

UCLA

UCLA Electronic Theses and Dissertations

Title

The Selective and Combinatorial Regulation of Toll-Like Receptor-Activated Transcriptional Cascades

Permalink

<https://escholarship.org/uc/item/7k44j89n>

Author

Liu, Xin Liu

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Selective and Combinatorial Regulation of Toll-Like
Receptor-Activated Transcriptional Cascades

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Molecular Biology

by

Xin Liu

2017

ABSTRACT OF THE DISSERTATION

The Selective and Combinatorial Regulation of Toll-Like Receptor-Activated Transcriptional Cascades

by

Xin Liu

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2017

Professor Stephen T. Smale, Chair

The immune system is essential for host defense to pathogen infection, tissue repair, stress response, and other physiological functions. An unbalanced immune system is detrimental to homeostasis and mammalian survival. Immunodeficiency can cause susceptibility to pathogen infection. A hyperactive immune system can cause autoimmunity and chronic inflammatory diseases including rheumatoid arthritis, psoriasis, and atherosclerosis. Deciphering the mechanisms regulating the immune response will illuminate on the pathogenesis of inflammatory diseases and promote the development of new therapies to treat these diseases. Because innate immunity is the first line of host defense and transcription is a critical contributor to the innate immune response, we focused our studies on mechanisms regulating transcriptional activation of the innate immune response.

Toll-like receptor (TLR) signaling is a classical model to study pathogen recognition and the activation of innate immunity. TLR transcriptional cascades have been extensively studied using conventional systems approaches. Because conventional systems approaches rely on statistics and large sample sizes to uncover the common regulatory mechanisms, they often miss gene-specific regulatory mechanisms. To reveal gene-specific regulatory mechanisms, we used a stringent systems approach to dissect the TLR transcriptional cascades in chapter 2. To prevent biases towards the majority of weakly induced genes, we separated strongly induced genes from weakly induced genes. Combining high-resolution transcriptional profiles and perturbation studies, we classified TLR4-activated genes by their activation mechanisms. By integrating RNA-seq, ChIP-seq, ATAC-seq, and motif data, we found that several key inflammatory genes are regulated by highly selective mechanisms. TLR4-activated nuclear factor kappa B (NF κ B) and interferon regulatory factor 3 (IRF3) selectively regulate a small subset of pro-inflammatory genes by chromatin and combinatorial regulation. We also found serum response factor (SRF) selectively regulates a few early transient primary response genes. In chapter 3, we used a gene-centric method and identified the motif rules governing the combinatorial regulation of serum response factor (SRF) and ternary complex factor (TCF). Ternary complex formation can enhance SRF binding and promote histone mark deposition. The active ternary complex can regulate transcription through either promoter or enhancer. Taken together, we demonstrated a gene-centric, stringent systems approach that can complement the conventional systems approach to unveil gene-specific regulatory mechanisms in ligand-induced transcriptional cascades.

The dissertation of Xin Liu is approved.

Arnold J. Berk

Harvey R. Herschman

John J. Colicelli

Robert L. Modlin

Stephen T. Smale, Committee Chair

University of California, Los Angeles

2017

Dedication

This work is dedicated to my parents, Yuee Jiang and Bing Liu.

TABLE OF CONTENTS

List of Figures and Tables	vii
Acknowledgments	ix
VITA	xi
CHAPTER 1	
Introduction: The Transcriptional Regulation of Innate Immune Response	1
CHAPTER 2	
A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation	50
CHAPTER 3	
A Gene-Centric Approach Unravels the Selective and Combinatorial Regulation of SRF Targets	73
CHAPTER 4	
Concluding Remarks: Conclusions and Future Directions	135

LIST OF FIGURES AND TABLES

CHAPTER 1

Figure 1-1: TLR4 Signaling Pathways	33
Figure 1-2: Cell-Specific and Stimulus-Specific Regulation of Gene Transcription	34
Figure 1-3: SRF Regulated by Rho-MRTF and MAPK-TCF Pathways	35

CHAPTER 2

Figure 2-1: Properties of the Lipid A-Induced Transcriptional Cascade	52
Figure 2-2: Analysis of IFNAR-Independent and –Dependent SRGs	54
Figure 2-3: Properties of PRGs	55
Figure 2-4: NFκB Interactions at the Promoters of Defined Gene Classes	56
Figure 2-5: Kinetic and Functional Analysis of Putative NFκB Target Genes	59
Figure 2-6: Analysis of IRF3 Target Genes	60
Figure 2-7: Analysis of SRF Target Genes	63
Figure 2-S1: IFNAR-Independent and IFNAR-Dependent Secondary Response Genes	66
Figure 2-S2 Highly Induced Primary Response Genes Ordered by Their Dependence on Various Signaling Pathways	67
Figure 2-S3: Promoter Motif Analysis of Primary Response Gene Clusters	68
Figure 2-S4: Putative NFκB Target Genes and the Genes that Exhibit Similar Kinetics and/or RelA Dependence	69
Figure 2-S5: The Position of RelA Peaks Relative to the TSSs of All Genes	70
Figure 2-S6: Final Classification of the Primary Response Genes	71
Figure 2-S7: Final Classification of the Secondary Response Genes	72

CHAPTER 3

Figure 3-1. General Features of SRF CHIP-seq in Lipid A-Activated Macrophages	118
Figure 3-2. Identify SRF Targets by RNA-seq and CHIP-seq	119
Figure 3-3. Characterize SRF Peaks by Peak Strength and Motif Strength	120
Figure 3-4. Define Promoter SRF-TCF Cassette of SRF targets	121
Figure 3-5. Identify Promoters with Defined SRF-TCF Cassettes Lacking SRF Binding	122
Figure 3-6. Interrogate the Potent SRF Binding at SRF Targets	123
Figure 3-7. Strong Promoter SRF Peaks Possess Proximate CREB Binding and Multiple SRF Motifs	124
Figure 3-8. SRF-TCF Motifs at Enhancer SRF Peaks Correlate with SRF Targets	125
Figure 3-9. The Role of MRTF in Regulating Promoters with Strong SRF Binding and Strong Motifs	126
Figure 3-10. Evaluate the Roles of the Ternary Complex at the <i>Nr4a1</i> Promoter by CRISPR	127
Figure 3-11. Evaluate the Roles of the Ternary Complex at the <i>Egr3</i> Enhancer by CRISPR	128
Figure 3-12. Robust Induction of SRF Targets by Different Stimuli in Different Cell Types	129
Table 3-1. List of Genes with SRF Binding or SRF Motifs	130

CHAPTER 4

Figure 4-1 The Distribution of SRF and RelA CHIP-seq Peak Scores	151
--	-----

ACKNOWLEDGMENTS

I thank my mentor, Dr. Stephen T. Smale, for his kind guidance and support, for living as a model of critical thinker, and for teaching me things beyond science. I thank my dissertation committee members, Dr. Arnold J. Berk, Dr. John J. Colicelli, Dr. Robert L. Modlin, and Dr. Harvey R. Herschman for the helpful discussion and insightful advice. I owe a lot to the past and current lab members who helped to develop my technical skills, inspired me through discussion, and grew together as scientists: Ann-Jay Tong, Kevin Doty, Brandon Thomas, Prabhat Purbey, Miguel Edwards, Justin Langerman, Philip Scumpia, Peter Kim, Teresita L. Arenzana, Hilde Schjerven, Ameya Champhekar, George Yeh, Jerry Lo, An-Chieh Feng, Cecilia Garcia, Richa Tewari, and Poonam Sachan.

I thank my previous mentors and supervisors for the kind instruction and support. I thank Dr. Zichun Hua, Huang Li, and Xiangyu Zhang for guiding me during my undergraduate research. I thank Dr. Karen Lyons and Kristine Estrada for helping with research, graduate school application, and adapting to a foreign environment during CSST summer program. I thank Dr. Dinesh Rao, Dr. Yousang Gwack, Dr. Sonal Srikanth, Dr. Guillaume Chanfreau for valuable discussion and support during my ACCESS rotation year. I also thank Jorge Contreras, Thilini Fernando, Camille Nery for guidance during my rotation.

