes. Trends in Biochemical Sciences, 14(3):105–110, March 1989.

[272] Kim D. Pruitt, Garth R. Brown, Susan M. Hiatt, Françoise Thibaud-Nissen, Alexander Astashyn, Olga Ermolaeva, Catherine M. Farrell, Jennifer Hart, Melissa J. Landrum, Kelly M. McGarvey, Michael R. Murphy, Nuala A. O'Leary, Shashikant Pujar, Bhanu Rajput, Sanjida H. Rangwala, Lillian D. Riddick, Andrei Shkeda, Hanzhen Sun, Pamela Tamez, Raymond E. Tully, Craig Wallin, David Webb, Janet Weber, Wendy Wu, Michael DiCuccio, Paul Kitts, Donna R. Maglott, Terence D. Murphy, and James M. Ostell. RefSeq: an update on mammalian reference sequences. Nucleic Acids Research, 42(Database issue):D756–763, January 2014.

[273] Eva Maria Putz, Dagmar Gotthardt, Gregor Hoermann, Agnes Csiszar, Silvia Wirth, Angelika Berger, Elisabeth Straka, Doris Rigler, Barbara Wallner, Amanda M. Jamieson, Winfried F. Pickl, Eva Maria Zebedin-Brandl, Mathias Müller, Thomas Decker, and Veronika Sexl. CDK8-Mediated STAT1-S727 Phosphorylation Restrains NK Cell Cytotoxicity and Tumor Surveillance. Cell Reports, 4(3):437–444, August 2013.

[274] Eva Maria Putz, Dagmar Gotthardt, Gregor Hoermann, Agnes Csiszar, Silvia Wirth, Angelika Berger, Elisabeth Straka, Doris Rigler, Barbara Wallner, Amanda M. Jamieson, Winfried F. Pickl, Eva Maria Zebedin-Brandl, Mathias Müller, Thomas Decker, and Veronika Sexl. CDK8-Mediated STAT1-S727 Phosphorylation Restrains NK Cell Cytotoxicity and Tumor Surveillance. Cell Reports, 4(3):437–444, August 2013.

[275] Aaron R. Quinlan and Ira M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6):841–842, March 2010.

[276] Owen J. L. Rackham, Jaber Firas, Hai Fang, Matt E. Oates, Melissa L. Holmes, Anja S. Knaupp, The FANTOM Consortium, Harukazu Suzuki, Christian M. Nefzger, Carsten O. Daub, Jay W. Shin, Enrico Petretto, Alistair R. R. Forrest, Yoshihide Hayashizaki, Jose M. Polo, and Julian Gough. A predictive computational framework for direct reprogramming between human cell types. Nature Genetics, 48(3):331–335, March 2016.

[277] Peter B. Rahl, Charles Y. Lin, Amy C. Seila, Ryan A. Flynn, Scott McCuine, Christopher B. Burge, Phillip A. Sharp, and Richard A. Young. c-Myc regulates transcriptional pause release. Cell, 141(3):432–445, April 2010.

[278] Ryan Raisner, Samir Kharbanda, Lingyan Jin, Edwin Jeng, Emily Chan, Mark Merchant, Peter M. Haverty, Russell Bainer, Tommy Cheung, David Arnott, E. Megan Flynn, F. Anthony Romero, Steven Magnuson, and Karen E. Gascoigne. Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. Cell Reports, 24(7):1722–1729, August 2018.

[279] Sheetal Raithatha, Ting-Cheng Su, Pedro Lourenco, Susan Goto, and Ivan Sadowski. Cdk8 regulates stability of the transcription factor Phd1 to control pseudohyphal differentiation of Saccharomyces cerevisiae. Molecular and Cellular Biology, 32(3):664–674, February 2012.

[280] Chilakamarti V. Ramana, M. Pilar Gil, Yulong Han, Richard M. Ransohoff, Robert D. Schreiber, and George R. Stark. Stat1-independent regulation of gene expression in response to IFN-. Proceedings of the National Academy of Sciences, 98(12):6674–6679, June 2001. Publisher: National Academy of Sciences Section: Biological Sciences.

[281] F. Ann Ran, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. Genome engineering using the CRISPR-Cas9 system. Nature Protocols, 8(11):2281–2308, November 2013.

[282] Timothy Read, Phillip A. Richmond, and Robin D. Dowell. A trans -acting Variant within the Transcription Factor RIM101 Interacts with Genetic Background to Determine its Regulatory Capacity. PLOS Genet, 12(1):e1005746, January 2016.

[283] Ho Sung Rhee and B. Franklin Pugh. ChIP-exo: A Method to Identify Genomic Location of DNA-binding proteins at Near Single Nucleotide Accuracy. Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.], 0 21, October 2012.

[284] Carlo Riccardi and Ildo Nicoletti. Analysis of apoptosis by propidium iodide staining and flow cytometry. Nature Protocols, 1(3):1458–1461, November 2006.

[285] P. Rickert, W. Seghezzi, F. Shanahan, H. Cho, and E. Lees. Cyclin C/CDK8 is a novel CTD kinase associated with RNA polymerase II. Oncogene, 12(12):2631–2640, June 1996.

[286] James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. Integrative genomics viewer. Nature Biotechnology, 29(1):24–26, January 2011.

[287] Jonathan D. Rubin, Jacob T. Stanley, Rutendo F. Sigauke, Cecilia B. Levandowski, Zachary L. Maas, Jessica Westfall, Dylan J. Taatjes, and Robin D. Dowell. Transcription factor enrichment analysis (TFEA): Quantifying the activity of hundreds of transcription factors from a single experiment. bioRxiv, page 2020.01.25.919738, February 2020.

[288] Jonathan D. Rubin and Dylan J. Taatjes. Molecular biology: Mediating transcription and RNA export. Nature, 526(7572):199–200, October 2015.

[289] Jonathan D. Rubin and Dylan J. Taatjes. Molecular biology: Mediating transcription and RNA export. Nature, 526(7572):199–200, October 2015.

[290] Seth M. Rubin. Deciphering the Rb phosphorylation code. Trends in biochemical sciences, 38(1):12–19, January 2013.

[291] Natalia Sacilotto, Kira M. Chouliaras, Leonid L. Nikitenko, Yao Wei Lu, Martin Fritzsche, Marsha D. Wallace, Svanhild Nornes, Fernando García-Moreno, Sophie Payne, Esther Bridges, Ke Liu, Daniel Biggs, Indrika Ratnayaka, Shane P. Herbert, Zoltán Molnár, Adrian L. Harris, Benjamin Davies, Gareth L. Bond, George Bou-Gharios, John J. Schwarz, and Sarah De Val. MEF2 transcription factors are key regulators of sprouting angiogenesis. Genes & Development, 30(20):2297–2309, October 2016.

[292] Nozomi Saito and Atsushi Kittaka. Highly potent vitamin D receptor antagonists: design, synthesis, and biological evaluation. Chembiochem: A European Journal of Chemical Biology, 7(10):1479–1490, October 2006.

[293] Amartya Sanyal, Bryan Lajoie, Gaurav Jain, and Job Dekker. The long-range interaction landscape of gene promoters. Nature, 489(7414):109–113, September 2012.

[294] Kai Schiemann, Aurélie Mallinger, Dirk Wienke, Christina Esdar, Oliver Poeschke, Michael Busch, Felix Rohdich, Suzanne A. Eccles, Richard Schneider, Florence I. Raynaud, Paul Czodrowski, Djordje Musil, Daniel Schwarz, Klaus Urbahns, and Julian Blagg. Discovery of potent and selective CDK8 inhibitors from an HSP90 pharmacophore. Bioorganic & Medicinal Chemistry Letters, 26(5):1443–1451, March 2016.

[295] Allison C. Schier and Dylan J. Taatjes. Structure and mechanism of the RNA polymerase II transcription machinery. Genes & Development, 34(7-8):465–488, April 2020. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[296] Eric C. Schirmer, John R. Yates, and Larry Gerace. MudPIT: A powerful proteomics tool for discovery. Discovery Medicine, 3(18):38–39, October 2003.

[297] E. V. Schneider, J. Böttcher, M. Blaesse, L. Neumann, R. Huber, and K. Maskos. The Structure of CDK8/CycC Implicates Specificity in the CDK/Cyclin Family and Reveals Interaction with a Deep Pocket Binder. Journal of Molecular Biology, 412(2):251–266, September 2011.

[298] Björn Schwalb, Margaux Michel, Benedikt Zacher, Katja Frühauf, Carina Demel, Achim Tresch, Julien Gagneur, and Patrick Cramer. TT-seq maps the human transient transcriptome. Science, 352(6290):1225–1228, June 2016. Publisher: American Association for the Advancement of Science Section: Report.

[299] Bradley J. Scott, Sami Qutob, Qing Y. Liu, and Cheng E. Ng. APM2 is a novel mediator of cisplatin resistance in a variety of cancer cell types regardless of p53 or MMR status. International Journal of Cancer, 125(5):1193–1204, September 2009.

[300] Benjamin S. Scruggs, Daniel A. Gilchrist, Sergei Nechaev, Ginger W. Muse, Adam Burkholder, David C. Fargo, and Karen Adelman. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Molecular Cell, 58(6):1101–1112, June 2015.

[301] Amy C. Seila, J. Mauro Calabrese, Stuart S. Levine, Gene W. Yeo, Peter B. Rahl, Ryan A. Flynn, Richard A. Young, and Phillip A. Sharp. Divergent transcription from active promoters. Science (New York, N.Y.), 322(5909):1849–1851, December 2008.

[302] Adrien Senecal, Brian Munsky, Florence Proux, Nathalie Ly, Floriane E. Braye, Christophe Zimmer, Florian Mueller, and Xavier Darzacq. Transcription Factors Modulate c-Fos Transcriptional Bursts. Cell reports, 8(1):75–83, July 2014.

[303] Jayasha Shandilya and Stefan G. E. Roberts. The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1819(5):391–400, May 2012.

[304] Libin Shang, She Chen, Fenghe Du, Shen Li, Liping Zhao, and Xiaodong Wang. Nutrient starvation elicits an acute autophagic response mediated by Ulk1 dephosphorylation and its subsequent dissociation from AMPK. Proceedings of the National Academy of Sciences, 108(12):4788–4793, March 2011.

[305] Wanqing Shao and Julia Zeitlinger. Paused RNA polymerase II inhibits new transcriptional initiation. Nature Genetics, 49(7):1045–1051, July 2017. Number: 7 Publisher: Nature Publishing Group.

[306] Sreenath V. Sharma, Diana Y. Lee, Bihua Li, Margaret P. Quinlan, Fumiyuki Takahashi, Shyamala Maheswaran, Ultan McDermott, Nancy Azizian, Lee Zou, Michael A. Fischbach, Kwok-Kin Wong, Kathleyn Brandstetter, Ben Wittner, Sridhar Ramaswamy, Marie Classon, and Jeff Settleman. A chromatin-mediated reversible drug tolerant state in cancer cell subpopulations. Cell, 141(1):69–80, April 2010.

[307] Eitan Shaulian and Michael Karin. AP-1 as a regulator of cell life and death. Nature Cell Biology, 4(5):E131–E136, May 2002.

[308] Ryan M. Sheridan and David L. Bentley. Selectable one-step PCR-mediated integration of a degron for rapid depletion of endogenous human proteins. BioTechniques, 60(2):69–74, February 2016.

[309] Mikyong Shin, Lilah M. Besser, James E. Kucik, Chengxing Lu, Csaba Siffel, Adolfo Correa, and Congenital Anomaly Multistate Prevalence and Survival Collaborative. Prevalence of Down syndrome among children and adolescents in 10 regions of the United States. Pediatrics, 124(6):1565–1571, December 2009.

[310] Daria Shlyueva, Gerald Stampfel, and Alexander Stark. Transcriptional enhancers: from properties to genome-wide predictions. Nature Reviews Genetics, 15(4):272–286, April 2014.

[311] Jakub Siednienko, Ashwini Maratha, Shuo Yang, Malgorzata Mitkiewicz, Sinéad M. Miggin, and Paul N. Moynagh. Nuclear Factor B Subunits RelB and cRel Negatively Regulate Toll-like Receptor 3-mediated -Interferon Production via Induction of Transcriptional Repressor Protein YY1. Journal of Biological Chemistry, 286(52):44750–44763, December 2011.

[312] Robert J. Sims, Rimma Belotserkovskaya, and Danny Reinberg. Elongation by RNA polymerase II: the short and long of it. Genes & Development, 18(20):2437–2468, October 2004. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[313] Erik Snapp. Design and Use of Fluorescent Fusion Proteins in Cell Biology. Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.], CHAPTER:Unit–21.4, July 2005.

[314] Thierry Soussi. TP53 (Tumour protein p53 (Li-Fraumeni syndrome)), 2016. 20(11):552-562.

[315] Brad Spellberg, Robert Guidos, David Gilbert, John Bradley, Helen W. Boucher, W. Michael Scheld, John G. Bartlett, John Edwards, and the Infectious Diseases Society of America. The Epidemic of Antibiotic-Resistant Infections: A Call to Action for the Medical Community from the Infectious Diseases Society of America. Clinical Infectious Diseases, 46(2):155–164, January 2008.

[316] Sabrina L. Spencer, Steven D. Cappell, Feng-Chiao Tsai, K. Wesley Overton, Clifford L. Wang, and Tobias Meyer. The Proliferation-Quiescence Decision Is Controlled by a Bifurcation in CDK2 Activity at Mitotic Exit. Cell, 155(2):369–383, October 2013.

[317] François Spitz and Eileen E. M. Furlong. Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics, 13(9):613–626, September 2012. Number: 9 Publisher: Nature Publishing Group.

[318] Anabella Srebrow and Alberto R. Kornblihtt. The connection between splicing and cancer. Journal of Cell Science, 119(13):2635–2641, July 2006.

[319] Iris Steinparzer, Vitaly Sedlyarov, Jonathan D. Rubin, Kevin Eislmayr, Matthew D. Galbraith, Cecilia B. Levandowski, Terezia Vcelkova, Lucy Sneezum, Florian Wascher, Fabian Amman, Renata Kleinova, Heather Bender, Zdenek Andrysik, Joaquin M. Espinosa, Giulio Superti-Furga, Robin D. Dowell, Dylan J. Taatjes, and Pavel Kovarik. Transcriptional Responses to IFN- Require Mediator Kinase-Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions. Molecular Cell, 76(3):485–499.e8, November 2019.

[320] W. Strober. Trypan blue exclusion test of cell viability. Current Protocols in Immunology, Appendix 3:Appendix 3B, May 2001.

[321] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102(43):15545–15550, October 2005.

[322] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43):15545–15550, October 2005.

[323] Zsolt Szilagyi, Gabor Banyai, Marcela Davila Lopez, Christopher J. McInerny, and Claes M. Gustafsson. Cyclin-Dependent Kinase 8 Regulates Mitotic Commitment in Fission Yeast. Molecular and Cellular Biology, 32(11):2099–2109, June 2012.

[324] Zsolt Szilagyi and Claes M. Gustafsson. Emerging roles of Cdk8 in cell cycle control. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1829(9):916–920, September 2013.

[325] Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, Alexander Roth, Alberto Santos, Kalliopi P. Tsafou, Michael Kuhn, Peer Bork, Lars J. Jensen, and Christian von Mering. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 43(Database issue):D447–452, January 2015.

[326] Damian Szklarczyk, John H. Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T. Doncheva, Alexander Roth, Peer Bork, Lars J. Jensen, and Christian von Mering. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Research, 45(D1):D362–D368, January 2017.

[327] Dylan J. Taatjes, Anders M. Näär, Frank Andel, Eva Nogales, and Robert Tjian. Structure, function, and activator-induced conformations of the CRSP coactivator. Science (New York, N.Y.), 295(5557):1058–1062, February 2002.

[328] Hidehisa Takahashi, Tari J. Parmely, Shigeo Sato, Chieri Tomomori-Sato, Charles A. S. Banks, Stephanie E. Kong, Henrietta Szutorisz, Selene K. Swanson, Skylar Martin-Brown, Michael P. Washburn, Laurence Florens, Chris W. Seidel, Chengqi Lin, Edwin R. Smith, Ali Shilatifard, Ronald C. Conaway, and Joan W. Conaway. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell, 146(1):92–104, July 2011.

[329] Moshe Talpaz, Richard T. Silver, Brian J. Druker, John M. Goldman, Carlo Gambacorti-Passerini, Francois Guilhot, Charles A. Schiffer, Thomas Fischer, Michael W. N. Deininger, Anne L. Lennard, Andreas Hochhaus, Oliver G. Ottmann, Alois Gratwohl, Michele Baccarani, Richard Stone, Sante Tura, Francois-Xavier Mahon, Sofia Fernandes-Reese, Insa Gathmann, Renaud Capdeville, Hagop M. Kantarjian, and Charles L. Sawyers. Imatinib induces durable hematologic and cytogenetic responses in patients with accelerated phase chronic myeloid leukemia: results of a phase 2 study. Blood, 99(6):1928–1937, March 2002.

[330] Helga Thorvaldsdóttir, James T. Robinson, and Jill P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics, 14(2):178–192, March 2013.

[331] Kazuhito Toyo-oka, Timothy J. Bowen, Shinji Hirotsune, Zirong Li, Sonia Jain, Sara Ota, Laure Escoubet Lozach, Ivan Garcia Bassett, Jean Lozach, Michael G. Rosenfeld, Christopher K. Glass, Robert Eisenman, Bing Ren, Peter Hurlin, and Anthony Wynshaw-Boris. Mnt-Deficient Mammary Glands Exhibit Impaired Involution and Tumors with Characteristics of Myc Overexpression. Cancer Research, 66(11):5565–5573, June 2006.

[332] Kazuhito Toyo-oka, Timothy J. Bowen, Shinji Hirotsune, Zirong Li, Sonia Jain, Sara Ota, Laure Escoubet Lozach, Ivan Garcia Bassett, Jean Lozach, Michael G. Rosenfeld, Christopher K. Glass, Robert Eisenman, Bing Ren, Peter Hurlin, and Anthony Wynshaw-Boris. Mnt-Deficient Mammary Glands Exhibit Impaired Involution and Tumors with Characteristics of Myc Overexpression. Cancer Research, 66(11):5565–5573, June 2006.

[333] Cole Trapnell, Brian A. Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J. van Baren, Steven L. Salzberg, Barbara J. Wold, and Lior Pachter. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nature biotechnology, 28(5):511–515, May 2010.

[334] Ignacio Javier Tripodi and Margaret Ann Gruca. Nascent-Flow. December 2018.

[335] Taiki Tsutsui, Rikiya Fukasawa, Aki Tanaka, Yutaka Hirose, and Yoshiaki Ohkuma. Identification of target genes for the CDK subunits of the Mediator complex. Genes to Cells, 16(12):1208–1218, December 2011.

[336] Taiki Tsutsui, Hiroyasu Umemura, Aki Tanaka, Fumitaka Mizuki, Yutaka Hirose, and Yoshiaki Ohkuma. Human mediator kinase subunit CDK11 plays a negative role in viral activator VP16-dependent transcriptional regulation. Genes to Cells, 13(8):817–826, August 2008.

[337] John W. Tullai, Michael E. Schaffer, Steven Mullenbrock, Gabriel Sholder, Simon Kasif, and Geoffrey M. Cooper. Immediate-early and delayed primary response genes are distinct in function and genomic architecture. The Journal of Biological Chemistry, 282(33):23981–23995, August 2007.

[338] Mikko Turunen, Jason M. Spaeth, Salla Keskitalo, Min Ju Park, Teemu Kivioja, Alison D. Clark, Netta Mäkinen, Fangjian Gao, Kimmo Palin, Helka Nurkkala, Anna Vähärautio, Mervi Aavikko, Kati Kämpjärvi, Pia Vahteristo, Chongwoo A. Kim, Lauri A. Aaltonen, Markku Varjosalo, Jussi Taipale, and Thomas G. Boyer. Uterine leiomyoma-linked MED12 mutations disrupt Mediator-associated CDK activity. Cell reports, 7(3):654–660, May 2014.

[339] Stefka Tyanova, Tikira Temu, and Juergen Cox. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nature Protocols, 11(12):2301, December 2016.

[340] Mathias Uhlen, Anita Bandrowski, Steven Carr, Aled Edwards, Jan Ellenberg, Emma Lundberg, David L. Rimm, Henry Rodriguez, Tara Hiltke, Michael Snyder, and Tadashi Yamamoto. A proposal for validation of antibodies. Nature Methods, 13(10):823–827, October 2016. Number: 10 Publisher: Nature Publishing Group.

[341] Lyubomir T. Vassilev, Binh T. Vu, Bradford Graves, Daisy Carvajal, Frank Podlaski, Zoran Filipovic, Norman Kong, Ursula Kammlott, Christine Lukacs, Christian Klein, Nader Fotouhi, and Emily A. Liu. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science (New York, N.Y.), 303(5659):844–848, February 2004.

[342] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1 0 Contributors. SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. arXiv:1907.10121 [physics], July 2019. arXiv: 1907.10121.

[343] Irene Visintin, Ziding Feng, Gary Longton, David C. Ward, Ayesha B. Alvero, Yinglei Lai, Jeannette Tenthorey, Aliza Leiser, Ruben Flores-Saaib, Herbert Yu, Masoud Azori, Thomas Rutherford, Peter E. Schwartz, and Gil Mor. Diagnostic Markers for Early Detection of Ovarian Cancer. Clinical Cancer Research, 14(4):1065–1072, February 2008.

[344] Tadashi Wada, Toshiyuki Takagi, Yuki Yamaguchi, Anwarul Ferdous, Takeshi Imai, Susumu Hirose, Seiji Sugimoto, Keiichi Yano, Grant A. Hartzog, Fred Winston, Stephen Buratowski, and Hiroshi Handa. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes & Development, 12(3):343–356, February 1998. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[345] Therese Wahlström and Marie Henriksson. Mnt takes control as key regulator of the myc/max/mxd network. Advances in Cancer Research, 97:61–80, 2007.

[346] William Walker, Zi-Qiang Zhou, Sara Ota, Anthony Wynshaw-Boris, and Peter J. Hurlin. Mnt–Max to Myc–Max complex switching regulates cell cycle entry. The Journal of Cell Biology, 169(3):405–413, May 2005.

[347] Dong Wang, Ivan Garcia-Bassets, Chris Benner, Wenbo Li, Xue Su, Yiming Zhou, Jinsong Qiu, Wen Liu, Minna Kaikkonen, Kenny H. Ohgi, Christopher K. Glass, Michael G. Rosenfeld, and Xiang-Dong Fu. Reprogramming Transcription via Distinct Classes of Enhancers Functionally Defined by eRNA. Nature, 474(7351):390–394, May 2011.

[348] Haoyi Wang, Hui Yang, Chikdu S. Shivalila, Meelad M. Dawlaty, Albert W. Cheng, Feng Zhang, and Rudolf Jaenisch. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell, 153(4):910–918, May 2013.

[349] Liguo Wang, Shengqin Wang, and Wei Li. RSeQC: quality control of RNA-seq experiments. Bioinformatics, 28(16):2184–2185, 2012.

[350] Taijin Wang, Zhuang Yang, Yongguang Zhang, Wei Yan, Fang Wang, Linhong He, Yuanyuan Zhou, and Lijuan Chen. Discovery of novel CDK8 inhibitors using multiple crystal structures in docking-based virtual screening. European Journal of Medicinal Chemistry, 129:275–286, February 2017.

[351] Weiping Wang and Jefferson Y. Chan. Nrf1 is targeted to the endoplasmic reticulum membrane by an N-terminal transmembrane domain. Inhibition of nuclear translocation and transacting function. The Journal of Biological Chemistry, 281(28):19676–19687, July 2006.

[352] Harpreet Wasan, Angela M Meade, Richard Adams, Richard Wilson, Cheryl Pugh, David Fisher, Benjamin Sydes, Ayman Madi, Bruce Sizer, Charles Lowdell, Gary Middleton, Rachel Butler, Richard Kaplan, and Tim Maughan. Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. The Lancet Oncology, 15(6):631–639, May 2014.

[353] Brian T. Weinert, Takeo Narita, Shankha Satpathy, Balaji Srinivasan, Bogi K. Hansen, Christian Schölz, William B. Hamilton, Beth E. Zucconi, Wesley W. Wang, Wenshe R. Liu, Joshua M. Brickman, Edward A. Kesicki, Albert Lai, Kenneth D. Bromberg, Philip A. Cole, and Chunaram Choudhary. Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. Cell, 174(1):231–244.e12, June 2018.

[354] Ronald C Wek and Kirk A Staschke. How do tumours adapt to nutrient stress? The EMBO Journal, 29(12):1946–1947, June 2010.

[355] Thomas Westerling, Emilia Kuuluvainen, and Tomi P. Mäkelä. Cdk8 Is Essential for Preimplantation Mouse Development. Molecular and Cellular Biology, 27(17):6177–6182, September 2007.

[356] A. J. Whitmarsh and R. J. Davis. Regulation of transcription factor function by phosphorylation. Cellular and molecular life sciences: CMLS, 57(8-9):1172–1183, August 2000.

[357] Lucy H. Williams, George Fromm, Nolan G. Gokey, Telmo Henriques, Ginger W. Muse, Adam Burkholder, David C. Fargo, Guang Hu, and Karen Adelman. Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. Molecular Cell, 58(2):311–322, April 2015.

[358] Lukas Windhager, Thomas Bonfert, Kaspar Burger, Zsolt Ruzsics, Stefan Krebs, Stefanie Kaufmann, Georg Malterer, Anne L'Hernault, Markus Schilhabel, Stefan Schreiber, Philip Rosenstiel, Ralf Zimmer, Dirk Eick, Caroline C. Friedel, and Lars Dölken. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome Research, 22(10):2031–2042, October 2012. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[359] J. A. Winkles. Serum- and polypeptide growth factor-inducible gene expression in mouse fibroblasts. Progress in Nucleic Acid Research and Molecular Biology, 58:41–78, 1998.

[360] Erin M. Wissink, Anniina Vihervaara, Nathaniel D. Tippens, and John T. Lis. Nascent RNA analyses: tracking transcription and its regulation. Nature Reviews Genetics, 20(12):705–723, December 2019. Number: 12 Publisher: Nature Publishing Group.

[361] Esther Witsch, Michael Sela, and Yosef Yarden. Roles for Growth Factors in Cancer Progression. Physiology (Bethesda, Md.), 25(2):85–101, April 2010.

[362] Lauren W. Wood, Nicole I. Cox, Cody A. Phelps, Shao-Chiang Lai, Arjun Poddar, Conover Talbot Jr, and David Mu. Thyroid Transcription Factor 1 Reprograms Angiogenic Activities of Secretome. Scientific Reports, 6:19857, February 2016.

[363] Carl Wu. The 5 ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature, 286(5776):854–860, August 1980. Number: 5776 Publisher: Nature Publishing Group.

[364] Xueqin Xi, Yanfen Yao, Na Liu, and Pibao Li. MiR-297 alleviates LPS-induced A549 cell and mice lung injury via targeting cyclin dependent kinase 8. International Immunopharmacology, 80:106197, March 2020.

[365] Jianguo Xia and David S. Wishart. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. In Current Protocols in Bioinformatics. John Wiley & Sons, Inc., 2002.

[366] Lan Xie, Amy L. Sullivan, Jana G. Collier, and Christopher K. Glass. Serum Response Factor Indirectly Regulates Type I Interferon-Signaling in Macrophages. Journal of Interferon & Cytokine Research, 33(10):588–596, October 2013.

[367] Wu Xu and Jun-Yuan Ji. Dysregulation of CDK8 and Cyclin C in tumorigenesis. Journal of Genetics and Genomics, 38(10):439–452, October 2011.

[368] Hai-Hui Xue, Julie Bollenbacher-Reilley, Zheng Wu, Rosanne Spolski, Xuefang Jing, Yi-Chen Zhang, J. Philip McCoy, and Warren J. Leonard. The Transcription Factor GABP Is a Critical Regulator of B Lymphocyte Development. Immunity, 26(4):421–431, April 2007.

[369] Yuki Yamaguchi, Toshiyuki Takagi, Tadashi Wada, Keiichi Yano, Akiko Furuya, Seiji Sugimoto, Jun Hasegawa, and Hiroshi Handa. NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. Cell, 97(1):41–51, April 1999.

[370] Seiji Yamamoto, Tomoko Hagihara, Yoshiyuki Horiuchi, Akira Okui, Shotaro Wani, Tokuyuki Yoshida, Takao Inoue, Aki Tanaka, Takashi Ito, Yutaka Hirose, and Yoshiaki Ohkuma. Mediator cyclin-dependent kinases upregulate transcription of inflammatory genes in cooperation with NF-B and C/EBP on stimulation of Toll-like receptor 9. Genes to Cells, pages n/a–n/a, January 2017.

[371] Hui Yang, Haoyi Wang, Chikdu S. Shivalila, Albert W. Cheng, Linyu Shi, and Rudolf Jaenisch. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell, 154(6):1370–1379, December 2013.

[372] Jiangbin Ye, Monika Kumanova, Lori S Hart, Kelly Sloane, Haiyan Zhang, Diego N De Panis, Ekaterina Bobrovnikova-Marjon, J Alan Diehl, David Ron, and Constantinos Koumenis. The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation. The EMBO Journal, 29(12):2082–2096, June 2010.

[373] Tyler Zarubin and Jiahuai Han. Activation and signaling of the p38 MAP kinase pathway. Cell Research, 15(1):11–18, January 2005.

[374] Julia Zeitlinger, Alexander Stark, Manolis Kellis, Joung-Woo Hong, Sergei Nechaev, Karen Adelman, Michael Levine, and Richard A. Young. RNA Polymerase Stalling at Developmental Control Genes in the Drosophila Embryo. Nature genetics, 39(12):1512–1516, December 2007.

[375] Huimin Zhang, Frank Rigo, and Harold G. Martinson. Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Molecular Cell, 59(3):437–448, August 2015.

[376] Jun-feng Zhang, Jian-shui Zhang, Zhao-hua Zhao, Peng-bo Yang, Sheng-feng Ji, Nan Li, Qin-dong Shi, Jing Tan, Xi Xu, Cang-bao Xu, and Ling-yu Zhao. MicroRNA-770 affects proliferation and cell cycle transition by directly targeting CDK8 in glioma. Cancer Cell International, 18(1):195, December 2018.

[377] Wei Zhang and Hui Tu Liu. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Research, 12(1):9–18, 2002.

[378] Xin-Chun Zhang, Hong-Feng Liang, Xiao-Dong Luo, Hua-Jun Wang, Ai-Ping Gu, Chun-Ye Zheng, Qiao-Zhen Su, and Jun Cai. YY1 promotes IL-6 expression in LPS-stimulated BV2 microglial cells by interacting with p65 to promote transcriptional activation of IL-6. Biochemical and Biophysical Research Communications, 502(2):269–275, 2018.

[379] Xiaoping Zhao, Daorong Feng, Qun Wang, Arian Abdulla, Xiao-Jun Xie, Jie Zhou, Yan Sun, Ellen S. Yang, Lu-Ping Liu, Bhavapriya Vaitheesvaran, Lauren Bridges, Irwin J. Kurland, Randy Strich, Jian-Quan Ni, Chenguang Wang, Johan Ericsson, Jeffrey E. Pessin, Jun-Yuan Ji, and Fajun Yang. Regulation of lipogenesis by cyclin-dependent kinase 8–mediated control of SREBP-1. The Journal of Clinical Investigation, 122(7):2417–2427, July 2012.

[380] Qianhe Zhou, Adnan Derti, David Ruddy, Daniel Rakiec, Iris Kao, Michelle Lira, Veronica Gibaja, HoMan Chan, Yi Yang, Junxia Min, Michael R. Schlabach, and Frank Stegmeier. A Chemical Genetics Approach for the Functional Assessment of Novel Cancer Genes. Cancer Research, 75(10):1949–1958, May 2015.

[381] C.-Y. Zhu, C.-Y. Li, Y. Li, Y.-Q. Zhan, Y.-H. Li, C.-W. Xu, W.-X. Xu, H. B. Sun, and X.-M. Yang. Cell growth suppression by thanatos-associated protein 11(THAP11) is mediated by transcriptional downregulation of c-Myc. Cell Death & Differentiation, 16(3):395–405, March 2009. Number: 3 Publisher: Nature Publishing Group.

[382] Ulf Andersson Ørom, Thomas Derrien, Malte Beringer, Kiranmai Gumireddy, Alessandro Gardini, Giovanni Bussotti, Fan Lai, Matthias Zytnicki, Cedric Notredame, Qihong Huang, Roderic Guigo, and Ramin Shiekhattar. Long non-coding RNAs with enhancer-like function in human. Cell, 143(1):46–58, October 2010.

Appendix A

Mediator Kinases within IFN-γ Supplementary Material

Description: What follows is the supplementary material that accompanied the main text of the Steinparzer 2018 manuscript. Within it are some figures that I generated including raw data tracks of genes exhibiting pausing (Figure S2B), eRNA transcription (Figure S3), and various GSEA analyses (Figure S2F and Figure S5F,G).