I thank Amy Pandya-Jones and Roberto Spreafico for discussion and optimization of chromatin fractionation protocol. I thank Dr. Siavash Kurdistani, Dr. Kathrin Plath, Chen

Cheng, Maria Vogelauer, Roberto Ferrari, Bernadett Papp, Konstantinos Chronis, David Lopez, and Roberto Spreafico for helping with ChIP-seq optimization and data analysis. I thank Dr. Douglas Black, Dr. Peter Tontonoz, Yi Ying, and Ayaka Ito, for allowing me to use their lab equipment. I thank UCLA Broad Stem Cell Research Center Core, especially Suhua Feng, for technical assistance with sequencing. I thank UCLA CSST program for supporting my graduate study at UCLA. I thank my funding support from China Scholarship Council and Whitcome pre-doctoral training program.

I thank my parents, Yuee Jiang and Bing Liu, for the timeless love and support across the ocean.

I thank Elsevier for permission to reprint a research article entitled “A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation” (license#: 4082780667820) for use in chapter 2. Ann-Jay Tong performed RNA-seq experiments and analysis of *Ifnar*^{-/-}, *Myd88*^{-/-}, *Trif*^{-/-}, *Irf3*^{-/-}, and *Rela*^{-/-} BMDMs. Grant Barish performed RelA ChIP-seq. Brandon Thomas performed ATAC-seq experiments and analysis. I thank Annual Reviews for permission to reprint their content in Figure 1-1 (license#: 4082740867171). I thank Elsevier for permission to reprint their content in Figure 1-2 (license#: 4082741209330) and Figure 1-3 (license#: 4082741323052).

VITA

2011	Bachelor of Science, Biological Science Nanjing University, China
2010-2011	CSST Undergraduate Research Scholar University of California, Los Angeles
2011-2012	ACCESS Program University of California, Los Angeles
2012-Present	Molecular Biology Institute University of California, Los Angeles
2012-2013	Teaching Assistant University of California, Los Angeles
2011-2015	China Scholarship Council Scholarship University of California, Los Angeles
2015-2016	Whitcome Fellowship University of California, Los Angeles

PUBLICATIONS

Tong, A.-J.*, **Liu, X.***, Thomas, B.J.*, Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. *Cell* 165, 165–179. *Co-first author.

Li, H., Zhang, X.-Y., Wu, T.-J., Cheng, W., **Liu, X.**, Jiang, T.-T., Wen, J., Li, J., Ma, Q.-L., and Hua, Z.-C. (2013). Endoplasmic Reticulum Stress Regulates Rat Mandibular Cartilage Thinning under Compressive Mechanical Stress. *J. Biol. Chem.* 288, 18172–18183.

PRESENTATIONS

Transcriptional Networks in the Immune Response to Leprosy P50 Grant Meeting, University of California, Los Angeles, May 23rd, 2016. Oral presentation.

Cold Spring Harbor Laboratory Meetings: Gene Expression and Signaling in the Immune System, Cold Spring Harbor, New York, Apr. 26nd-30th, 2016. Poster presentation.

MBI Program Retreat, University of California, Los Angeles. Apr. 22nd-24th, 2016. Poster presentation.

MBIDP Gene Regulation Recruitment, University of California, Los Angeles, Feb. 27th, 2016. Oral presentation.

JCCC Gene regulation monthly meeting. University of California, Los Angeles, Feb. 2nd, 2016. Oral presentation.

UCLA Global Collaboration Network of Translational Research in Hansen's Disease Symposium, University of California, Los Angeles, Jan. 16th, 2015. Poster presentation.

Immunology LA Conference, Skirball Cultural Center, Los Angeles, May 22nd, 2015. Poster presentation.

CHAPTER 1

Introduction

The Transcriptional Regulation of Innate Immune Response

A. Immune Responses

The immune system is indispensable for vertebrate survival. Immune cells circulate throughout the body and participate in a wide variety of roles including tissue development and repair, host defense against pathogen infection, and stress response. Upon pathogen infection, the immune system can trigger two types of responses: the innate immune response and the adaptive immune response. The innate immune response typically involves neutrophils, macrophages, and dendritic cells. They express pattern recognition receptors that can recognize non-self molecules unique to pathogens. Upon activation, they rapidly produce numerous cytokines and chemokines, which in turn recruit other effector cells to sites of inflammation. Furthermore, macrophages and dendritic cells can bridge the innate and adaptive immunity by presenting antigens to adaptive immune cells. Adaptive immune cells are composed of T and B lymphocytes. Adaptive immune cells undergo DNA rearrangement to generate pathogen-specific responses. They can also generate memory cells that respond quickly upon subsequent encounters to the same pathogen. While innate immune cells detect pathogens by pre-programmed systems such as pattern recognition receptors, the adaptive immune cells go through DNA arrangement that allows for plasticity, selectivity, and memory. But recent discoveries suggest that macrophages can also develop trained “memory” to previous stimuli (Netea et al., 2016; Ostuni et al., 2013).

Because the immune system is constantly reacting to intrinsic and extrinsic stimuli, it is in a dynamic equilibrium and requires proper balance between host defense and

homeostasis (Eberl, 2016). Insufficient immune activity can cause immunodeficiency and susceptibility to pathogen infection whereas over-activation of the immune system can cause allergy, autoimmunity, and chronic inflammatory diseases including rheumatoid arthritis, inflammatory bowel disease, and psoriasis (Hato and Dagher, 2014; Eberl, 2016). Because the immune system is involved in maintaining homeostasis in many tissues, its deficiency or over-activation also contribute to the pathogenesis of many “modern diseases” including diabetes, Alzheimer’s disease, atherosclerosis, and cancer (Kotas and Medzhitov, 2015; Okin and Medzhitov, 2012; Lampron et al., 2013; Hansson and Libby, 2006). Demystifying the regulation of the immune system is critical to understanding disease pathogenesis and developing novel therapies for inflammatory, autoimmune, and other diseases. Because pathogen recognition by innate immune cells is the first step of the immune response, understanding the regulation of innate immune cell is key to unraveling the immune response cascades.

B. Pattern Recognition Receptors and TLR Signaling

Innate immune cells can recognize pathogen-associated molecular patterns or danger-associated molecular patterns by pattern recognition receptors. They can sense non-self components and initiate the innate immune response, which in turn activates the adaptive immune response (Janeway, 1989). They can recognize extracellular and intracellular pathogen components such as lipopolysaccharides, lipoprotein, zymosan, and pathogenic nucleic acids. To recognize a broad range of pathogen components, pattern recognition receptors have diversified into many different receptor families. These include Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), nucleotide-binding

oligomerization domain, leucine-rich repeat-containing receptors (NLRs), and C-type lectin receptors (CLRs) (Iwasaki and Medzhitov, 2015; Thompson et al., 2011). Other newly discovered families include AIM2-like receptors, DAI, and cGAS (Bürckstümmer et al., 2009; Takaoka et al., 2007; Cai et al., 2014). TLRs are a well-characterized founding family of pathogen recognition receptors consisting of 9 members. Depending on the location of their ligands, they reside in different cellular compartments. TLR1, TLR2, TLR4, TLR5, TLR6, TLR11 (only in mouse) are located on the cell surface membrane to detect extracellular pathogen components. For example, TLR4 can recognize lipopolysaccharides, mannan, and glycoinositol phospholipids from Gram-negative bacteria. TLR3, TLR7, TLR8, and TLR9 reside on the membrane of endolysosome. They are specialized in sensing intracellular pathogens such as virus and mycobacteria. For example, TLR3 can recognize viral RNA; TLR9 can recognize viral DNA. RLRs are intracellular sensors of viral RNA (Loo and Gale, 2011). NLRs consist of 23 members, located in cytoplasm, plasma membrane, or endosome. They detect a diverse range of intracellular pathogen components including bacterial peptides, bacterial toxin, and bacterial flagellin (Motta et al., 2015). CLRs reside on the cell membrane and detect fungal and bacterial components including glucan, mannan, and glycoprotein (Hoving et al., 2014). The redundancy within and between families allows them to act in concert to amplify innate immune responses (Kawai and Akira, 2011).

C. TLR Signaling and Transcriptional Cascades

Toll-like receptors were discovered 20 years ago, as the first mammalian pattern

recognition receptor family (Lemaitre et al., 1996). Although 11 mammalian TLR members are activated by different ligands, they share several common signaling pathways. TLR4 represents a typical pattern recognition receptor (Figure 1-1). TLR4 is located on the plasma membrane to recognize extracellular microorganisms. Its extracellular domain contains leucine-rich domains adopting a horseshoe-like structure (Kim et al., 2007). With the assistance of other trans-membrane molecules including MD2, LPS binding protein, and CD14, TLR4 can recognize and bind to ligands such as lipopolysaccharides. This induces TLR4 homo-dimerization and conformation changes of the intracellular domain. The intracellular domain belongs to the Toll/IL-1R (TIR) domain family, which can induce intracellular signaling. The selectivity of intracellular pathways is dependent on the ability to recruit different adaptors. Through TIR domain, TLR4 can recruit at least two adaptors: myeloid differentiation primary response gene 88 (MyD88) and TIR-domain-containing adapter-inducing interferon- β (TRIF).

MyD88-dependent pathways trigger a cascade of signal molecules, leading to the activation of the MAPK and NF κ B pathways (Figure 1-1). Activated TLR4 can interact with TIRAP through homotypic interaction of TIR domains (Akira and Takeda, 2004). TIRAP can then recruit MyD88 to membrane TLR4. The activated MyD88 will assemble into a signaling complex called myddosome by recruiting IRAKs and TRAF6 (Brubaker et al., 2015). Upon Phosphorylation and activation by IRAK4, IRAK1 can phosphorylate the E3 ubiquitin ligase TRAF6. Active TRAF6 can induce the polyubiquitination of TAK1 and the release of TAK1 into the cytosol. TAK1 acts upstream of the MAPK and NF κ B pathways (Brubaker et al., 2015; O'Neill et al., 2013). TAK1 phosphorylates IKK β , which

induces phosphorylation and proteasome degradation of NF κ B inhibitor I κ B α . Dissociation from I κ B α allows NF κ B to translocate into the nucleus and activate pro-inflammatory genes. TAK1 also functions as MAPKKK to activate the MAPK pathway, which in turn induces its downstream transcription factors including CREB, AP1, and SRF. These transcription factors also contribute to the induction of many early and late inflammatory genes. Because MyD88-dependent signaling starts at the plasma membrane and does not require receptor internalization, it can rapidly induce gene transcription, generating an early wave of TLR4-activated transcription (Warner and Núñez, 2013).

Another TLR4 signaling pathway, TRIF-dependent pathway, follows the internalization of TLR4 and induces the NF κ B and IRF pathways, leading to a late wave of transcriptional activation. Although its regulation is still unclear, CD14 can mediate activated TLR4 trafficking from plasma membrane to endosome through the ITAM pathway. After internalization to the endosome membrane, TLR4 can interact with the sorting adaptor TRAM and recruit TRIF. Similar to MyD88, TRIF can assemble a signaling complex, which recruits RIPK1, TRADD, caspase-8 complex, and IKKs. Activated IKKs trigger NF κ B signaling, thus promoting sustained NF κ B activation following MyD88-activated NF κ B signaling. TRIF also recruits TRAF3, TANK, TBK1, IKK γ , and IKK ϵ . TBK1 and IKKs phosphorylate and activate IRF3, which can translocate into the nucleus and induce type I IFN production (Yamamoto and Takeda, 2010; Brubaker et al., 2015; Yamamoto and Takeda, 2010).

Intriguingly, TLR4 signaling pathways also induce negative regulators that inhibit TLR4 signaling at multiple stages including the signaling, transcriptional, and post-transcriptional stages. For example, TLR4 signaling can strongly induce genes encoding the negative regulators of the NFκB pathway such as A20. A20 can inhibit NFκB pathway by several mechanisms. It can remove the K63 ubiquitination from TRAF6 or inhibit TRAF6 E3 ligase activity, thus preventing NFκB activation through the MyD88 pathway (Boone et al., 2004; Shembade et al., 2010). It can also inhibit TRAF3 interaction with TBK1 and IKKs, thus preventing NFκB activation through the TRIF pathway (Parvatiyar et al., 2010). TLR4 signaling also rapidly induces transcription of ATF3, which can recruit HDACs and restrict chromatin accessibility to prevent transcriptional induction (Gilchrist et al., 2006; Whitmore et al., 2007). Lastly, TLR4 signaling also strongly induces TTP (also known as ZFP36 or TIS11), an RNA-binding protein. By inducing transcription and phosphorylation, TTP can bind to the AU-rich elements in the 3'-UTR of many pro-inflammatory mRNAs and mediate their degradation (Deleault et al., 2008; Lai et al., 2006). Recent studies also found that TLR4 signaling can produce anti-inflammatory fatty acids, which contribute to inflammation resolution (Oishi et al., 2017). Thus, TLR4 signaling can robustly induce two waves of pro-inflammatory and IFN gene induction, subject to auto-regulatory negative feedback mechanisms.

D. The Contribution of Transcription to TLR4 Activation

Besides transcriptional activation, TLR4 signaling also triggers many changes at the levels of mRNA splicing, mRNA degradation, translation, and protein degradation

(Vogel and Marcotte, 2012). Therefore, the rate of protein accumulation is disproportionate to the rate of transcription. For example, because TLR4 signaling can also activate factors controlling translation, some genes appear weakly or barely inducible by transcription, but their protein levels exhibit significant increase. Because TLR4 signaling also induces factors controlling mRNA degradation or protein degradation, some genes appear unaffected by transcription but show reduced protein abundance (Rabani et al., 2014; Jovanovic et al., 2015). Splicing rates of nascent transcripts can also affect protein abundance (Jovanovic et al., 2015; Bhatt et al., 2012). Because of the technical limitations of microarray and un-labeled mass spectrometry, the contribution of transcription in cellular response was underestimated. With the emergence of RNA-seq, biochemical labeling methods, and improved mass spectrometry, several studies revealed that transcription is the major contributor to cellular responses in both mouse and human (Li and Biggin, 2015; Vogel et al., 2010; Vogel and Marcotte, 2012). Transcription has a huge impact on protein expression. In yeast, every mRNA molecule can produce approximately 5000 protein molecules (Lu et al., 2007). In TLR signaling, mRNA can explain 90% of the changes in protein level, while translation and protein degradation only explain 10% of the changes (Jovanovic et al., 2015). Acting as the upstream of protein synthesis, transcription is also associated with most of phenotypic variance (Battle et al., 2015). Because transcription can significantly affect the output of stimulus-induced responses, we will focus our studies on the transcriptional regulatory mechanisms in macrophages.

E. Primary and Secondary Response Genes Induced by TLR4 Signaling

Ligand-induced transcriptional cascades are universal in eukaryotic cells. By their activation mechanisms, ligand-induced genes can be roughly divided into primary and secondary response genes (K R Yamamoto and Alberts, 1976; Herschman, 1991). Activation of primary response genes depends on existing signaling molecules and is independent of new protein synthesis. They are immediately downstream of receptor-induced signaling through post-translational modifications or nuclear translocation of transcription factors. Therefore, primary response genes are often rapidly induced. In contrast, activation of secondary response genes requires new protein synthesis, which results in delayed transcriptional induction.