Molecular Cell, Volume *76*

Supplemental Information

Transcriptional Responses to IFN-γ Require Mediator Kinase-Dependent Pause Release and Mechanistically Distinct CDK8 and CDK19 Functions

Iris Steinparzer, Vitaly Sedlyarov, Jonathan D. Rubin, Kevin Eislmayr, Matthew D. Galbraith, Cecilia B. Levandowski, Terezia Vcelkova, Lucy Sneezum, Florian Wascher, Fabian Amman, Renata Kleinova, Heather Bender, Zdenek Andrysik, Joaquin M. Espinosa, Giulio Superti-Furga, Robin D. Dowell, Dylan J. Taatjes, and Pavel Kovarik

Figure S1

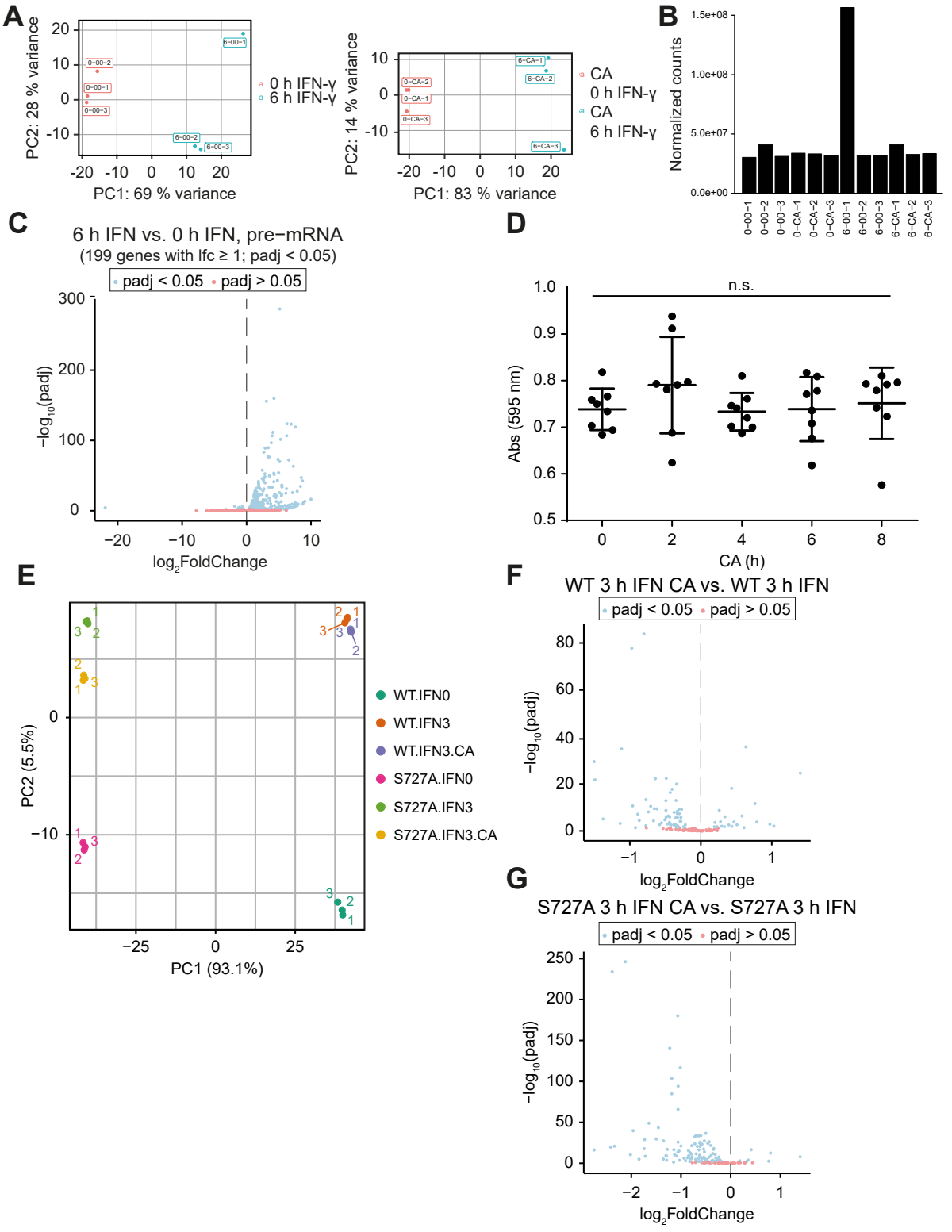

A
PC2: 28 % variance
PC1: 69 % variance
0 h IFN-γ
6 h IFN-γ
PC2: 14 % variance
PC1: 83 % variance
CA
0 h IFN-γ
CA
6 h IFN-γ
B
Normalized counts
0–00–1
0–00–2
0–00–3
0–CA–1
0–CA–2
0–CA–3
6–00–1
6–00–2
6–00–3
6–CA–1
6–CA–2
6–CA–3
C
6 h IFN vs. 0 h IFN, pre−mRNA
(199 genes with lfc ≥ 1; padj < 0.05)
padj < 0.05
padj > 0.05
−log10(padj)
log2FoldChange
D
n.s.
Abs (595 nm)
CA (h)
E
PC2 (5.5%)
PC1 (93.1%)
WT.IFN0
WT.IFN3
WT.IFN3.CA
S727A.IFN0
S727A.IFN3
S727A.IFN3.CA
F
WT 3 h IFN CA vs. WT 3 h IFN
padj < 0.05
padj > 0.05
−log10(padj)
log2FoldChange
G
S727A 3 h IFN CA vs. S727A 3 h IFN
padj < 0.05
padj > 0.05
−log10(padj)
log2FoldChange

Figure S1 (related to Figure 1). Mediator kinase inhibition impairs IFN-γ-induced gene expression independently of STAT1 S727 phosphorylation

(**A**) Principal component analysis (PCA) of RNA-Seq experiment described in Figure 1A-E (WT MEFs, 0 or 6 h IFN-γ, with or without 1 h CA pre-treatment). Left panel: PCA of three biological IFN-γ-stimulated (6-00-1, 6-00-2, 6-00-3) and unstimulated (0-00-1, 0-00-2, 0-00-3) replicates. Right panel: PCA of three biological CA-pre-treated IFN-γ-stimulated (6-CA-1, 6-CA-2, 6-CA-3) and unstimulated (0-CA-1, 0-CA-2, 0-CA-3) replicates.

(**B**) Read counts normalized to library size and composition. Y-axis: The sum of all gene counts per sample normalized to library size and composition. X-axis: Samples. Sample 6-00-1 was thereby identified as outlier and removed from subsequent analysis.

(**C**) Gene expression (pre-mRNA level) changes in WT MEFs upon 6 h IFN-γ treatment (blue, padj < 0.05; red, padj > 0.05). Genes with padj < 0.05, $\log_2$FoldChange (lfc) ≥ 1, FPKM in IFN-γ stimulated samples ≥ 1 were regarded as IFN-γ-induced (199 genes).

(**D**) CA has no effects on viability of MEFs in relevant time window. WT MEFs were treated for 2, 4, 6 and 8 h with CA or left untreated followed by crystal violet staining and absorbance measurements at 595 nm (n=8 per time point).

(**E**) Principal component analysis (PCA) of RNA-Seq experiment described in Figure 1F-H (WT and S727A MEFs, 0 or 3 h IFN-γ, with or without 1 h CA pre-treatment). PCA is shown for three biological replicates (replicates 1, 2, and 3) for each treatment.

(**F** and **G**) Effects of CA on expression of IFN-γ-induced gene expression (mRNA level) in WT (F) and S727A (G) MEFs. Cells (± 100 nM CA, 1 h pre-treatment) were stimulated with IFN-γ (3 h) followed by RNA-Seq and differential expression analyses for effects of CA (blue, padj < 0.05; red, padj > 0.05).

Figure S2

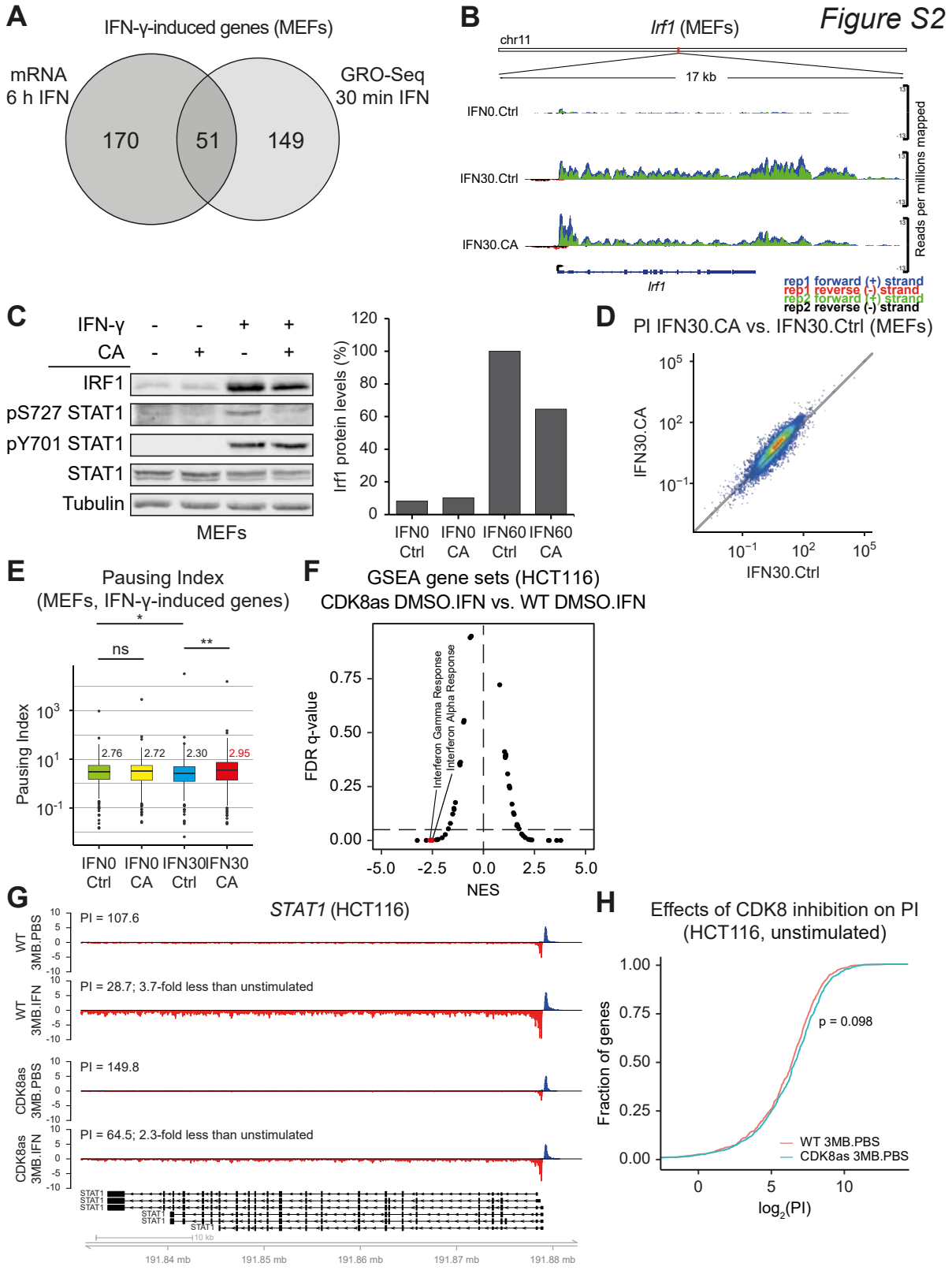

A
IFN-γ-induced genes (MEFs)
mRNA
6 h IFN
GRO-Seq
30 min IFN
170
51
149
B
Irf1 (MEFs)
chr11
17 kb
IFN0.Ctrl
IFN30.Ctrl
IFN30.CA
Irf1
Reads per millions mapped
rep1 forward (+) strand
rep1 reverse (-) strand
rep2 forward (+) strand
rep2 reverse (-) strand
C
IFN-γ - - + +
CA - + - +
IRF1
pS727 STAT1
pY701 STAT1
STAT1
Tubulin
MEFs
Irf1 protein levels (%)
IFN0 Ctrl
IFN0 CA
IFN60 Ctrl
IFN60 CA
D
PI IFN30.CA vs. IFN30.Ctrl (MEFs)
IFN30.CA
IFN30.Ctrl
E
Pausing Index
(MEFs, IFN-γ-induced genes)
Pausing Index
ns
*
**
2.76
2.72
2.30
2.95
IFN0 Ctrl
IFN0 CA
IFN30 Ctrl
IFN30 CA
F
GSEA gene sets (HCT116)
CDK8as DMSO.IFN vs. WT DMSO.IFN
FDR q-value
Interferon Gamma Response
Interferon Alpha Response
NES
G
STAT1 (HCT116)
WT 3MB.PBS
PI = 107.6
WT 3MB.IFN
PI = 28.7; 3.7-fold less than unstimulated
CDK8as 3MB.PBS
PI = 149.8
CDK8as 3MB.IFN
PI = 64.5; 2.3-fold less than unstimulated
STAT1
191.84 mb
191.85 mb
191.86 mb
191.87 mb
191.88 mb
H
Effects of CDK8 inhibition on PI
(HCT116, unstimulated)
Fraction of genes
p = 0.098
WT 3MB.PBS
CDK8as 3MB.PBS
log2(PI)

Figure S2 (related to Figure 2). Mediator kinase inhibition increases RNAPII pausing at IFN-γ-induced genes in MEFs and human HCT116 cells

(**A** – **E**) GRO-Seq analysis in MEFs (relates to Figure 2A-F).

(**A**) Overlap of genes induced upon IFN-γ stimulation for 6 h (mRNA, Figure 1B) and 30 min (GRO-Seq).

(**B**) Overlay (Genome browser view) of replicates of GRO-Seq reads at *Irf1* locus for IFN0.Ctrl, IFN30.Ctrl and IFN30.CA conditions.

(**C**) Western blot (left) showing effect of CA on induction of IRF1 by IFN-γ (60 min). Right: quantitative evaluation of Western blot signal.

(**D**) Density plot of PI for IFN-γ-stimulated MEFs ± CA. Dot density: red, high; blue, low.

(**E**) Median PI and statistical assessment of PI changes for IFN-γ-induced genes (lfc > 0 and padj < 0.05, 200 genes). Median PI value shown for each condition (red = highest PI value). Mann-Whitney U test, p-value ns ≥ 0.05; * < 0.05; ** < 0.01. Note: statistically significant IFN-γ-mediated drop in PI from 2.76 (green) to 2.30 (blue), consistent with the read count distribution analysis at *Irf1* in (B); inhibition of the IFN-γ-mediated drop in PI by CA treatment (2.76, green vs. 2.95, red); no effect of CA treatment in cells not stimulated with IFN-γ (2.76 , green vs. 2.72, yellow).

(**F** – **H**) PRO-Seq analysis in HCT116 cells (relates to Figure 2G-J).

(**F**) GSEA plot of CDK8as effects on IFN-γ-induced gene set changes as compared to WT cells (CDK8as DMSO.IFN vs. WT DMSO.IFN).

(**G**) Genome browser view of PRO-Seq reads (pooled replicates) at *STAT1* locus in WT and CSK8as cells treated with 3MB-PP1 ± IFN-γ stimulation. Red: forward strand, blue: reverse strand. PI values are indicated.

(**H**) PI distribution upon CDK8 inhibition (CDK8as 3MB.PBS, blue) and not inhibited control (WT 3MB.PBS, red) for same genes as in Figure 2J.

Figure S3

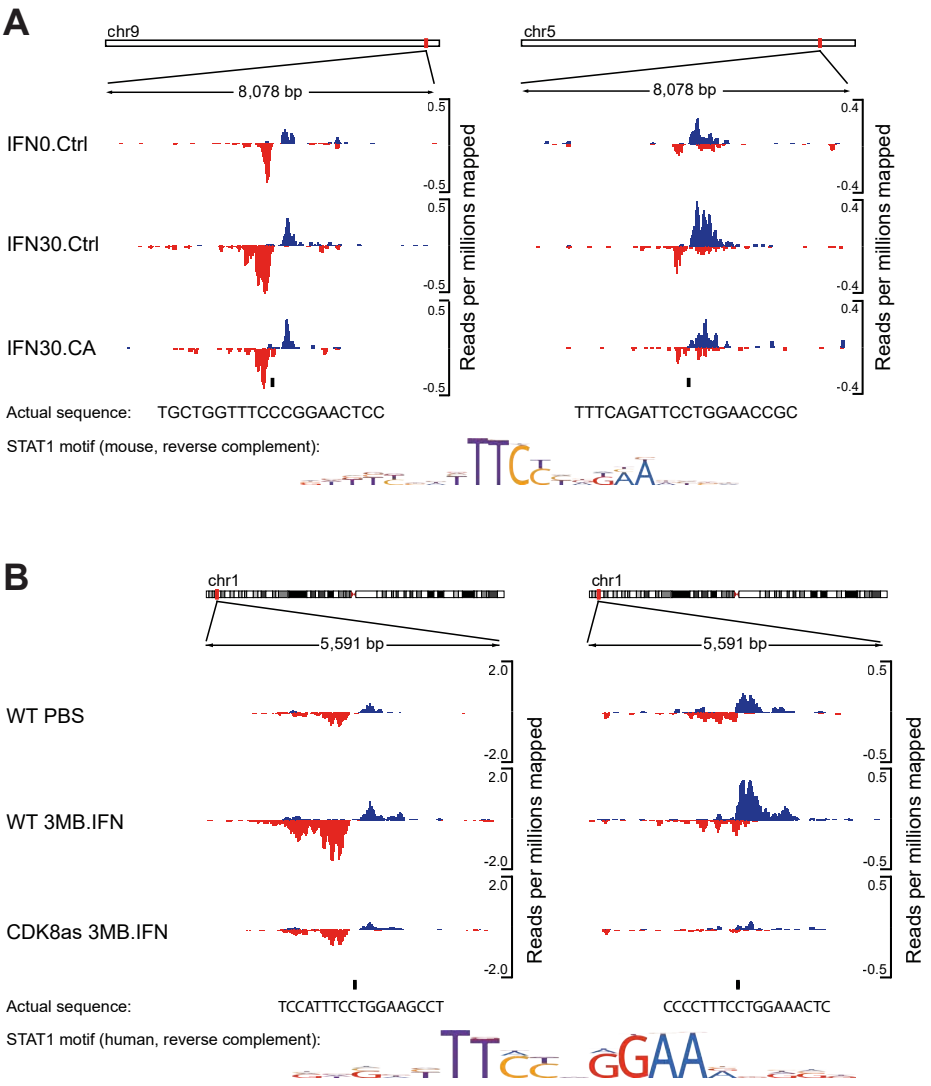

A
chr9
8,078 bp
chr5
8,078 bp
IFN0.Ctrl
IFN30.Ctrl
IFN30.CA
Reads per millions mapped
Actual sequence: TGCTGGTTTCCCGGAACTCC
TTTCAGATTCCTGGAACCGC
STAT1 motif (mouse, reverse complement):
B
chr1
5,591 bp
chr1
5,591 bp
WT PBS
WT 3MB.IFN
CDK8as 3MB.IFN
Reads per millions mapped
Actual sequence: TCCATTTCCTGGAAGCCT
CCCCTTTCCTGGAAACTC
STAT1 motif (human, reverse complement):

Figure S3 (related to Figure 3). eRNA counts implicate Mediator kinases in eRNA transcription during IFN-γ response

(**A**) eRNA traces at two loci in unstimulated (IFN0.Ctrl), IFN-γ-stimulated (IFN30.Ctrl) and IFN-γ-stimulated CA-treated (IFN30.CA) MEFs. The sequence of the locus, TF motif annotation for the sequence (STAT1) and STAT1 motif logo (mouse) are shown.