Besides their differences in transcriptional kinetics and regulatory mechanisms, primary and secondary response genes also have other distinct characteristics. Many primary response genes possess CpG-island promoters that are constitutively accessible (Ramirez-Carrozzi et al., 2009). Their promoters also associated with active histone marks including H3K4me3, H3S10ph, and histone acetylation (Fowler et al., 2011). Their promoters also have paused RNA Pol II with serine 5-phosphorylation, which can quickly transition into active elongation upon stimulus-induced P-TEFb recruitment (Hargreaves et al., 2009; Medzhitov and Horng, 2009). Many primary response genes are relatively small genes that contain fewer exons than secondary response genes, leading to rapid transcriptional and splicing rates (Tullai et al., 2007). In contrast, secondary response genes often lack CpG-island promoters and require stimulus-induced transcription factors that can recruit chromatin remodeling complex to initiate

chromatin remodeling (Ramirez-Carrozzi et al., 2009; Ramirez-Carrozzi et al., 2006). Besides mechanistic differences, primary response genes and secondary response genes also differ in their biological functions. In TLR4 signaling, many primary response genes encode signaling molecules, transcription factors, and effector molecules. Secondary response genes often encode molecules regulating more specific biological functions such as antigen presentation and T cell activation (Tong et al., 2016).

Primary and secondary response genes can be interdependent. Primary response genes can also encode activators of secondary response genes. For instance, LPS can activate the MAPK pathway and induce the primary response genes *Fos*, *Jun*, and *Atf3*, which encode AP1/ATF transcription factors (Tong et al., 2016; Bhatt et al., 2012). AP1/ATF transcription factors can potentiate the expression of several secondary response genes. LPS can also induce primary response gene *Ifnb1*, which encodes type I interferon. Type I interferon is a key immune-regulatory molecule that induces at least half of the secondary response genes regulating the interferon response (Raza et al., 2014; Amit et al., 2009).

However, the boundary between primary and secondary response genes is often blurred. Some genes can behave like both primary and secondary response genes. For example, *Gbp5* contains a canonical IRF binding site. In early TLR4 signaling, it is a primary response gene directly activated by IRF3. Later in TLR4 signaling, it can be activated as a secondary response gene by interferon-induced IRF9-STAT1-STAT2 complex (Ourthiague et al., 2015; Krapp et al., 2016). Limited by the cell-specific

expression of key regulators, one gene can be either primary or secondary response gene in different cell contexts. For instance, *Il6* is a primary response gene induced by LPS in macrophages (Ramirez-Carrozzi et al., 2009). But *Il6* is a secondary response gene in fibroblasts, which is sensitive to cycloheximide, an inhibitor of translation. As an initial clue to transcriptional regulatory mechanisms, separating primary and secondary response genes is often the first step of gene classification.

F. Selective Regulation of Transcriptional Cascades in Inflammation

Almost all cells in their native environment are constantly encountering extracellular and intracellular stimuli. How cells react to these stimuli depends on cell context and the nature of stimuli (Figure 1-2). It is believed that the lineage-determining factors and stimulus-regulated transcription factors together shape the transcriptional networks in macrophages (Glass and Natoli, 2016; Ostuni and Natoli, 2013; Heinz et al., 2010). During cellular development, the environmental cues will elicit the expression of lineage-determining transcription factors. Some of these factors can act as pioneer factors, which can recruit chromatin-remodeling complexes, bind extensively to many cell-specific *cis*-regulatory regions, and activate key regulators to promote lineage commitment. Subsequent environmental factors can further shape cell identity by inducing polarizing factors, which can further differentiate precursor cells into certain cell types. By opening chromatin or modifying other chromatin features, both pioneer factors and polarizing factors can establish cell-specific *cis*-regulatory elements including enhancers and promoters. Once the mature cells enter peripheral tissues, peripheral signals can activate stimulus-specific effector transcription factors, which can bind to

cell-specific *cis*-regulatory elements established during development. Different combinations of effector transcription factors can activate unique sets of genes for stimulus-specific responses. Deciphering these selective regulation mechanisms is essential for understanding the logic of selective gene regulation and developing cell-specific or gene-specific therapeutics.

G. Cell-Specific Regulation of Transcription

Cell identity is continually evolving in changing microenvironments. In bone marrow, hematopoietic stem cells can differentiate into macrophages in the presence of many differentiating factors including colony-stimulating factors, chemokines, morphogens, and contact signals (Ostuni and Natoli, 2013; Sánchez-Martín et al., 2011; Amit et al., 2016). In the presence of tissue-specific factors, tissue-resident macrophages will undergo transcriptional and phenotypic changes to adapt to the local microenvironment. For example, peritoneal retinoic acids drive unique expression signatures of the peritoneal macrophages; brain TGF β induces genes regulating microglia-specific functions (Gosselin et al., 2014; Amit et al., 2016). Besides tissue-specific phenotypes, cytokines can prime and polarize macrophages into M1 and M2 macrophages (Martinez and Gordon, 2014; Benoit et al., 2008; Mantovani and Locati, 2009). IFN γ can polarize macrophages into M1 type macrophages. M1 macrophages are pro-inflammatory, typically expressing genes regulating antibacterial functions such as *Nos2*. On the other hand, IL-4 can polarize macrophages into M2 type macrophages. M2 macrophages are anti-inflammatory, expressing genes regulating tissue repair such as *Arg1*. Besides transcriptional control, environmental factors also employ epigenetic strategies to shape

cell identities.

DNA methylation. DNA methylation is an epigenetic modification that can regulate gene transcription. In vertebrates, DNA methylation usually occurs at the cytosine of CpG dinucleotide except CpG islands at the promoter. Alteration of DNA methylation by developmental and environmental stimuli can cause changes in gene expression and cellular functions (Philibert et al., 2012). During blood formation, methylome changes are correlated with lineage-specific gene expression and lineage commitment (Rönnerblad et al., 2014; Reinius et al., 2012). Knockout of DNA methyltransferase 1 impaired lineage development (Bröske et al., 2009). Loss of DNA methylation often prevents heterochromatin formation, leading to the chromatin accessibility, binding of key transcription factors, and marking of active histone. During macrophage differentiation, DNA de-methylation promotes the activation of macrophage-specific *cis*-regulatory elements and expression of macrophage-specific genes (Wallner et al., 2016). DNA methylation also plays a role in shaping cell identity by environmental signals. For example, DNA methylation participates in M1/M2 macrophage polarization. In M1 macrophages, pro-inflammatory genes have hypo-methylated promoters, and anti-inflammatory genes have hyper-methylated promoters. This methylation pattern is correlated with the skewed expression of pro-inflammatory genes in M1 macrophages. M2 macrophages show the opposite methylation and expression patterns at the promoters of pro-inflammatory and anti-inflammatory genes (Babu et al., 2015). Thus, DNA methylation can refine cell identity by altering the chromatin features and gene expression during development and environmental stimulation.

Pioneer factors and polarizing factors. Pioneer factors often refer to transcription factors that can bind to closed chromatin and modify chromatin features during development. Pioneer factor binding and the consequent chromatin changes can facilitate or repress the binding of other transcription factors at key *cis*-regulatory elements, leading to transcriptional and phenotypic changes. For example, during hematopoiesis, progenitor cells undergo significant changes in transcription factor binding, chromatin features, and gene expression (Kowalczyk et al., 2015; Lara-Astiaso et al., 2014). Pioneer factors play a pivotal role in maintaining cell identity through chromatin regulation. In macrophages and B cells, PU.1 functions as a pioneer factor binding to thousands of genomic sites regulating cell-specific gene expression (Ghisletti et al., 2010). PU.1 can prevent heterochromatin formation and maintain open chromatin by interacting with p300/CBP (Nerlov and Graf, 1998; Yamamoto et al., 1999; Tagore et al., 2015). PU.1 can establish *de novo* enhancers with H3K4me1 and H3K27ac marks. Some of these enhancers are constitutively active by directly regulating macrophage-specific genes; some enhancers are inducible by recruiting other stimulus-induced transcription factors such as NFκB, AP1, and IRF. PU.1 can also collaborate with other stimulus-induced factors and induce transcriptional and phenotypic changes to adapt to the local environment. For example, in peritoneal macrophages, PU.1 can collaborate with GATA-6, a transcription factor induced by retinoic acids in the peritoneal cavity, and promote the expression of genes regulating gut IgA production (Glass and Natoli, 2016; Rosas et al., 2014; Gautier et al., 2014). Besides pioneer factors expressed during development, polarizing factors can also regulate gene expression through modifying chromatin features at *cis*-regulatory elements. Polarizing factors are often transcription

factors that are induced by microenvironment signals. In M2 macrophage polarization, IL-4-induced STAT6 can activate the H3K27 demethylase Jmjd3. Jmjd3 removes H3K27me_{2/3} marks on inactive *cis*-regulatory elements and promotes the expression of M2 marker genes (Ishii et al., 2009). Thus, by establishing novel *cis*-regulatory elements, both pioneer factors and polarizing factors can shape cell identity by acting alone or collaborating with other stimulus-induced transcription factors.