(**B**) eRNA traces at two loci in chromosome 1 in unstimulated (WT.PBS) and 3MB-PP1-treated IFN-γ-stimulated (WT 3MB.IFN) WT HCT116 cells, and 3MB-PP1-treated IFN-γ-stimulated CDK8as HCT116 cells (CDK8as 3MB.IFN). The sequence of the locus, TF motif annotation for the sequence (STAT1) and STAT1 motif logo (human) are shown.

Figure S4

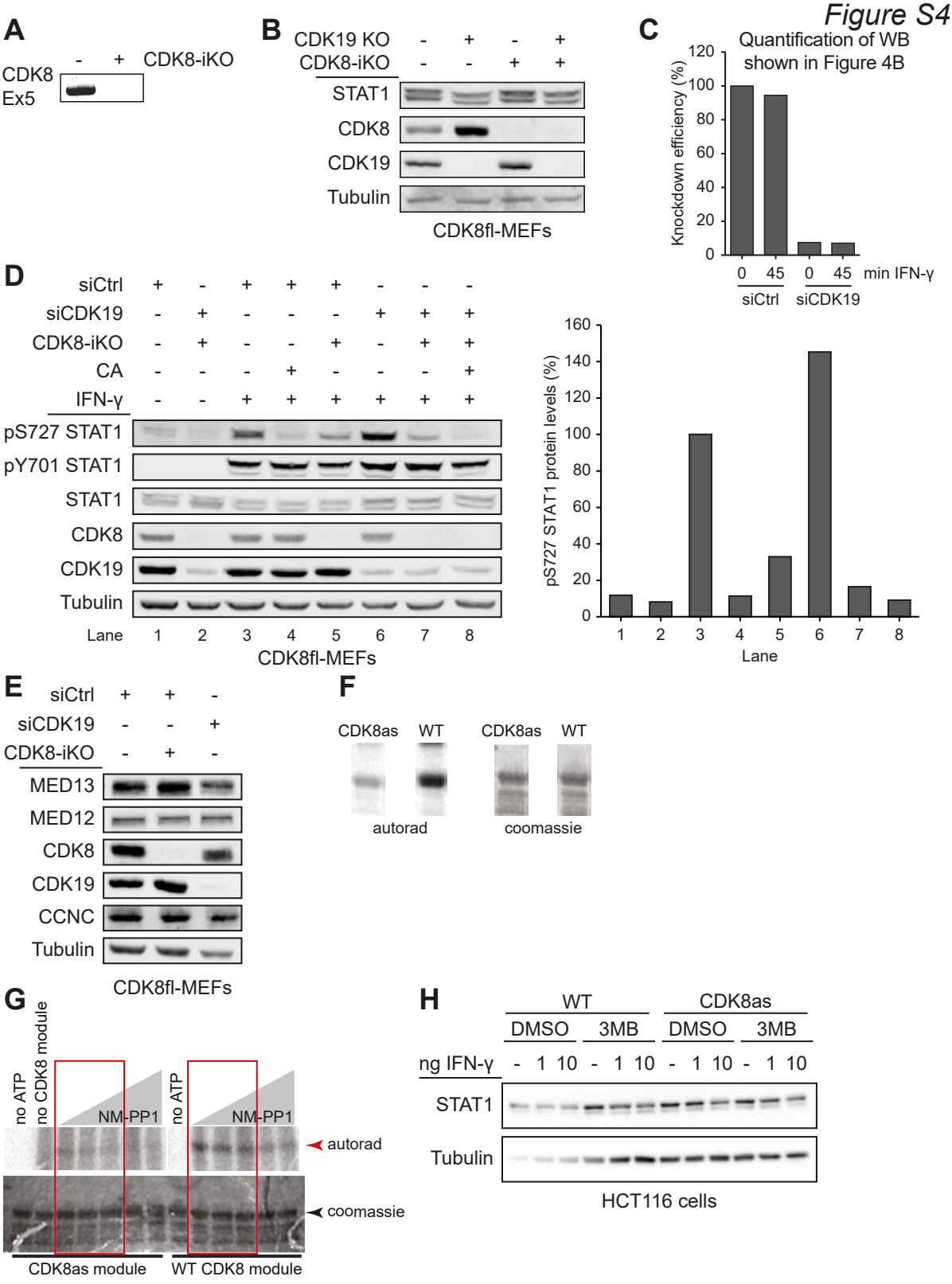

A
CDK8 Ex5
- + CDK8-iKO
B
CDK19 KO - + - +
CDK8-iKO - - + +
STAT1
CDK8
CDK19
Tubulin
CDK8fl-MEFs
C
Quantification of WB shown in Figure 4B
Knockdown efficiency (%)
0 45 0 45 min IFN-γ
siCtrl siCDK19
D
siCtrl + - + + + - - -
siCDK19 - + - - - + + +
CDK8-iKO - + - - + - + +
CA - - - + - - - +
IFN-γ - - + + + + + +
pS727 STAT1
pY701 STAT1
STAT1
CDK8
CDK19
Tubulin
Lane 1 2 3 4 5 6 7 8
CDK8fl-MEFs
pS727 STAT1 protein levels (%)
Lane
E
siCtrl + + -
siCDK19 - - +
CDK8-iKO - + -
MED13
MED12
CDK8
CDK19
CCNC
Tubulin
CDK8fl-MEFs
F
CDK8as WT CDK8as WT
autorad
coomassie
G
no ATP
no CDK8 module
NM-PP1
no ATP
NM-PP1
autorad
coomassie
CDK8as module
WT CDK8 module
H
WT CDK8as
DMSO 3MB DMSO 3MB
ng IFN-γ - 1 10 - 1 10 - 1 10 - 1 10
STAT1
Tubulin
HCT116 cells

Figure S4 (related to Figure 4). Supporting evidence for the role of CDK8, not CDK19, in IFN-γ-induced STAT1 AD phosphorylation at S727

(**A**) PCR showing inducible CDK8 exon 5 deletion (CDK8-iKO) in CDK8fl-MEFs (regime described in Figure 4A).

(**B**) Constitutive deletion of CDK19 in clonally selected cells can cause upregulation of CDK8. CDK19 was deleted in CDK8fl-MEFs using CRISPR/Cas9, and CDK19-KO cells were clonally selected followed by Western blot analysis using STAT1, CDK8, CDK19 and tubulin antibodies. CDK8 deletion was induced in CDK19 KO and parental CDK8fl-MEFs in lanes 3 and 4. Note that CDK8 was upregulated in CDK19 KO compared to CDK8fl-MEFs (lane 2 vs. 1).

(**C**) Quantitation (using Bio-Rad Image Lab) of Western blot shown in Figure 4B to demonstrate >90% CDK19 knockdown efficiency. CDK19 signal (normalized to tubulin) is shown as percentage of control (siCtrl, 0 IFN-γ).

(**D**) A second representative experiment demonstrating a key role of CDK8, not CDK19, in IFN-γ-induced STAT1 S727 phosphorylation. Left panel shows Western blot analysis, right panel shows quantitation of the Western blot. The setting was the same as described in Figure 1C. Similar to Figure 1C, siCDK19 did not cause reduction of IFN-γ-induced STAT1 S727 phosphorylation (lane 6 vs. 3).

(**E**) Mediator kinase module subunits are not affected by CDK8-iKO or siCDK19 as revealed by immunoblotting for MED12, MED13, CDK8, CDK19, and CCNC.

(**F** and **G**) In vitro kinase assays showing STAT1 phosphorylation by WT CDK8 and analog-sensitive CDK8 (CDK8as) (i.e. F97G mutant). Purified kinase modules containing WT CDK8 or CDK8as were incubated with STAT1 substrate together with [γ-^{32}P]-ATP (F) or [γ-^{32}P]-ATP in the presence of the ATP analog NM-PP1 (1-100 μM) (G) and subsequently analyzed by autoradiography and coomassie staining of gels. Note that lower NM-PP1 concentrations inhibited CDK8as but not WT CDK8 (red frames); higher NM-PP1 concentrations competed with ATP thus inhibited also WT CDK8.

(**H**) Loading control for STAT1 levels corresponding to Western blot shown in Figure 4E (HCT116 WT and CDK8as cells).

Figure S5

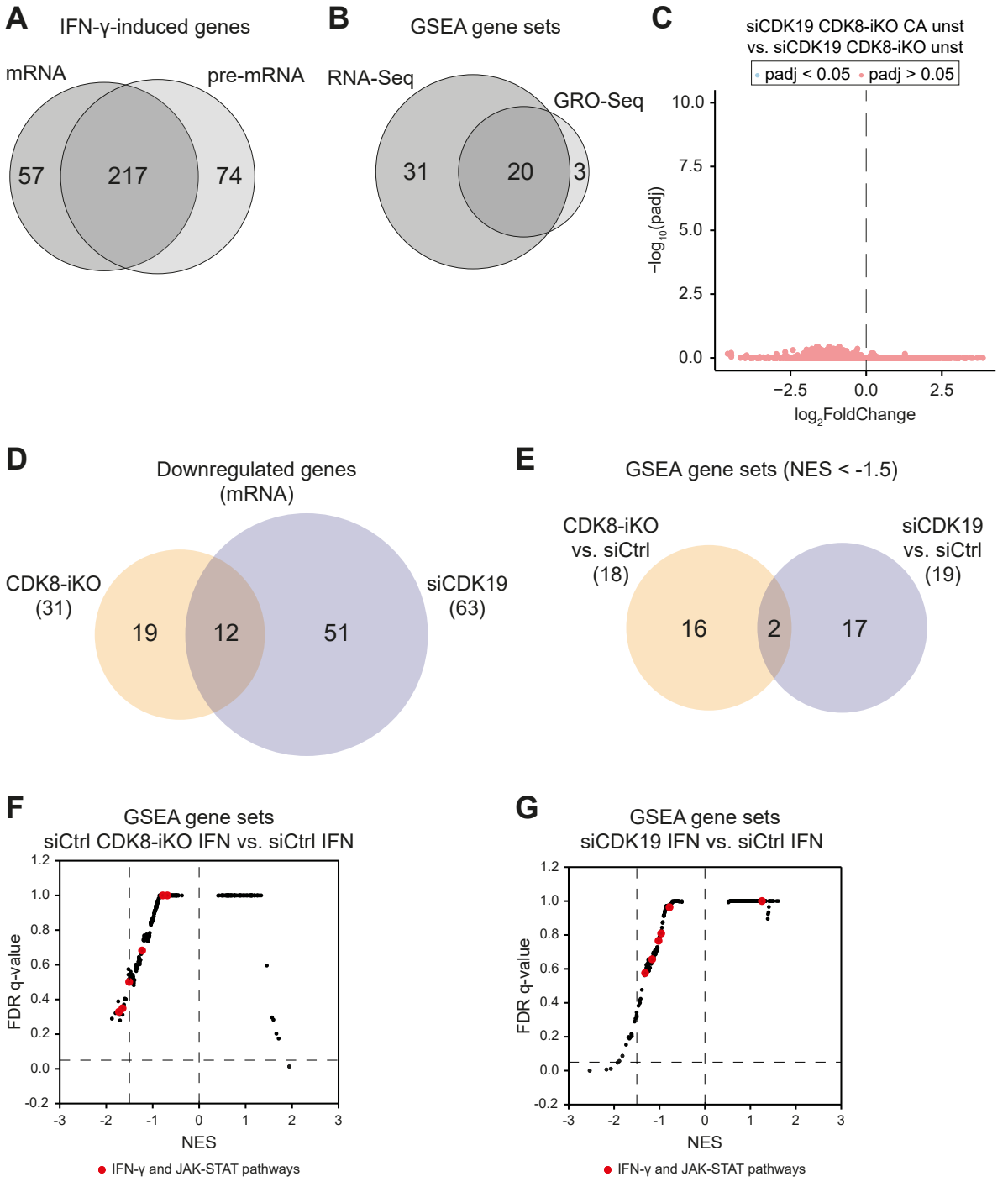

A
IFN-γ-induced genes
mRNA
pre-mRNA
57
217
74
B
GSEA gene sets
RNA-Seq
GRO-Seq
31
20
3
C
siCDK19 CDK8-iKO CA unst
vs. siCDK19 CDK8-iKO unst
padj < 0.05
padj > 0.05
−log10(padj)
log2FoldChange
D
Downregulated genes
(mRNA)
CDK8-iKO
(31)
siCDK19
(63)
19
12
51
E
GSEA gene sets (NES < -1.5)
CDK8-iKO
vs. siCtrl
(18)
siCDK19
vs. siCtrl
(19)
16
2
17
F
GSEA gene sets
siCtrl CDK8-iKO IFN vs. siCtrl IFN
FDR q-value
NES
IFN-γ and JAK-STAT pathways
G
GSEA gene sets
siCDK19 IFN vs. siCtrl IFN
FDR q-value
NES
IFN-γ and JAK-STAT pathways

Figure S5 (related to Figure 5 and 6). CDK8 and CDK19 have distinct effects on transcriptional response to IFN-γ

(**A**) Overlap of genes induced in siCtrl cells after 3 h of IFN-γ stimulation (siCtrl 3 h IFN vs. siCtrl 0 h IFN) at mRNA vs. pre-mRNA levels (lfc ≥ 1; padj < 0.05; FPKM in IFN-γ stimulated samples ≥ 1).

(**B**) GSEA overlap between IFN-γ-induced gene set changes at mRNA level (Figure 5C) and in GRO-Seq (Figure 2A).

(**C**) CA effects on gene expression changes (mRNA level) in the absence of CDK8 and CDK19 in cells not stimulated with IFN-γ (siCDK19 CDK8-iKO CA vs. siCDK19 CDK8-iKO). Note that no genes significantly (blue) up- or down-regulated by CA were found, similar to IFN-γ-stimulated cells (Figure 5F).

(**D**) Overlap of IFN-γ-induced genes downregulated by CDK8 knockout (CDK8-iKO) vs. CDK19 knockdown (siCDK19). Note that only minor fraction of downregulated genes overlap between the two conditions.

(**E**) GSEA overlap between gene set changes caused by CDK8 knockout (CDK8-iKO vs. siCtrl) and CDK19 knockdown (siCDK19 vs. siCtrl) during IFN-γ response. Note that only two gene sets are similarly regulated by CDK8 and CDK19.

(**F**) GSEA plot of gene set changes caused by CDK8 knockout in cells stimulated with IFN-γ (siCtrl CDK8-iKO IFN vs. siCtrl IFN).

(**G**) GSEA plot of gene set changes caused by CDK19 knockdown in cells stimulated with IFN-γ (siCDK19 IFN vs. siCtrl IFN).

Figure S6

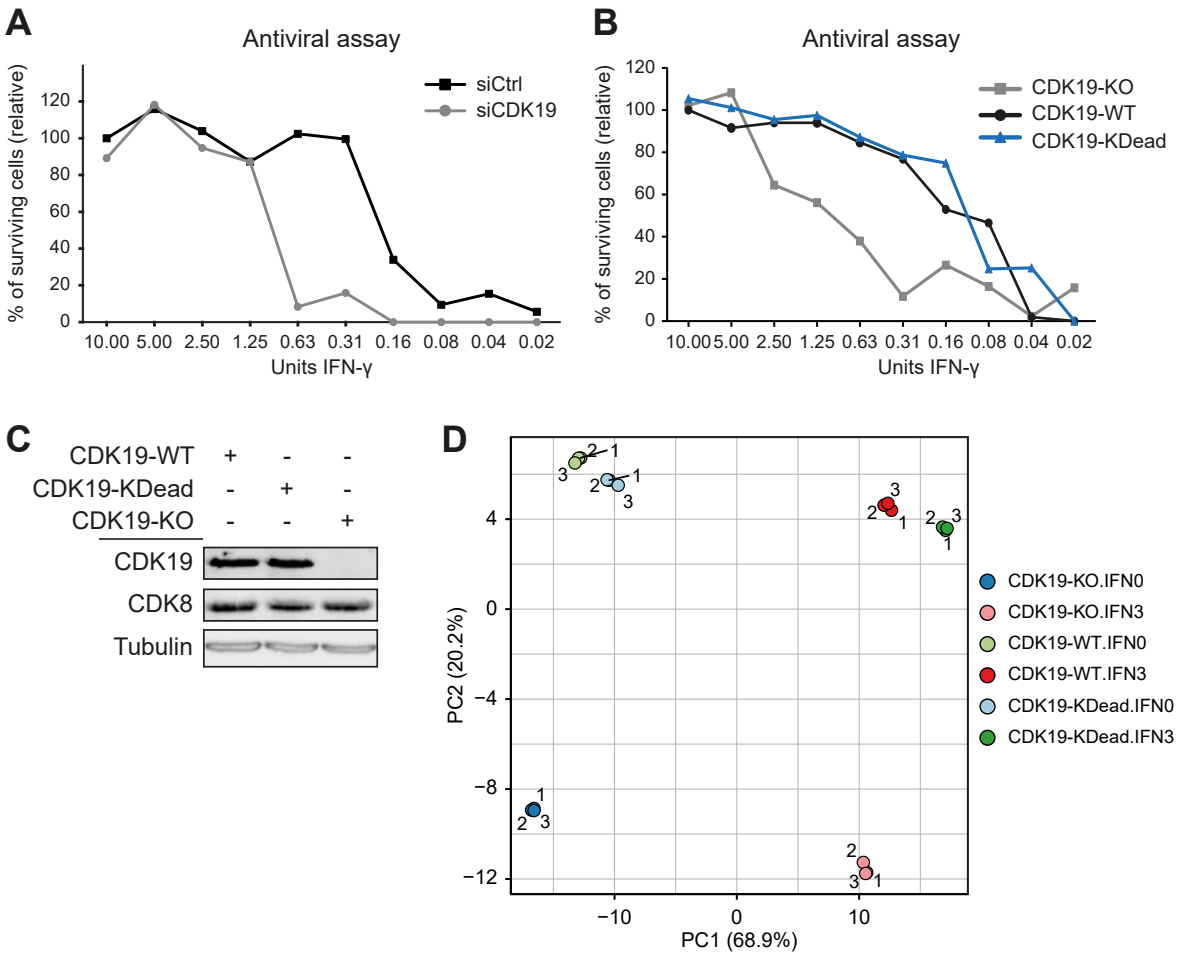

A
Antiviral assay
% of surviving cells (relative)
siCtrl
siCDK19
Units IFN-γ
B
Antiviral assay
% of surviving cells (relative)
CDK19-KO
CDK19-WT
CDK19-KDead
Units IFN-γ
C
CDK19-WT
CDK19-KDead
CDK19-KO
CDK19
CDK8
Tubulin
D
PC2 (20.2%)
PC1 (68.9%)
CDK19-KO.IFN0
CDK19-KO.IFN3
CDK19-WT.IFN0
CDK19-WT.IFN3
CDK19-KDead.IFN0
CDK19-KDead.IFN3

Figure S6 (related to Figure 7). Supporting evidence for kinase-independent role of CDK19 in IFN-γ-responses

(**A**) Assessment of IFN-γ-dependent antiviral response in the absence of CDK19. The panel corresponds to Figure 7A and shows percentages of surviving cells.

(**B**) Assessment of IFN-γ-dependent antiviral response of CDK19-KO cells rescued with CDK19-WT or CDK19-KDead. The panel corresponds to Figure 7B and shows percentages of surviving cells.

(**C**) Western blot showing expression levels of CDK19-WT and CDK19-KDead proteins in rescue CDK19-KO cells. CDK8 protein levels were not affected in CDK19-KO cells and rescued cell pools.

(**D**) Principal component analysis (PCA) of RNA-Seq experiment described in Figure 7C and D. PCA is shown for three biological replicates (replicates 1, 2, and 3) for each condition (CDK19-KO, CDK19-WT and CDK19-KDead, each stimulated with IFN-γ for 3 h).

Appendix B

TFEA Supplementary Material

Description: What follows is the supplementary material that accompanied the main text of the TFEA manuscript. Within it are all datasets used, some additional diagrams detailing *muMerge* and the tests performed on *muMerge*, GC-Correction methods for TFEA, runtime and memory performance of TFEA, and some additional analyses performed.

Supplementary Material

Transcription factor enrichment analysis (TFEA): Quantifying the activity of hundreds of TFs from a single experiment

Jonathan D. Rubin, Jacob T. Stanely, Rutendo F. Sigauke,
Cecilia B. Levandowski, Zachary M. Maas, Dylan J. Taatjes,
Robin D. Dowell

[1] Department of Biochemistry, University of Colorado, Boulder CO 80309

[2] BioFrontiers Institute, University of Colorado, Boulder CO 80309

[3] Computational Bioscience Program, Anschutz Medical Campus, University of Colorado, Aurora, CO 80045

[4] Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder CO 80309

[5] Department of Computer Science, University of Colorado, Boulder CO 80309

∗ Corresponding author: robin.dowell@colorado.edu

1 Accession Tables

1.1 Figure 3 Accession Table

Target	Treatment	Data Type	Cell Type	Accession
Nascent RNA	DMSO 1hr	GRO-seq	HCT116	SRR1105736
Nascent RNA	DMSO 1hr	GRO-seq	HCT116	SRR1105737
Nascent RNA	Nutlin 1hr	GRO-seq	HCT116	SRR1105738
Nascent RNA	Nutlin 1hr	GRO-seq	HCT116	SRR1105739

Supplemental Table 1: Accession numbers used in Figure 3, data from [1].

1.2 Figure 5 Accession Table

Target	Treatment	Data Type	Cell Type	Accession
Capped RNA	LPS 0hr	CAGE	Macrophage donor2	12796-136F6
Capped RNA	LPS 0.25hr	CAGE	Macrophage donor2	12797-136F7
Capped RNA	LPS 0.5hr	CAGE	Macrophage donor2	12798-136F8
Capped RNA	LPS 0.75hr	CAGE	Macrophage donor2	12799-136F9
Capped RNA	LPS 1hr	CAGE	Macrophage donor2	12800-136G1
Capped RNA	LPS 1.3hr	CAGE	Macrophage donor2	12801-136G2
Capped RNA	LPS 2hr	CAGE	Macrophage donor2	12803-136G4
Capped RNA	LPS 2.5hr	CAGE	Macrophage donor2	12804-136G5
Capped RNA	LPS 3hr	CAGE	Macrophage donor2	12805-136G6
Capped RNA	LPS 3.5hr	CAGE	Macrophage donor2	12806-136G7
Capped RNA	LPS 4hr	CAGE	Macrophage donor2	12807-136G8
Capped RNA	LPS 5hr	CAGE	Macrophage donor2	12808-136G9
Capped RNA	LPS 8hr	CAGE	Macrophage donor2	12811-136H3
Capped RNA	LPS 10hr	CAGE	Macrophage donor2	12812-136H4
Capped RNA	LPS 12hr	CAGE	Macrophage donor2	12813-136H5
Capped RNA	LPS 14hr	CAGE	Macrophage donor2	12814-136H6
Capped RNA	LPS 16hr	CAGE	Macrophage donor2	12815-136H7
Capped RNA	LPS 18hr	CAGE	Macrophage donor2	12816-136H8
Capped RNA	LPS 20hr	CAGE	Macrophage donor2	12817-136H9
Capped RNA	LPS 22hr	CAGE	Macrophage donor2	12818-136I1
Capped RNA	LPS 24hr	CAGE	Macrophage donor2	12819-136I2
Capped RNA	LPS 36hr	CAGE	Macrophage donor2	12820-136I3
Capped RNA	LPS 48hr	CAGE	Macrophage donor2	12821-136I4

Supplemental Table 2: Project numbers used in Figure 5, data from [6, 4].

1.3 Figure 6 Accession Table

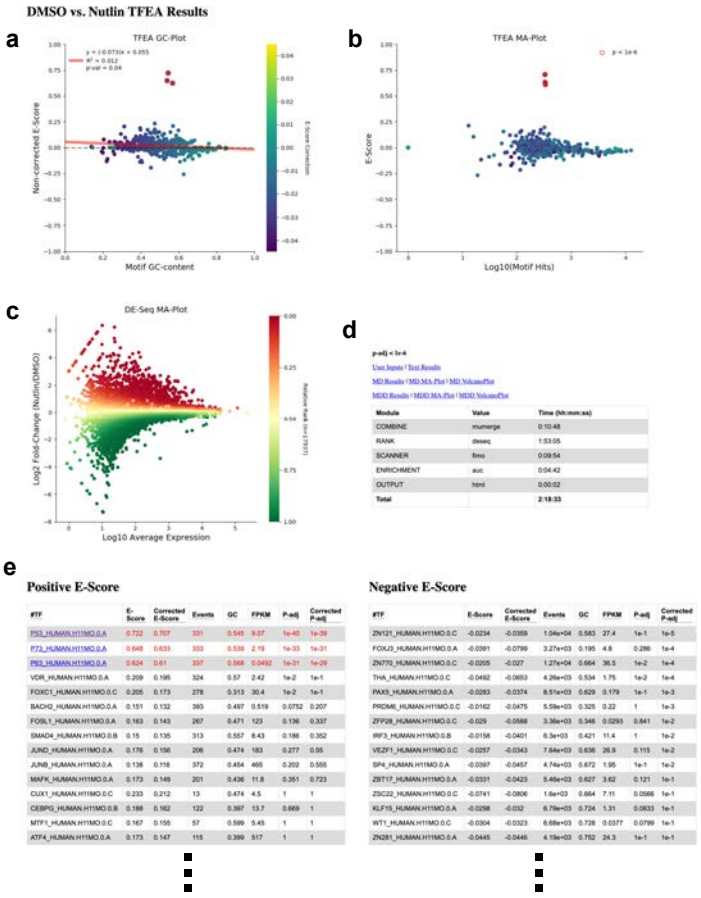
Target	Treatment	Data Type	Cell Type	Accession
ATAC	Dex 0hr	ATAC-seq	A549	ENCSR220ASC
ATAC	Dex 1hr	ATAC-seq	A549	ENCSR139OYS
ATAC	Dex 4hr	ATAC-seq	A549	ENCSR288YMH
ATAC	Dex 8hr	ATAC-seq	A549	ENCSR074AHH

ATAC	Dex 12hr	ATAC-seq	A549	ENCSR265ZXX
DNase	Dex 0hr	DNase-seq	A549	ENCSR136DNA
DNase	Dex 30min	DNase-seq	A549	ENCSR406EMB
DNase	Dex 1hr	DNase-seq	A549	ENCSR384KCZ
DNase	Dex 2hr	DNase-seq	A549	ENCSR837VHE
DNase	Dex 3hr	DNase-seq	A549	ENCSR294XUZ
DNase	Dex 4hr	DNase-seq	A549	ENCSR599WJC
DNase	Dex 5hr	DNase-seq	A549	ENCSR565WPR
DNase	Dex 6hr	DNase-seq	A549	ENCSR077EYC
DNase	Dex 7hr	DNase-seq	A549	ENCSR347CEH
DNase	Dex 8hr	DNase-seq	A549	ENCSR660OQE
DNase	Dex 10hr	DNase-seq	A549	ENCSR128IVG
DNase	Dex 12hr	DNase-seq	A549	ENCSR523FJT
EP300	Dex 0hr	ChIP-seq	A549	ENCSR886OEO
EP300	Dex 5min	ChIP-seq	A549	ENCSR602BTS
EP300	Dex 10min	ChIP-seq	A549	ENCSR174FJD
EP300	Dex 15min	ChIP-seq	A549	ENCSR788VKG
EP300	Dex 20min	ChIP-seq	A549	ENCSR167QIJ
EP300	Dex 25min	ChIP-seq	A549	ENCSR044IFH
EP300	Dex 30min	ChIP-seq	A549	ENCSR260WCE
EP300	Dex 1hr	ChIP-seq	A549	ENCSR358ELZ
EP300	Dex 2hr	ChIP-seq	A549	ENCSR770OTI
EP300	Dex 3hr	ChIP-seq	A549	ENCSR047EVQ
EP300	Dex 4hr	ChIP-seq	A549	ENCSR145YCX
EP300	Dex 5hr	ChIP-seq	A549	ENCSR610RKF
EP300	Dex 6hr	ChIP-seq	A549	ENCSR841ASB
EP300	Dex 7hr	ChIP-seq	A549	ENCSR467VXG
EP300	Dex 8hr	ChIP-seq	A549	ENCSR561ZRE
EP300	Dex 10hr	ChIP-seq	A549	ENCSR792VMN
EP300	Dex 12hr	ChIP-seq	A549	ENCSR124VXG
H3K27ac	Dex 0hr	ChIP-seq	A549	ENCSR778NQS
H3K27ac	Dex 5min	ChIP-seq	A549	ENCSR734FLK
H3K27ac	Dex 10min	ChIP-seq	A549	ENCSR027BPE
H3K27ac	Dex 15min	ChIP-seq	A549	ENCSR325VCV
H3K27ac	Dex 20min	ChIP-seq	A549	ENCSR864KVZ
H3K27ac	Dex 25min	ChIP-seq	A549	ENCSR480OHP

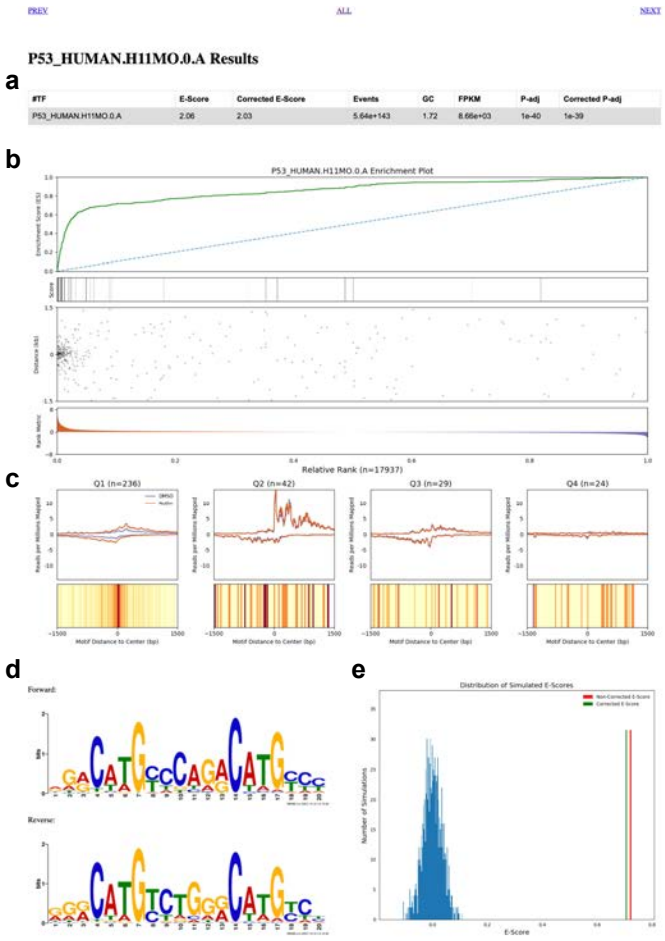
H3K27ac	Dex 30min	ChIP-seq	A549	ENCSR102XUM
H3K27ac	Dex 1hr	ChIP-seq	A549	ENCSR242TBH
H3K27ac	Dex 2hr	ChIP-seq	A549	ENCSR614NPG
H3K27ac	Dex 3hr	ChIP-seq	A549	ENCSR350EFV
H3K27ac	Dex 4hr	ChIP-seq	A549	ENCSR543ZVZ
H3K27ac	Dex 5hr	ChIP-seq	A549	ENCSR716XDB
H3K27ac	Dex 6hr	ChIP-seq	A549	ENCSR340NAL
H3K27ac	Dex 7hr	ChIP-seq	A549	ENCSR569IBY
H3K27ac	Dex 8hr	ChIP-seq	A549	ENCSR250EHC
H3K27ac	Dex 10hr	ChIP-seq	A549	ENCSR180YHA
H3K27ac	Dex 12hr	ChIP-seq	A549	ENCSR435JKM
H3K4me1	Dex 0hr	ChIP-seq	A549	ENCSR636PIN
H3K4me1	Dex 30min	ChIP-seq	A549	ENCSR593RGY
H3K4me1	Dex 1hr	ChIP-seq	A549	ENCSR537FVU
H3K4me1	Dex 2hr	ChIP-seq	A549	ENCSR726MAP
H3K4me1	Dex 3hr	ChIP-seq	A549	ENCSR171ZJG
H3K4me1	Dex 4hr	ChIP-seq	A549	ENCSR949IDI
H3K4me1	Dex 5hr	ChIP-seq	A549	ENCSR225AOO
H3K4me1	Dex 6hr	ChIP-seq	A549	ENCSR868MLT
H3K4me1	Dex 7hr	ChIP-seq	A549	ENCSR462JVS
H3K4me1	Dex 8hr	ChIP-seq	A549	ENCSR404OLV
H3K4me1	Dex 10hr	ChIP-seq	A549	ENCSR954HUB
H3K4me1	Dex 12hr	ChIP-seq	A549	ENCSR529YKU
H3K4me2	Dex 0hr	ChIP-seq	A549	ENCSR410BCN
H3K4me2	Dex 30min	ChIP-seq	A549	ENCSR215DID
H3K4me2	Dex 1hr	ChIP-seq	A549	ENCSR692JHM
H3K4me2	Dex 2hr	ChIP-seq	A549	ENCSR124YCC
H3K4me2	Dex 3hr	ChIP-seq	A549	ENCSR918VQU
H3K4me2	Dex 4hr	ChIP-seq	A549	ENCSR834LCU
H3K4me2	Dex 5hr	ChIP-seq	A549	ENCSR555EAA
H3K4me2	Dex 6hr	ChIP-seq	A549	ENCSR905REY
H3K4me2	Dex 7hr	ChIP-seq	A549	ENCSR016PSC
H3K4me2	Dex 8hr	ChIP-seq	A549	ENCSR774KCU
H3K4me2	Dex 10hr	ChIP-seq	A549	ENCSR766NHB
H3K4me2	Dex 12hr	ChIP-seq	A549	ENCSR428DFL
H3K4me3	Dex 0hr	ChIP-seq	A549	ENCSR203XPU