H. Stimulus-Specific Regulation of Transcription

Environmental stimuli can activate different receptors to promote stimulus-specific cellular responses. The activated receptors use different combinations of signaling molecules and transcription factors to induce genes regulating stimulus-specific responses. For example, different TLRs selectively activated anti-bacterial or anti-viral genes by using different adaptors and downstream signaling molecules. The membrane receptors TLR1, TLR2, TLR5, and TLR6 use adaptors MyD88 to activate the MAPK and NFκB pathways, inducing pro-inflammatory genes encoding cytokines, chemokines, and anti-bacterial peptides. The intracellular receptor TLR3 uses TRIF to activate the NFκB and IRF pathways, inducing pro-inflammatory genes and type I interferon response genes. TLR4 uses both MyD88 and TRIF adaptors to activate pro-inflammatory and type I interferon response genes. However, the limited number of transcription factors cannot fully explain the differential expression induced by different stimuli. While there are around 30,000 genes in the mouse genome, there are fewer than 3,000 transcription factors. Therefore, cells might leverage the limited number of transcription factors for selective gene activation using other strategies. These strategies are often

prevalent for regulating important transcription factors and important genes.

In TLR signaling, some key transcription factors can undergo multiple post-translational modifications that affect their activity and stability. One example is IRF3. It can induce type I IFN production in activated macrophages. Depending on the specific stimuli, IRF3 can have at least three different types of post-translational modifications. In TLR4 signaling, TBK1 or IKK ϵ can phosphorylate IRF3 C-terminal Ser/Thr clusters. IRF3 phosphorylation increases the negative charges and induces structural changes. This leads to IRF3 dimerization, nuclear translocation, DNA binding, and interaction with coactivators CBP/p300 (Panne et al., 2007; Dragan et al., 2007). IRF3 activity is also negatively regulated by SUMOylation at Lys152. Mutation of the SUMOylation site or the SUMO-conjugating enzyme UBC9, enhances transcriptional activity of IRF3 and elevates type I interferon production (Kubota et al., 2008; Decque et al., 2015). Additionally, IRF3 stability is regulated by polyubiquitination. Pin1 can bind to phosphorylated IRF3 and promote IRF3 degradation by proteasomes, resulting in a transient expression of type I IFN (Saitoh et al., 2006; Lin et al., 1998). On the other hand, HERC5 can prevent Pin1 binding and stabilize IRF3 by coupling ISG15 to IRF3, which results in a sustained expression of type I interferon and robust antiviral response (Shi et al., 2010). Therefore, by carefully balancing different post-translational modifications of IRF3, cells can precisely regulate the temporal expression of type I interferon and anti-viral response genes.

Besides regulating key transcription factors, selective regulatory mechanisms also

include the combinatorial interaction of transcription factors. Combinatorial regulation is widely used to regulate diverse biological functions including immune responses, cell proliferation and differentiation, and vascular development. One typical example of combinatorial regulation involves the transcription factor SRF. Although SRF is widely expressed, it can selectively induce transcription by interacting with two different cofactors: TCF transcription factors and MRTF transcription factors. Upon MAPK activation, SRF can partner with TCF to rapidly activate many primary response genes including *c-fos*, *Egr1*, and *Nr4a1* (Treisman et al., 1992; Gregg and Fraizer, 2011; Costello et al., 2004). SRF and TCF motifs have short and flexible spacing (Treisman et al., 1992; Posern and Treisman, 2006). At the promoter of *c-fos*, TCF binds to an Ets motif with a short core motif GGA; SRF binds to a core motif CCATATTAGG. The distance between Ets motif and SRF motif on *c-fos* promoter is only 3bp, thus allowing SRF to form a ternary complex with TCF (Mo et al., 2001). The cooperative binding of SRF and TCF stabilizes the binding of both factors on DNA (Kukushkin et al., 2002; Treisman et al., 1992). The binding of the complex causes a stronger bending of DNA towards SRF than SRF binding alone (Hassler, 2001). This probably allows SRF to make more extensive contact with DNA in the ternary complex than SRF binding alone. Both SRF and TCF are required for transcriptional activation. Mutation of either site will abolish gene induction. Another SRF cofactor, MRTF, can compete with TCF for the same binding domain on SRF (Murai and Treisman, 2002; Wang et al., 2004). Unlike TCF, MRTF does not bind to DNA in a sequence-specific manner. It indirectly binds to SRF motif through SRF recruitment.

IFNB1, encoding type I interferon in human, is a primary response gene induced by TLR4 signaling. It is also an autocrine molecule that triggers half of the secondary response genes, which regulate the interferon response and antiviral response. IFNB1 transcription relies on the cooperative binding of three different transcription factors: ATF-2/c-Jun, IRF3/IRF7, and NFκB. Their binding sites on IFNB1 promoter comprise a tightly packed 55-bp motif including one AP1 site, two IRF sites, and one NFκB site. This motif is highly conserved in mammals, suggesting that they are functionally important. Although these transcription factors lack physical contact in crystal structure, they can bind cooperatively to the promoter and form a large functional complex called the enhanceosome (Panne et al., 2007; Thanos et al., 1993). The formation of the enhanceosome requires the architectural protein HMGI(Y) and the orderly recruitment of histone acetylase, chromatin remodeling complexes, and transcription initiation machinery (Agalioti et al., 2000; Merika et al., 1998). In embryonic stem cells, Oct4 and Sox2 transcription factors can also bind cooperatively to a composite site and regulate genes essential for maintaining pluripotency (Reményi et al., 2004). In antigen-presenting cells, CIITA complexes with RFX, X2BP, and NF-Y transcription factors, acting as a master control factor of MHCII genes (Ludigs et al., 2015; LeibundGut-Landmann et al., 2004). In endothelial cells, Forkhead and Ets transcription factors bind cooperatively to a highly conserved Ets:Fox motif composite and selectively activate genes regulating vascular development (Val et al., 2008).

Combinatorial regulation requires fulfilling rules in motif, dimerization, and post-translational modification. In combinatorial regulation, transcription factors often need

strong motifs for binding, but their motifs may also diverge slightly from the canonical motifs for optimal protein interaction. And because the cooperative binding may stabilize protein binding in the complex, transcription factors can sometimes bind to weak motifs. For example, ternary complex formation allows TCF binding to a suboptimal Ets site near SRF site (Treisman et al., 1992). Short motif spacing is often required for physical interaction between transcription factors. Some have strict motif spacing, but some have flexible spacing. For example, the cooperative binding of Forkhead factor and Ets factor requires no spacing between the two motifs (Val et al., 2008). In contrast, TCF interacts with SRF through a flexible linking domain and forms a complex with SRF with variable spacing (Treisman et al., 1992). Besides motif requirements, combinatorial regulation might also prefer selective dimers. For example, the AP1 site on human IFNB1 enhanceosome prefers the AP1 dimer formed by c-Jun and ATF2. A selective dimer further ensures the precise of transcriptional activation by selective stimuli. Lastly, combinatorial regulation can also depend on selective translational modifications of transcription factors. For example, SRF and TCF can form a ternary complex on DNA prior to stimulation, but the complex is not activated until TCF phosphorylation by MAPK (Herrera et al., 1989; Gineitis and Treisman, 2001).

In eukaryotic cells, DNA wraps around histone proteins and assembles into a unit called a nucleosome. Each nucleosome is composed of 146bp DNA wrapped around an octamer consists of two copies each of H2A, H2B, H3, and H4. Between neighboring nucleosomes, there is usually a 10-50bp linker DNA bound by histone H1. To pack 3 billion base pairs of human genome into 2nm fibers, nucleosomes also fold into higher

order of structures (Woodcock and Ghosh, 2010; Luger et al., 2012). Because many *cis*-regulatory elements are buried in the nucleosome structure, protein binding to these inaccessible sites requires chromatin-remodeling complexes such as SWI/SNF. They can open chromatin by sliding or removing nucleosomes (Venkatesh and Workman, 2015; Clapier and Cairns, 2009). Histone modifiers can covalently modify the histone residues and alter chromatin structure by changing the electric charges of key residues or recruiting other histone regulators (Venkatesh and Workman, 2015; Bannister and Kouzarides, 2011). These chromatin regulation strategies increase the complexity of selective transcriptional regulation.

Chromatin accessibility is largely affected by the intrinsic properties of promoters. About 70% of mammalian promoters contain CpG islands (Saxonov et al., 2006). Because CpG-island promoters are too rigid to form stable nucleosomes, they are more likely to be nucleosome-free and accessible to protein binding (Ramirez-Carrozzi et al., 2009; Tazi and Bird, 1990; Choi, 2010). Thus, high CpG content is often sufficient to maintain open chromatin at promoters. Because they do not require chromatin remodeling, these CpG-island promoters are frequently associated with ubiquitously expressed genes. Almost all housekeeping genes possess CpG-island promoters (Zhu et al., 2008). CpG-island promoters often contain the active promoter mark H3K4me3 and a paused Pol II, which is often associated with high basal expression (Hargreaves et al., 2009). Conversely, low CpG-island promoters are more likely to form stable nucleosomes, thus preventing transcription factor binding and gene activation. Low CpG-island promoters often rely on pioneer factors or stimulus-induced factors to recruit chromatin-remodeling

complexes and histone modifiers. For example, interferon- γ can synergize with TLR agonists to activate pro-inflammatory genes by inducing the transcription factors STAT1 and IRF1. STAT1 and IRF1 recruit remodeling complexes to low CpG-island promoters, resulting in TLR-activated transcription factors binding to accessible promoters and transcriptional activation (Qiao et al., 2013; Pattenden et al., 2002). Because low CpG-island promoters require chromatin remodeling, they often have delayed transcriptional activation.

In TLR signaling, chromatin regulation also plays a role in separating primary and secondary response genes regulating different biological functions. Many primary response genes containing CpG-island promoters encode transcription factors and effector molecules. Because they do not require chromatin remodeling, they can be quickly induced to activate secondary response genes or auto-regulate primary response genes. Conversely, many secondary response genes encode cytokines regulating adaptive immune functions. They possess low CpG-island promoters that require chromatin remodeling, leading to delayed transcriptional induction (Ramirez-Carrozzi et al., 2009).

Cell-specific and stimulus-specific regulatory mechanisms often act collaboratively to regulate key inflammatory genes. One example is the *IL12b* gene. It encodes IL-12p40, a subunit shared by IL-12 and IL-23. IL-12p40 can dimerize with the constitutively expressed IL-12p35 subunit to form IL-12, a cytokine that bridges the innate and adaptive immunity by promoting Th1 cell development (Trinchieri, 2003; Trinchieri,

1995). IL-12p40 can also dimerize with IL-23p19 to form IL-23, a cytokine essential for Th17 development (Vignali and Kuchroo, 2012; Teng et al., 2015). *Il12b* is inducible by cytokine and pattern recognition receptor agonists only in myeloid cells. During hematopoiesis, PU.1 binds to the Ets site of *Il12b* enhancer and maintains an accessible enhancer. In cells lacking PU.1, polycomb repressive complex 2 binds to the *Il12b* enhancer and establishes heterochromatin with the repressive histone mark H3K27me3 (Tagore et al., 2015). The PU.1-established enhancer ensures the cell-specific expression of *Il12b*. *Il12b* gene transcription requires activation of both the promoter and the enhancer. Upon LPS stimulation, Oct proteins and CEBP β bind to the *Il12b* enhancer to induce *Il12b* transcription (Zhou et al., 2007; Bradley et al., 2003). *Il12b* enhancer activity is also negatively regulated by IL-10-induced NFIL3 (Smith et al., 2011). IRF5 can facilitate the synergy between the *Il12b* enhancer and promoter by binding to the IRF sites on both the enhancer and the promoter of *Il12b* (Koshiba et al., 2013). These transcription factors have differential roles in regulating *Il12b* transcription. While CEBP β mutation does not strongly affect *Il12b* induction, Oct and IRF5 mutation severely impaire *Il12b* induction (Takaoka et al., 2005; Zhou et al., 2007; Bradley et al., 2003). The *Il12b* promoter has binding sites for CEBP, AP1, and NF κ B. Among them, the NF κ B site is most critical for the transcriptional induction of *Il12b*. The non-canonical NF κ B site at the *Il12b* promoter has much a stronger preference for c-Rel:p50 dimer than other NF κ B dimers. Mutating c-Rel abolished the expression of *Il12b*, but RelA mutation hardly affected *Il12b* expression (Sanjabi et al., 2005). While the *Il12b* enhancer is constitutively open, *Il12b* promoter lacks CpG islands and requires TLR-activated chromatin remodeling for transcriptional activation (Weinmann et al., 2001;

Weinmann et al., 1999). However, only about 20% of the macrophages undergo chromatin remodeling at promoters after LPS stimulation (Gjidoda et al., 2014). This suggests that nucleosomes might compete with transcription factors for DNA binding. Interestingly, although c-Rel knockout cells have diminished *I12b* induction, it does not affect promoter chromatin remodeling (Weinmann et al., 2001). This indicates that c-Rel binding is a downstream event of chromatin remodeling. Although there might be other as-yet-unknown mechanisms regulating *I12b*, *I12b* represents a complicated selective regulation model combining cell-specific and stimulus-specific mechanisms, which involves transcription factor regulation, chromatin regulation, and promoter-enhancer interaction.

Cell-specific and stimulus-specific regulation can be interdependent and can share similar mechanisms. Pioneer factors can collaborate with stimulus-induced transcription factors at cell-specific *cis*-regulatory elements to activate transcription; and stimulus-induced factors can also change chromatin status and further shape cell identity. For example, differentiated macrophages maintain a certain degree of plasticity. Upon repeated bacterial infection, macrophages can develop trained immunity or immune tolerance (Netea et al., 2016; Ifrim et al., 2014). LPS-activated transcription factors can bind to intergenic sites lacking enhancer marks and establish *de novo* enhancers by promoting the removal of repressive histone marks and the deposition of H3K4me1 marks (Ostuni et al., 2013; Yoshida et al., 2015). These *de novo* enhancers are also called latent enhancers. Upon the second LPS challenge, they can induce transcription more robustly and rapidly than the first challenge. Thus, macrophages are able to gain

partial short-term memory to certain environmental stimuli. On the other hand, repeated LPS stimulation can also cause weaker or no response, a phenomenon called LPS tolerance. LPS can reduce histone acetylation at *cis*-regulatory elements of some induced genes, leading to the non-response or tolerance to the subsequent LPS stimulation (Hargreaves et al., 2009). Trained immunity and immune tolerance demonstrate that stimulus-specific changes can also shape cell identity and overlap with the functions of cell-specific regulatory mechanisms.

I. Selective Regulation of NFκB

NFκB transcription factors play a pivotal role in innate and adaptive immunity. They are widely expressed in many cell types and can be activated by a diverse range of stimuli. Dysregulation of NFκB transcription factors is associated with many chronic inflammatory diseases and other diseases including cancer, Alzheimer's disease, diabetes, and atherosclerosis (Hoesel and Schmid, 2013; Patel and Santani, 2009; Pamukcu et al., 2011).