H3K4me3	Dex 30min	ChIP-seq	A549	ENCSR677QYM
H3K4me3	Dex 1hr	ChIP-seq	A549	ENCSR928FDN
H3K4me3	Dex 2hr	ChIP-seq	A549	ENCSR252FZA
H3K4me3	Dex 3hr	ChIP-seq	A549	ENCSR483JJT
H3K4me3	Dex 4hr	ChIP-seq	A549	ENCSR901AAW
H3K4me3	Dex 5hr	ChIP-seq	A549	ENCSR524UOX
H3K4me3	Dex 6hr	ChIP-seq	A549	ENCSR646OPC
H3K4me3	Dex 7hr	ChIP-seq	A549	ENCSR285FZP
H3K4me3	Dex 8hr	ChIP-seq	A549	ENCSR618MUP
H3K4me3	Dex 10hr	ChIP-seq	A549	ENCSR139DGM
H3K4me3	Dex 12hr	ChIP-seq	A549	ENCSR944WVU
H3K9me3	Dex 0hr	ChIP-seq	A549	ENCSR775TAI
H3K9me3	Dex 30min	ChIP-seq	A549	ENCSR109KEL
H3K9me3	Dex 1hr	ChIP-seq	A549	ENCSR954SQX
H3K9me3	Dex 2hr	ChIP-seq	A549	ENCSR936UEX
H3K9me3	Dex 3hr	ChIP-seq	A549	ENCSR354ERB
H3K9me3	Dex 4hr	ChIP-seq	A549	ENCSR037DAC
H3K9me3	Dex 5hr	ChIP-seq	A549	ENCSR299MNA
H3K9me3	Dex 6hr	ChIP-seq	A549	ENCSR032SCO
H3K9me3	Dex 7hr	ChIP-seq	A549	ENCSR873VJA
H3K9me3	Dex 8hr	ChIP-seq	A549	ENCSR791BNU
H3K9me3	Dex 10hr	ChIP-seq	A549	ENCSR013RTF
H3K9me3	Dex 12hr	ChIP-seq	A549	ENCSR451MJX

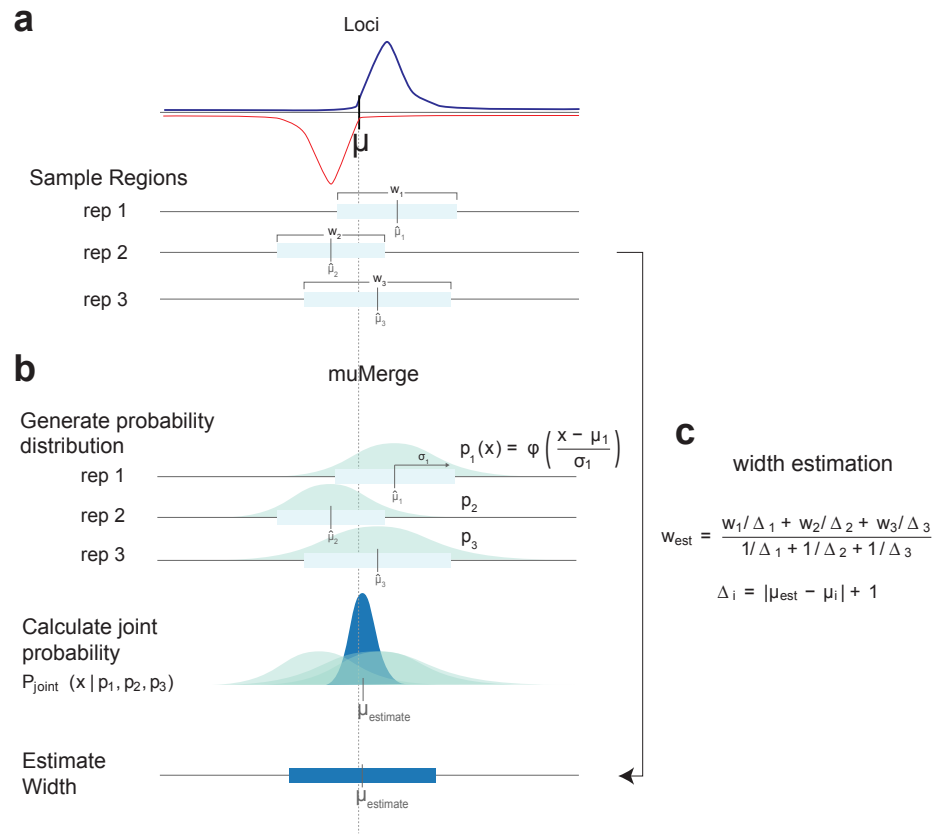
Supplemental Table 3: Accession numbers used in Figure 6, data from [5, 8].



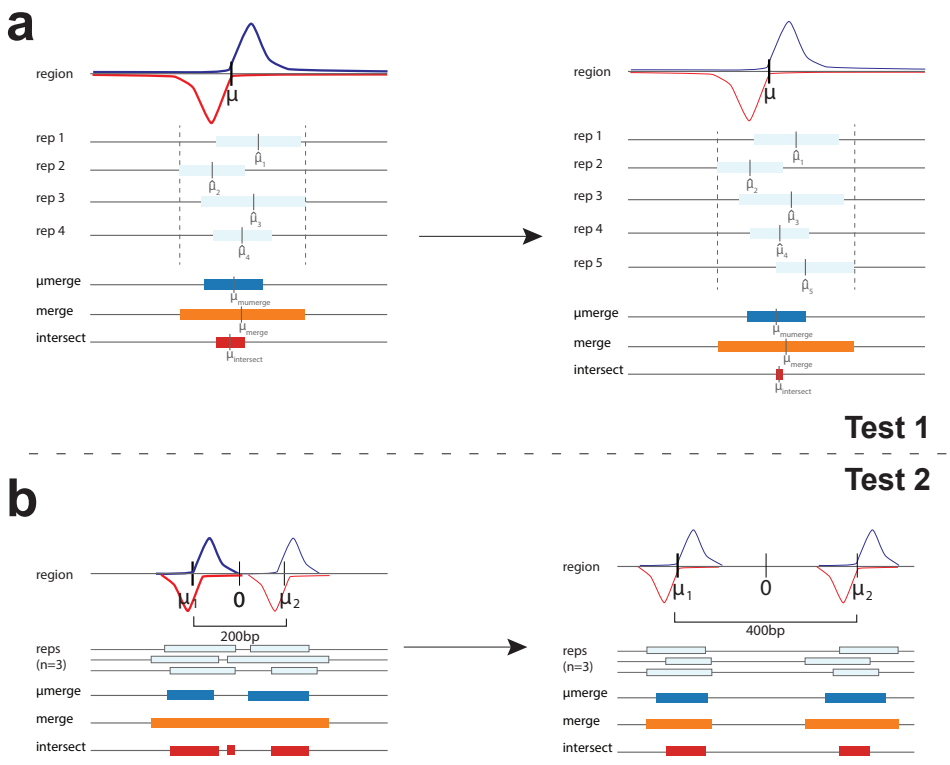
Supplemental Figure 1: **An example of TFEA HTML results page.** (a) Pre-GC correction showing the E-Score of each motif (y-axis) as a function of GC-content (x-axis). Red line: linear regression fit; dots colored by the amount to correct. (b) A scatter plot (colored as in a), similar to an MA-plot, showing the GC-corrected E-Scores (y-axis) vs number of motif hits within regions (<1.5kb; x-axis) for each motif analyzed. (c) An MA-plot of the ROIs generated from DE-Seq2. (d) A table listing the inputs, text results, MD-Score and MDD-Score results (as clickable links), as well as the time taken to complete each step of the TFEA process. (e) A list of motifs that exhibit positive or negative enrichment ordered by adjusted p-value. Significant motifs appear as red and have clickable links to individual results pages with more detailed information. Data is HCT116 dataset, as used in Figure 1a[1].



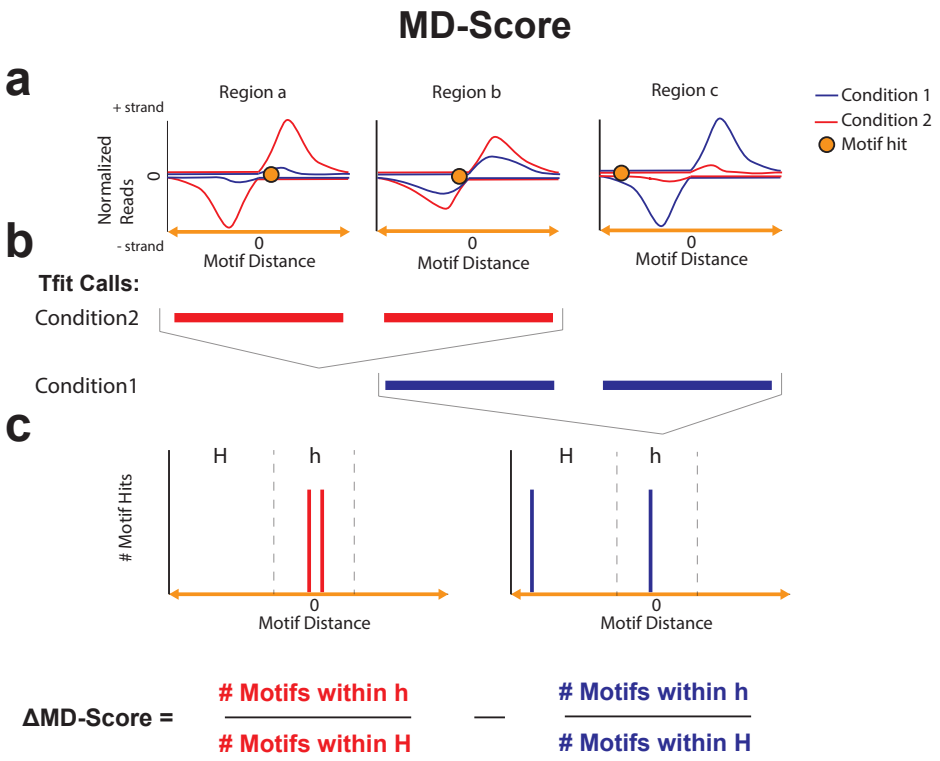
Supplemental Figure 2: **An example of a TFEA individual motif results page.** This page is reached by clicking on the corresponding motif in Figure 1e. (a) Summary statistics for the motif of interest, in this case p53 from HOCOMOCO v11. (b) Enrichment plot showing (from top to bottom) the running sum statistic (green line), the individual scores of each ROI (as heatmap), scatter plot of motif hits within ROIs relative to the reference point (labeled 0), and the ranking of ROIs based on differential transcription (red: positive; blue: negative). (c) For each quartile, summarize motif containing ROI within the quartile via Top: Meta plot of read coverage over ROIs. Bottom: Motif displacement distribution (as heatmap) summarizing the motif positions relative to the reference point. (Yellow is background to Red at max instances). (d) Forward and reverse complement position specific scoring matrix of the motif analyzed. (e) Histogram of E-Scores from randomly shuffling the rank order of ROIs (blue) with true non-corrected E-score (red) and GC-corrected E-score (green).



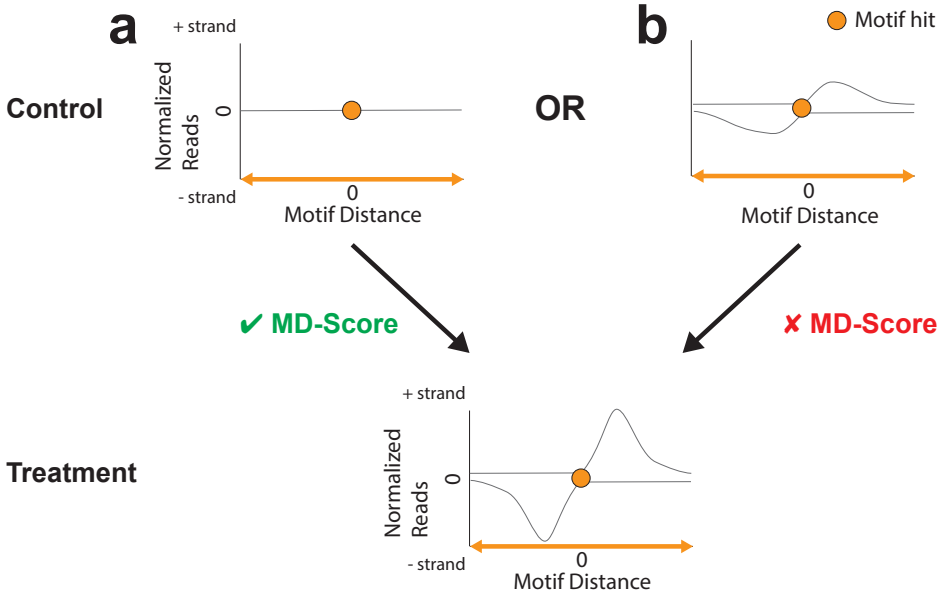
Supplemental Figure 3: **Diagrammatic description of the *muMerge* method** (a) The goal of *muMerge* is to combine multiple sample regions (light blue boxes) which originate from different replicates and/or conditions, but are measurements of the same underlying loci μ, into a consensus set of ROIs. Red and blue lines are hypothetical data. (b) *muMerge* assumes that each sample region is an estimate on the location of a genomic loci of interest and models this probability (p_i) as a normal distribution (light blue) with μ_i equal to the center of each sample region. Subsequently, a joint probability (p_{joint}, dark blue) is calculated from the samples, and the estimate for the consensus position ($\mu_{estimate}$) is the maxima of this joint distribution. (c) Finally, to calculate the best estimate for the width of the ROI, a weighted average of the sample region widths is calculated. It is assumed that the sample regions closest to the consensus position are the most accurate representation of the underlying loci, so the weighted average of the widths is calculated such that more weight is given to the sample regions closer to $\mu_{estimate}$.



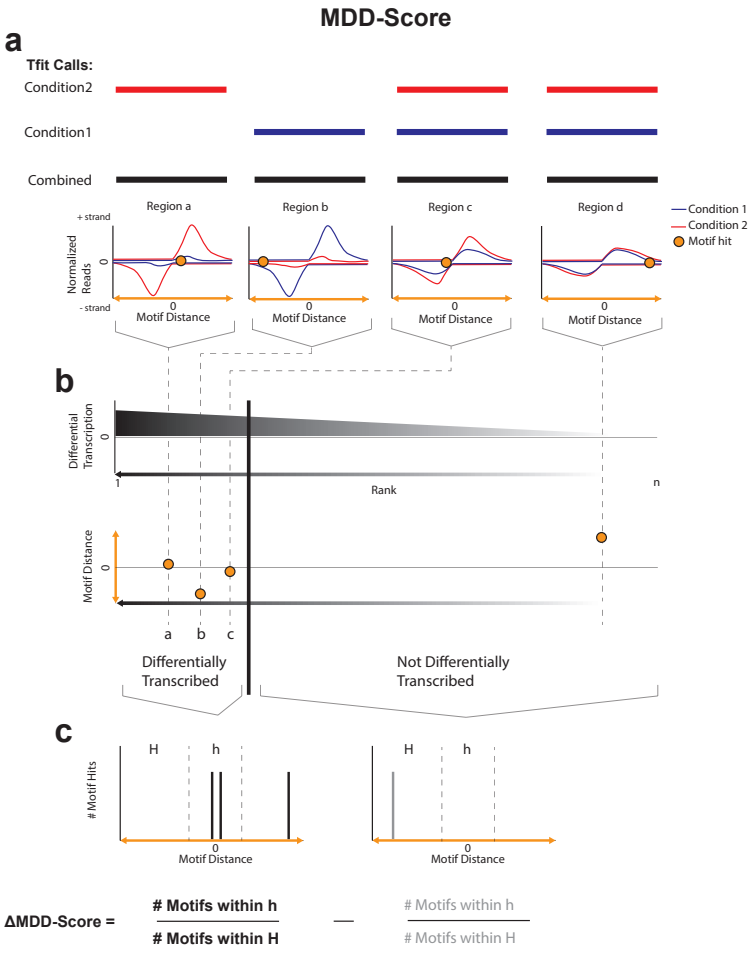
Supplemental Figure 4: **Tests performed to compare the performance of *muMerge* with bedtools merge and bedtools intersect.** (a) The first test involves sampling regions from a single theoretical loci with increasing replicates (light blue). *muMerge* retains correct length and *mu* position, bedtools merge tends to increase ROI length and bedtools intersect tends to decrease ROI length as more replicates are included. (b) A second test to determine performance when sampling from two theoretical loci as a function of inter-loci spacing. For closely spaced loci, *muMerge* correctly separates the two loci whereas bedtools merge is more likely to generate a single ROI, and bedtools intersect is more likely to generate multiple separate ROI (in this example, three). For both tests, top cartoon depicts sequencing data histograms on two strands (blue: positive strand; red (negative strand). Regions inferred from individual replicates in light blue. ROI ascertained by *muMerge* (dark blue), *bedtools merge* (orange) and *bedtools intersect* (red) shown for comparison.



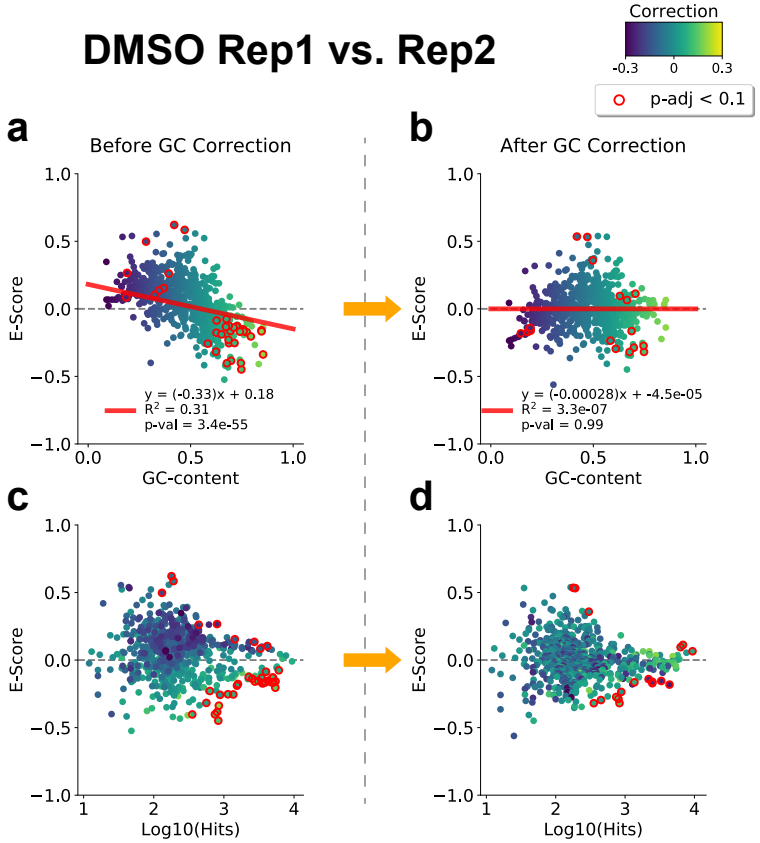
Supplemental Figure 5: **Cartoon diagram description of MD-Score method.** (a) Cartoon depicting typical histograms of nascent transcription data for three example regions. Orange dot is motif location. (b) Tfit called sites of RNA polymerase initiation in each dataset (red, blue) as called by Tfit[3]. These regions are the inputs to the MD-score approach[2]. (c) Motif displacement distribution histograms plot position of motif (vertical bars) relative to reference point (labeled 0) for both conditions (red and blue). The MD-Score is the fraction of motif instances within the inner window (h) divided by the total motif hits in the larger window (H; note H encompasses h). MD-Scores are calculated independently in each of the two conditions to obtain the difference.



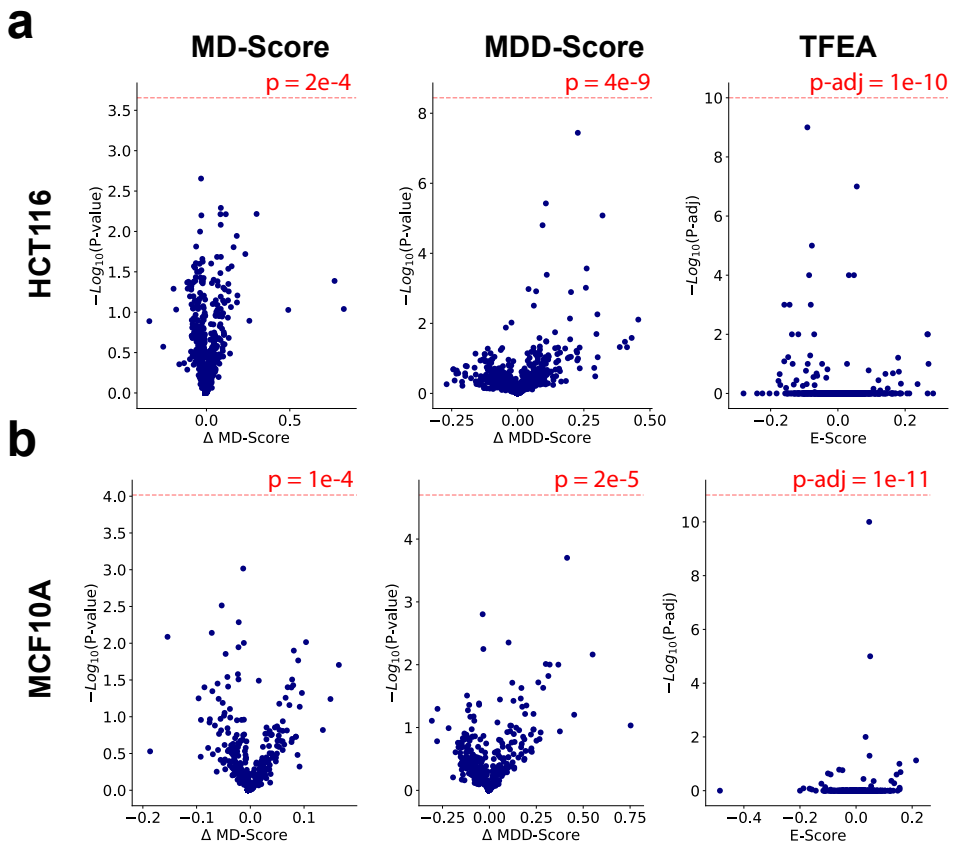
Supplemental Figure 6: **The MD-Score approach only detects gain or loss of transcribed regions.** A given locus in the treatment can arise from either (a) a region of no signal in the control; or (b) increase in signal at a pre-existing region within the control sample. Importantly, the first case increases the Δ MD-Score whereas the second does not alter the Δ MD-Score.



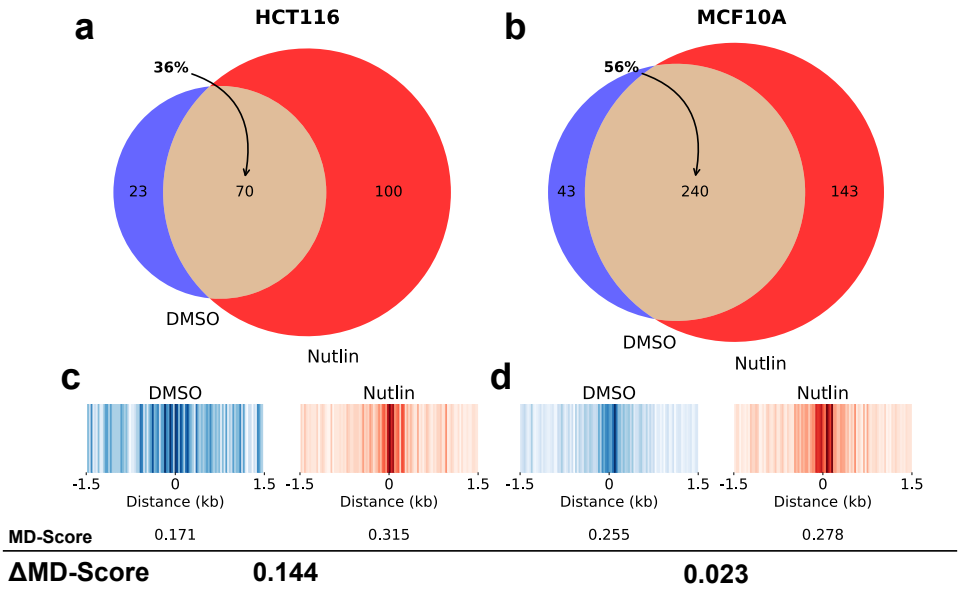
Supplemental Figure 7: **Cartoon diagram depicting the MDD-Score method.** The differential MD-Score method (referred to as MDD-Score)[9, 7] begins with (a) a collection of regions called in one or more conditions (red and blue). (b) Regions are ranked by DESeq or DESeq2 p-value (depending on replicate number) and a cutoff segregates identifies the differentially transcribed subset. (c) The differential MD-Score is calculated similarly to the MD-Score but between the differentially transcribed set (black) and the not differentially transcribed (in grey).



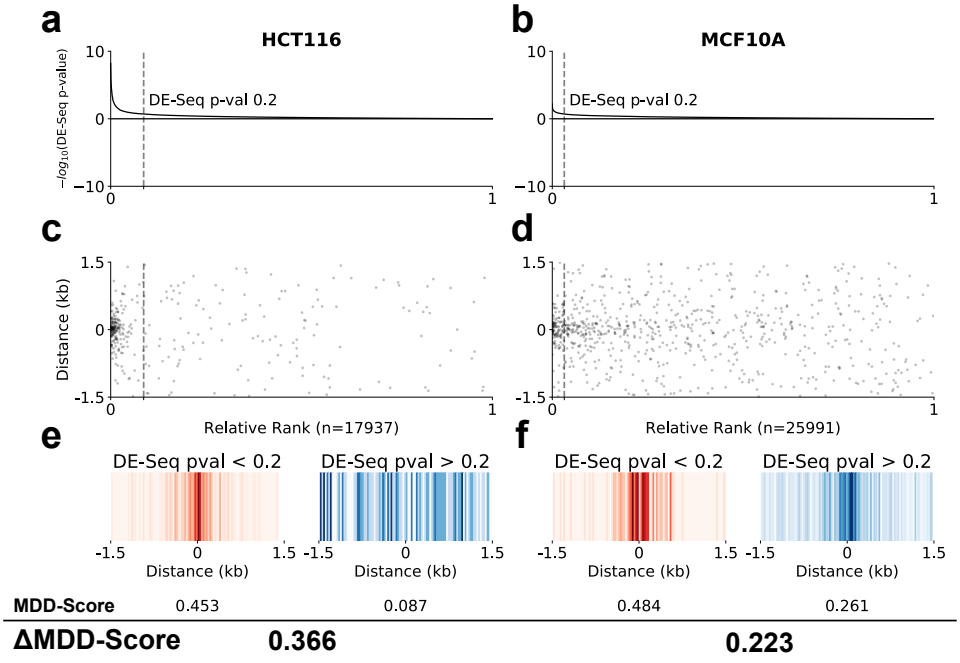
Supplemental Figure 8: **E-Scores are adjusted based on the GC content bias using linear regression.** We observed that motif E-Scores often correlated with their GC-content. (a) Scatter plot of E-Score (y-axis) vs. GC-content (x-axis) of motifs, comparing replicate 1 vs. replicate 2 (DMSO condition) before GC-correction (red line: linear regression fit). (b) Scatter plot of E-Score (y-axis) vs. GC-content (x-axis) of motifs after GC correction (red line: linear regression fit). (c) MA-plot of E-Score (y-axis) vs. Log10 of number of motif hits within regions of interest (x-axis) before GC correction. (d) MA-like plot of E-Score (y-axis) vs. Log10 of number of motif hits within regions of interest (x-axis) after GC-correction. These MA-Plots show that the underlying distribution of E-Scores relative to number of motif hits does not significantly change after GC-correction. All panels are data in HCT116 DMSO condition (SRR1105736, SRR1105737 [1], dots are colored by the amount to be corrected due to GC-bias, red outline dots are p-adj < 0.1).



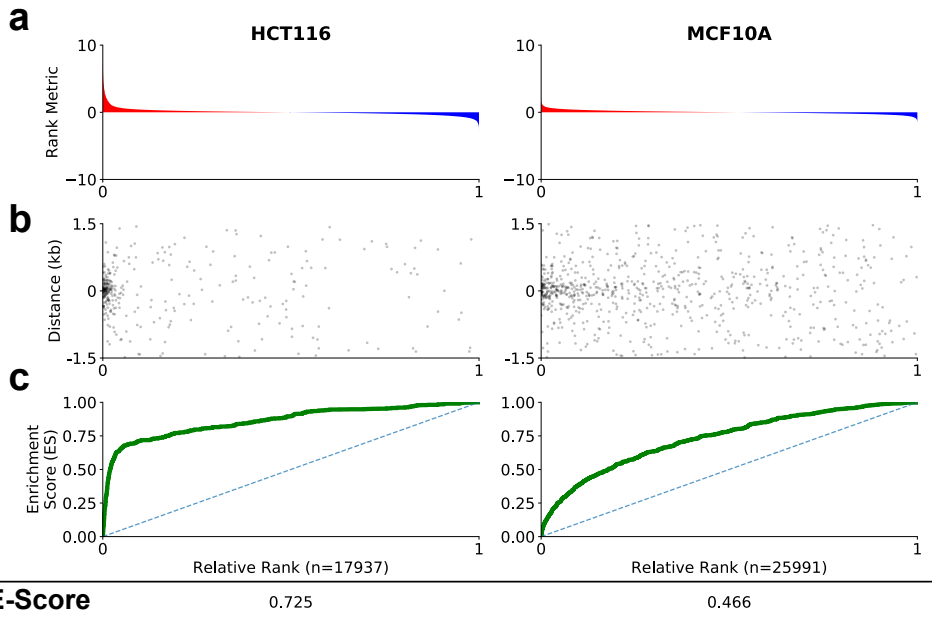
Supplemental Figure 9: **Choosing thresholds for MD-Score, MDD-Score, and TFEA.** To choose a threshold cutoff for the three methods used here, DMSO replicates were compared and the threshold at which no false positives are obtained was determined. To be conservative, an additional order of magnitude is added for stringency. We performed this for each method in either (a) HCT116 or (b) MCF10A cells.



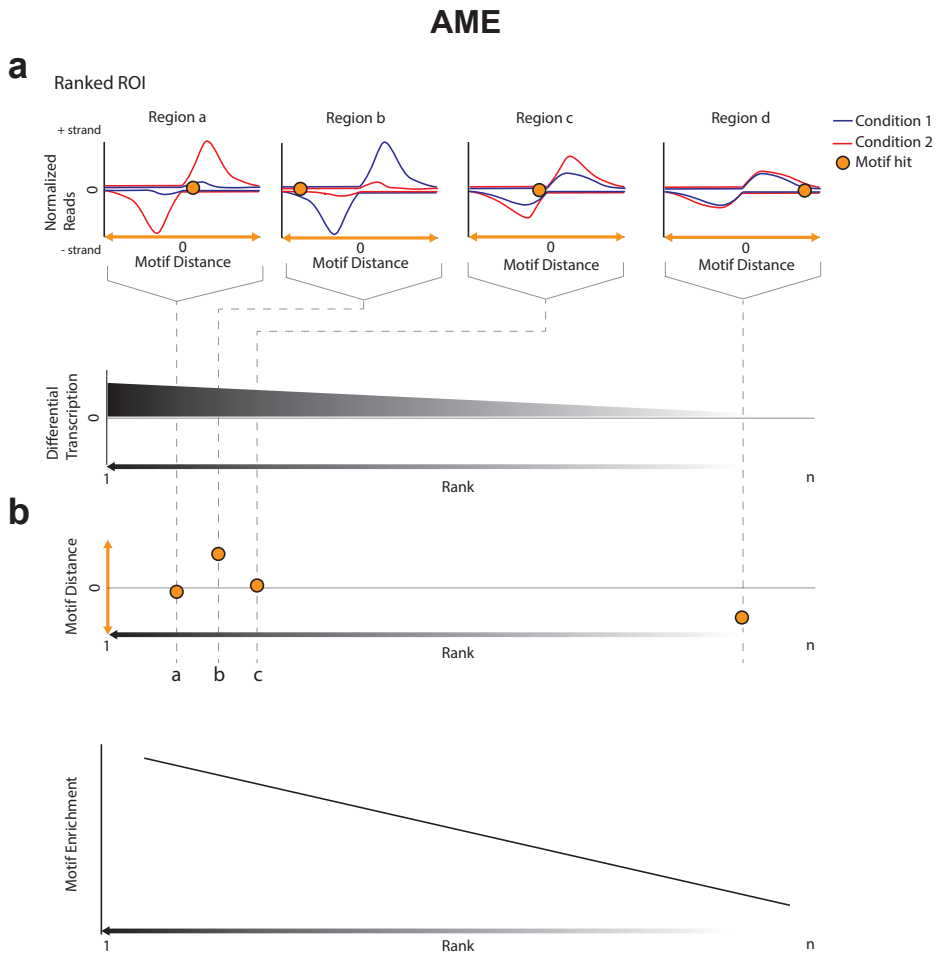
Supplemental Figure 10: **The MD-score approach fails to capture p53 after Nutlin treatment in MCF10A cells.** The response to Nutlin-3a visualized as Venn diagrams of (a) HCT116 and (b) MCF10a cells shows distinct p53 response, with a larger proportion (in MCF10A cells) of existing sites of RNA polymerase initiation that respond to Nutlin-3a. In both cases, only regions with p53 motif within 150 bps of the point of interest (midpoint of ROI) are shown. Motif displacement distributions of TP53 motif within 1.5 kb of ROI midpoints for (c) HCT116 or (d) MCF10A cells shows a higher co-localization of p53 in DMSO treated MCF10A cells. Bottom: MD-Score quantification for each condition followed by the observed ΔMD-Score for the Nutlin-3a response in each cell type.