NFκB family transcription factors consist of five members: Rel A, RelB, c-Rel, p50, and p52. Structurally, all NFκB members share a Rel homology domain (RHD), which is responsible for interaction with IκB inhibitor proteins, dimerization, nuclear localization, and DNA binding (Hoesel and Schmid, 2013; Lawrence, 2009; Zhang et al., 2017). While Rel A, RelB, and c-Rel share an additional transcriptional activation domain, p50 and p52 precursors share many copies of ankyrin repeats in their C-terminal domains. These sequences are responsible for self-inhibition. Because p50 and p52 lack

transcription activation domains, their homodimers can bind to NFκB sites as dominant negative proteins to inhibit transcription of NFκB targets. NFκB members can form 15 dimeric species that can bind to DNA. The most important dimers in the immune response are RelA:p50, c-Rel:p50, and RelB:p52. NFκB dimers can bind to a consensus κB motif: 5'-GGGRN(Y)YYCC-3' (Hoffmann and Baltimore, 2006). However, different dimers favor slightly different κB motifs (Wang et al., 2012). Protein binding microarrays reveal three classes of consensus κB motifs. RelA and c-Rel homodimers prefer binding to a 9-bp motif; heterodimers prefer binding to a 10-bp motif; and p50 and p52 homodimers prefer binding to an 11-bp or a 12-bp motif (Siggers et al., 2012). The differential binding of dimers represents one of the strategies to selectively regulate NFκB targets. For example, the *I12b* promoter has a non-canonical NFκB site that favors c-Rel:p50 binding. While *I12b* induction was abolished by c-Rel mutation, it was not affected by the mutation of other NFκB members (Sanjabi et al., 2005; Tong et al., 2016).

NFκB transcription factors are regulated by the canonical and non-canonical pathways. In the canonical pathway, NFκB is sequestered in the cytoplasm by inhibitory IκB proteins including IκBα, IκBβ, IκBγ, and IκBε. IκB proteins mask the nuclear translocation signals of NFκB and thereby prevent nuclear translocation. Extracellular stimuli like pathogen components and cytokines can activate the IKK complex, which consists of IKKα, IKKβ, and regulatory subunit IKKγ. The active IKKβ can phosphorylate IκB proteins, inducing IκB protein ubiquitination and proteasome degradation. The degradation of IκB proteins releases NFκB, allowing it to dimerize, translocate into

nucleus, and bind to NF κ B sites regulating pro-inflammatory genes (Shih et al., 2011; Hayden and Ghosh, 2012). In the non-canonical pathway, p52 precursor, p100, remains in the cytoplasm because of self-inhibitory sequences at the C-terminus. TNF family cytokines such as CD40L, BAFF, and RANKL, can activate a complex composed of NIK and IKK α . Activated IKK α can phosphorylate p100 and induce removal of these inhibitory sequences. The processed p100 releases p52, which can dimerize with RelB and translocate into the nucleus (Sun, 2011; Cildir et al., 2016). While the canonical pathway activates genes regulating host response to pathogen infection, the non-canonical pathway often controls genes regulating lymph-organogenesis, B cell maturation, and bone metabolism.

NF κ B transcription factors are regulated at many steps of gene expression including transcription and post-translational modification steps. NF κ B members and I κ B proteins can induce their own transcription in an auto-regulatory manner (Tong et al., 2016). The rapid induction of I κ B proteins can export nuclear NF κ B to the cytoplasm and replenish the cytoplasmic pool of NF κ B proteins. This results in the transient expression of NF κ B targets, but it also ensures robust cellular response to repeated stimuli (Kearns et al., 2006; Nelson et al., 2004; Kellogg and Tay, 2015). NF κ B proteins also undergo different post-translational modifications that positively or negatively regulate NF κ B activity. For example, RelA Ser-276 can be phosphorylated by the cytoplasmic protein kinase A or the nuclear MSK1. Ser-276 phosphorylation can prevent the folding of two terminal domains and facilitate the interaction with the co-activator p300/CBP (Perkins, 2006; Zhong et al., 2002). Phosphorylation of Thr-254 of RelA is required for the binding of

Pin1, which can activate RelA by promoting RelA dissociation from I κ B α . RelA can also be acetylated by p300/CBP at multiple sites including Lys-281, Lys-221, and Lys-310. While Lys-281 and Lys-221 prevent nuclear I κ B α from exporting of RelA, Lys-310 enhances the transcriptional activity of RelA (Chen et al., 2002). Post-translational modifications can also negatively regulate RelA. For example, RelA methylation by Set9 or ubiquitination by SOCS-1 can result in proteasome-mediated protein degradation and termination of transcription (Yang et al., 2009; Ryo et al., 2003; Huang et al., 2010). Thus, by combining post-translational modifications and interactions with co-factors, NF κ B can both broadly and selectively regulate hundreds of pro-inflammatory genes.

J. Serum Response Factor (SRF) and Combinatorial Regulation

Serum response factor (SRF) is a ubiquitously expressed transcription factor that regulates a wide variety of biological functions including cell proliferation and differentiation, circadian rhythm, cytoskeleton, cell migration, muscle cell development, and inflammation (Olson and Nordheim, 2010; Posern and Treisman, 2006; Knöll and Nordheim, 2009). SRF mutation is embryonic lethal during gastrulation. Tissue-specific deletion of SRF also causes defects in vascular development, skeletal muscle development, neuronal development, B cell and T cell maturation, and macrophage functions (Miano, 2010). It is interesting that such a ubiquitously expressed transcription factor can selectively regulate genes involved in diverse functions in different cell types.

SRF belongs to the MADS domain protein family that is conserved from yeast to human. It binds to DNA as a homodimer and induces a 72° bending of DNA towards SRF (Pellegrini et al., 1995; Mo et al., 2001). SRF binds to a consensus motif named CArG

box: CC[A/T]₆GG. Degenerate CArG box motifs are widespread in the genome. Computational analyses predict more than 3 million such CArG box motifs in human genome (Sun et al., 2006). SRF ChIP-seq, however, revealed only thousands of peaks in each cell type (Kim et al., 2010; Sullivan et al., 2011; Esnault et al., 2014). This suggests the functional binding of SRF may be regulated by other mechanisms such as chromatin regulation. SRF can be regulated by transcriptional induction, post-translational modification, and cofactor interaction. SRF is constitutively expressed, and it can activate its targets without new protein synthesis. But its transcription can also be further induced by Rho-actin pathway activators (Esnault et al., 2014; Misra et al., 1991). SRF has two phosphorylation sites: Thr-159 and Thr-162. Phosphorylation of Thr-159 in the MADS domain is required for transcriptional activation of the α -actin gene (Iyer et al., 2003). Phosphorylation of Thr-162 prevents SRF binding at the actin promoter, but it does not prevent SRF binding at the promoter of *c-fos* (Iyer et al., 2006). This is probably because SRF activity requires different cofactors at these two genes. And different phosphorylation can affect the choice of cofactor to selectively activate SRF targets.

SRF activity is mainly regulated through cofactor interaction (Figure 1-3). SRF can interact with two cofactors: ternary complex factor (TCF) and myocardin-related transcription factor (MRTF) transcription factors. TCF transcription factors consist of three members: SAP-1, Elk1, and Net. TCF belongs to the ETS family transcription factors. In the presence of nearby SRF sites, TCF can bind cooperatively with SRF and form a ternary complex on DNA in the absence of stimulation (Posem and Treisman,

2006; Herrera et al., 1989). Because TCF contacts SRF through a flexible domain, they can form a complex with variable spacing between the two binding sites (Treisman et al., 1992). The MAPK pathway can activate TCF in the ternary complex through phosphorylation. Active TCF can further recruit CBP/p300, histone modifiers, and transcription initiation machinery for transcriptional activation (Ramirez et al., 1997; Esnault et al., 2017). The MAPK-TCF pathway can be activated by a broad range of stimuli including mitogens, cytokines, UV irradiation, and TLR agonists (Chai and Tarnawski, 2002). SRF and TCF control the expression of many primary response genes that regulate cell proliferation and differentiation, such as *c-fos*, *Egr1*, and *Nr4a1*.