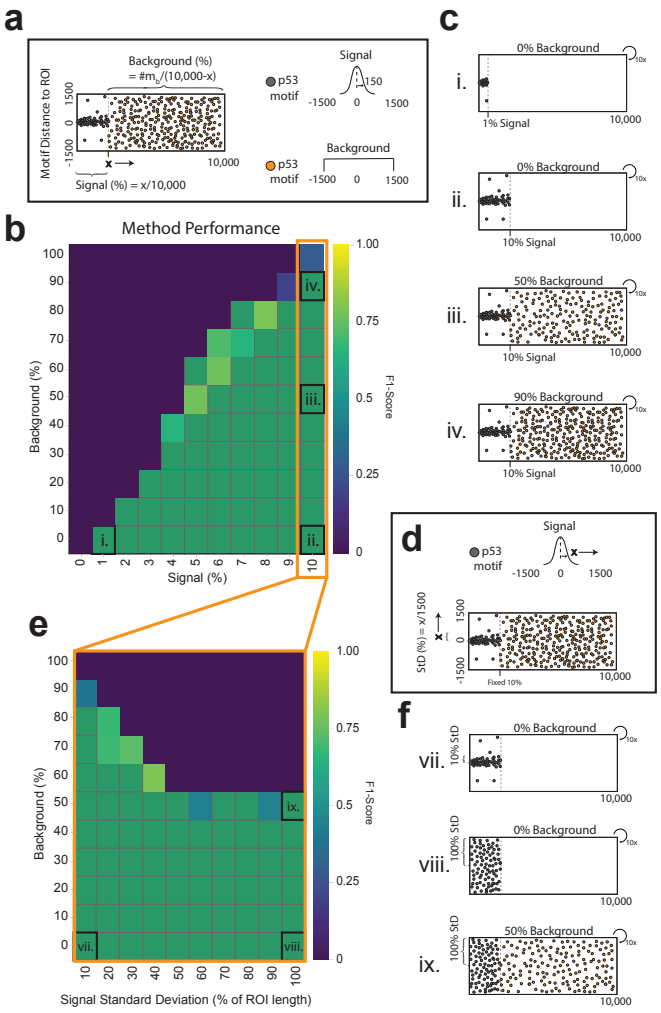
Supplemental Figure 11: **The MDD-Score method detects p53 following Nutlin treatment in both cell types.** The MDD-Score approach detects p53 response in both (a) HCT116 and (b) MCF10a cells. By default, a loose DESeq2 p-value of 0.2 is chosen to identify the set of differentially transcribed ROI. Scatterplots show instances of TP53 motif across ranked ROI for (c) HCT116 and (d) MCF10A cells. The presence of constitutive TP63 activity leads MCF10a cells to have a higher background signal around TP53 motifs. Motif displacement distribution heatmaps for (e) HCT116 and (f) MCF10A cells, further emphasize the increased background presence of the TP53 motifin MCF10A cells. Red is control (DMSO), blue is Nutlin treated. All panels are HCT116 data from SRR1105736, SRR1105737, SRR1105738, SRR1105739.



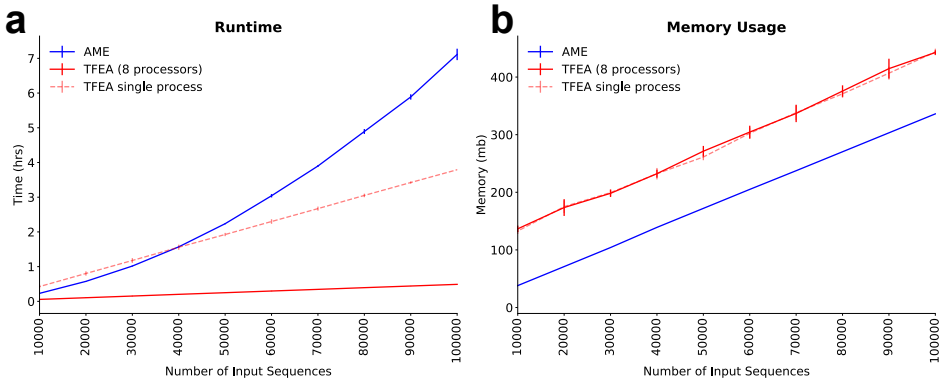
Supplemental Figure 12: **TFEA detects p53 in both HCT116 cells and MCF10A cells without the use of fixed thresholds.** (a) ROI are ranked by differential transcription. Red: increased transcription, blue: decreased. (b) Instances of the TP53 motif are detected within ranked ROIs. (c) TFEA measures motif enrichment as the E-Score, calculated as 2* AUC (ie. area under the curve) between the running sum of ROI scores (green line) and the uniform distribution (dashed blue line). HCT116 data from SRR1105736, SRR1105737, SRR1105738, SRR1105739.



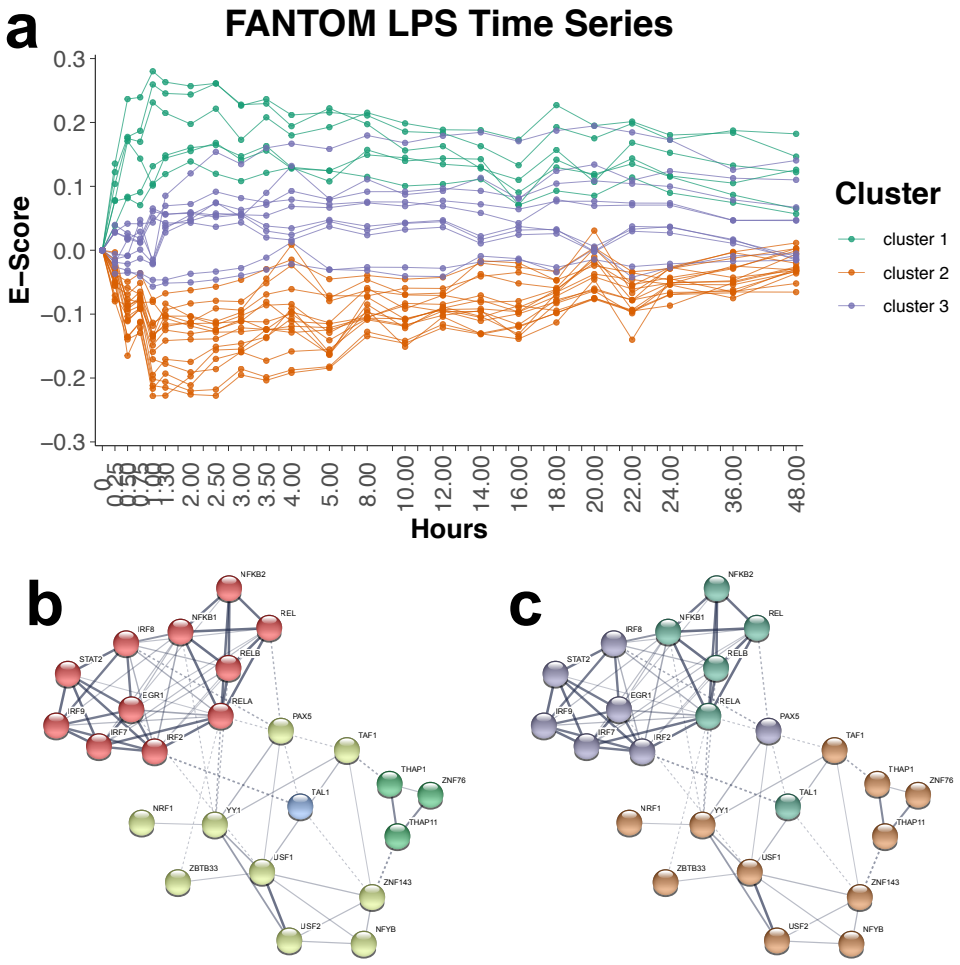
Supplemental Figure 13: **A cartoon diagram depicting the AME method.** Analysis of Motif Enrichment (AME) is part of the MEME suite and requires (a) a ranked list of ROIs as input. AME then performs (b) linear regression on the motifs as a function of rank.



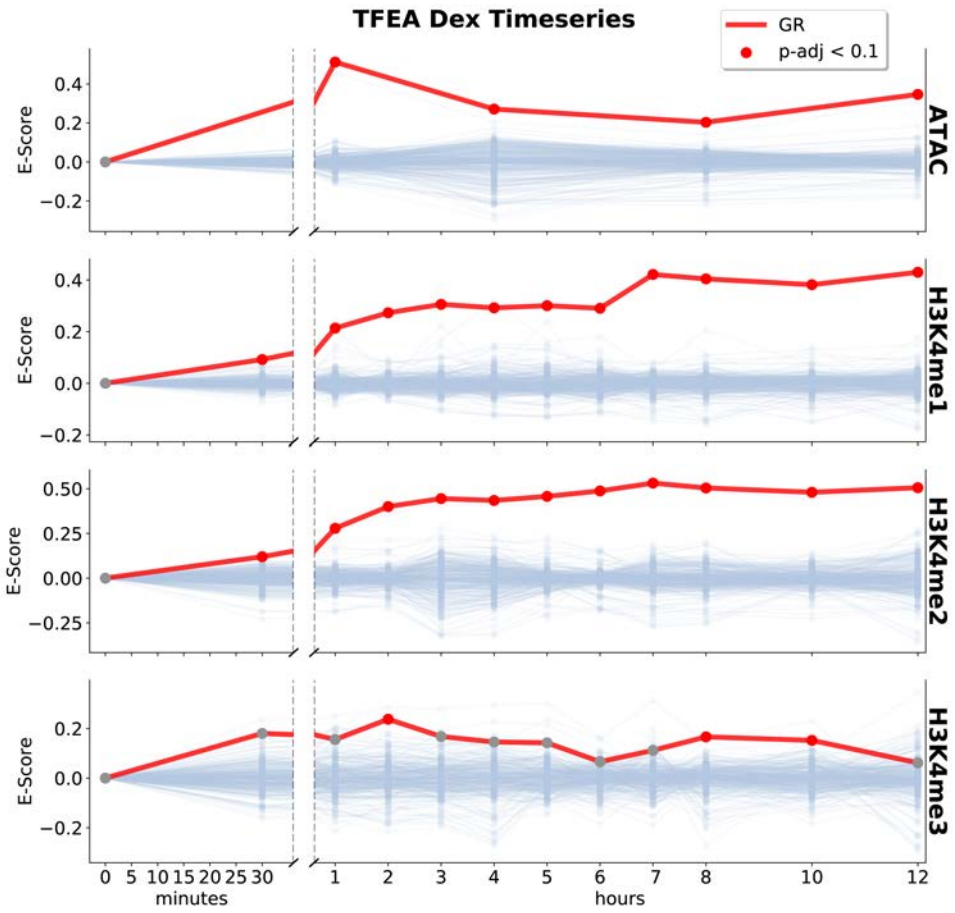
Supplemental Figure 14: **Diagram depicting the benchmark strategy utilized in Figure 4.** (a) A description of key concepts of motif embedding strategy for both signal (grey) and background (orange). (b) F1-Score (as heatmap) for benchmark varying fraction of ROI with signal (x-axis) and background (y-axis). Representative tests cases are labeled (i-iv) and their (c) respective embedding strategies are shown. (d) A description of additional criteria utilized for altering variability of signal embedding. (e) For 10% signal, we additionally alter the signal standard deviation (x-axis) vs background (y-axis). Representative cases (vii-ix) are labeled and their (f) respective embedding strategies are shown.



Supplemental Figure 15: **TFEA is fast and memory efficient.** (a) Runtime statistics for AME (solid blue; parallel processing not supported) and TFEA (8 processors: solid red; 1 processor: dashed red) with varying numbers of input ROI (bars = standard deviation of 10 runs). (b) Memory usage statistics comparing AME to TFEA.



Supplemental Figure 16: **Clustering LPS induced TFs based on dynamics over time.** We applied k-means clustering to the subset of TFs that were significant (by TFEA) in at least 15 time points (~2/3 of all timepoints; n=32 TFs). (a) Time series traces of significant TFs colored by resulting cluster. The three main clusters correspond to the immediate increased responders (cluster 1, green), the immediate decreased responders (cluster 2, orange) and the later responding TFs (cluster 3, purple). (b) Alternatively the TFs can be analyzed using the String database using the Markov cluster algorithm. (c) Superposition of the coloring scheme in (a) onto the network cluster of (b).



Supplemental Figure 17: **TFEA recovers the glucocorticoid receptor (GR) following treatment with Dexamethasone.** TFEA is able to recover GCR in many distinct data sets including ATAC, H3K4me1, and H3K4me2. Interestingly, TFEA only detects moderate enrichment of GR in H3K4me3, in agreement which previous results indicating that GR primarily binds to enhancers (which do not have the H3K4me3 mark).

References

[1] M. A. Allen, H. Mellert, V. Dengler, Z. Andryzik, A. Guarnieri, J. A. Freeman, X. Luo, W. L. Kraus, R. D. Dowell, and J. M. Espinosa. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *eLife*, 3:e02200, 2014. doi: 10.7554/eLife.02200.

[2] J. G. Azofeifa, M. A. Allen, J. R. Hendrix, T. Read, J. D. Rubin, and R. D. Dowell. Enhancer RNA profiling predicts transcription factor activity. *Genome research*, feb 2018.

[3] J. G. Azofeifa and R. D. Dowell. A generative model for the behavior of RNA polymerase. *Bioinformatics*, 33(2):227–234, 09 2016.

[4] J. K. Baillie, E. Arner, C. Daub, M. De Hoon, M. Itoh, H. Kawaji, T. Lassmann, P. Carninci, A. R. R. Forrest, Y. Hayashizaki, G. J. Faulkner, C. A. Wells, M. Rehli, P. Pavli, K. M. Summers, and D. A. Hume. Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. *PLoS Genetics*, 13(3), Mar. 2017.

[5] C. A. Davis, B. C. Hitz, C. A. Sloan, E. T. Chan, J. M. Davidson, I. Gabdank, J. A. Hilton, K. Jain, U. K. Baymuradov, A. K. Narayanan, K. C. Onate, K. Graham, S. R. Miyasato, T. R. Dreszer, J. S. Strattan, O. Jolanki, F. Y. Tanaka, and J. M. Cherry. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research*, 46(D1):D794–D801, 2018.

[6] A. R. R. Forrest, H. Kawaji, M. Rehli, J. Kenneth Baillie, M. J. L. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jørgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. A. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. C. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. Maxwell Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici,

Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drabløs, A. S. B. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furuno, J.-i. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. Peter Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. J. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-sim, R.-i. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohmiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. D. Prendergast, O. J. L. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, H. Satoh, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. C. 't Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyoda, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verardo, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida,

S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci, Y. Hayashizaki, and The FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature*, 507(7493):462–470, Mar. 2014.

[7] M. A. Gruca, M. A. Gohde, and R. D. Dowell. Annotation agnostic approaches to nascent transcription analysis: Fast read stitcher and transcription fit. *Methods in Molecular Biology*, to appear, 2019.

[8] I. C. McDowell, A. Barrera, A. M. D'Ippolito, C. M. Vockley, L. K. Hong, S. M. Leichter, L. C. Bartelt, W. H. Majoros, L. Song, A. Safi, D. D. Koçak, C. A. Gersbach, A. J. Hartemink, G. E. Crawford, B. E. Engelhardt, and T. E. Reddy. Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. *Genome Research*, Aug. 2018.

[9] S. K. Sasse, M. Gruca, M. A. Allen, V. Kadiyala, T. Song, F. Gally, A. Gupta, M. A. Pufall, R. D. Dowell, and A. N. Gerber. Nascent transcript analysis of glucocorticoid crosstalk with tnf defines primary and cooperative inflammatory repression. *Genome Research*, 2019.