Myocardin-related transcription factors (MRTFs) belong to another family of SRF cofactors. This family consists of myocardin, MRTFA, and MRTFB. Myocardin is selectively expressed in muscle cells. Myocardin is constitutively active and controls the expression of many muscle-specific genes (Wang et al., 2003). MRTFA and MRTFB are expressed ubiquitously in many cell types. MRTF is sequestered by monomeric actin in the resting state. Activation of the Rho-actin pathway induces actin polymerization, which dissociates actin from MRTF. MRTF then translocates to the nucleus and complexes with SRF (Miralles et al., 2003; Olson and Nordheim, 2010). Although MRTF also makes DNA contact in the SRF-MRTF complex, it does not bind to DNA in a sequence-specific manner; it is recruited to the SRF motif by SRF (Zaromytidou et al., 2006). The MRTF-SRF complex regulates genes involved in actin cytoskeleton function, cell migration, and muscle function (Olson and Nordheim, 2010; Medjkane et al., 2009). Because MRTF and TCF bind to a common interface on SRF, they can compete for

SRF interaction, resulting in either the activation or repression of transcription (Zaromytidou et al., 2006; Murai and Treisman, 2002; Wang et al., 2004). The competition of different cofactors enables SRF to integrate multiple signals and selectively activate transcription.

In conclusion, a proper immune response requires a fine balance between host defense and homeostasis. Clarifying the mechanisms underlying the immune activation will improve our understanding of the immune system and will shed light on treating immune-related diseases. Because innate immunity orchestrates a chain of innate and adaptive immune responses during inflammation, we began our studies on the innate immune response. Innate immune cells can sense pathogen components through pattern recognition receptors, which reside in different cellular compartments and trigger potent transcriptional induction integrating discrete signaling pathways, chromatin regulators, and transcription factors. The highly coordinated events regulating the transcriptional cascades require both common and gene-specific strategies that allow innate immune cells to respond in cell-specific and stimulus-specific manners. Although many studies have investigated the transcriptional mechanisms of different pattern recognition receptors, they relied on statistics and conventional systems approaches that focus on common regulatory mechanisms. Their approaches assume that one transcription factor or pathway can regulate many genes, and combinatorial regulation leads to gene-specific regulation. However, their statistic methods require large sample sizes, and therefore, they often suffer from the loss of selective mechanisms that only regulate a small number of genes. In chapter 2, I describe a stringent system approach

to classify primary and secondary response genes. This approach demonstrates that key inflammatory genes are regulated by highly selective mechanisms involving chromatin regulation and combinatorial regulation. In chapter 3, I examine a combinatorial regulation model of SRF and TCF transcription factors and identify the strict motif requirements for cooperative binding and transcriptional induction. Taken together, our stringent systems approach can complement the conventional systems approach by uncovering selective mechanisms and by elucidating deeper mechanistic details in complicated contexts.

Figure Legends

Figure 1-1: TLR4 Signaling Pathways

LPS-activated TLR4 signaling is mediated through two adaptors: TIRAP/MyD88 and TRIF. Dimerization of TLR4 recruits TIRAP/MyD88, which in turn recruits IRAKs, TRAF6, TAB1/2/3, TAK1, IKK α , IKK β , and the regulatory subunit IKK γ . Active IKKs can phosphorylate I κ Bs and activate the NF κ B pathway. Active TAK1 in the cytoplasm can activate the MAPK pathway. Through CD14-mediated activation of ITAM, Syk and PLC γ , TLR4 can internalize and transport to endosome. In endosome, TLR4 interacts with the adaptor TRAM/TRIF, which in turn recruits RIPK1, TRADD, caspase-8, FADD, IKK α , IKK β , and IKK γ . Active IKKs can induce the NF κ B pathway. TRAM/TRIF can also recruit TRAF3, which can interact with TANK, IKK γ , IKK ϵ , and TBK1. TBK1 can activate IRF3, a key transcription factor responsible for type I IFN production.

Figure 1-2: Cell-Specific and Stimulus-Specific Regulation of Gene Transcription

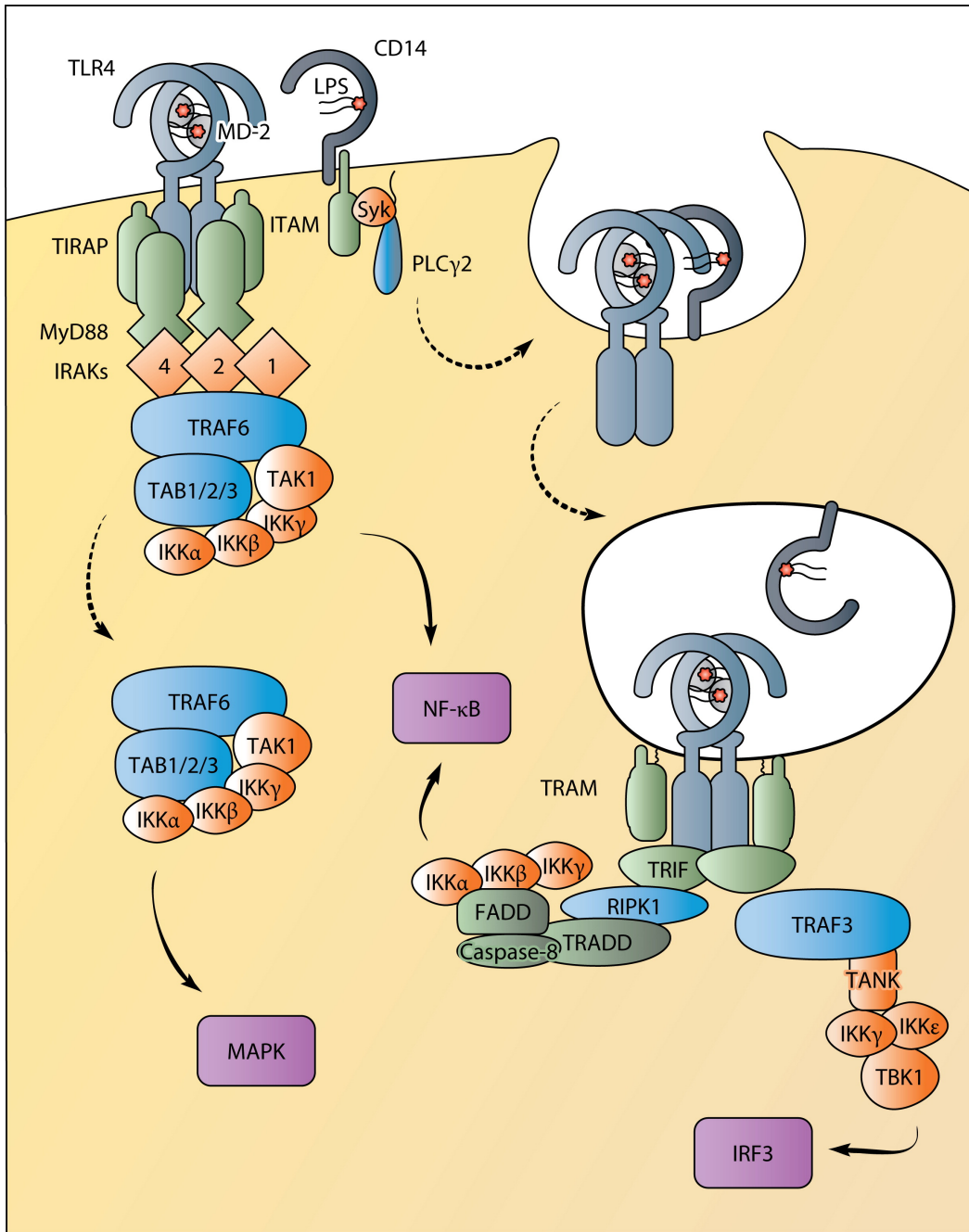
A model explains the cell-specific and stimulus-specific regulation of transcription. During cell development, micro-environmental cues induce the expression of pioneer transcription factors and polarizing transcription factors. They can promote cell-specific gene expression program and drive lineage commitment and cell differentiation. In the peripheral tissue, terminally differentiated cells can respond to extracellular and intracellular stimuli by inducing different effector transcription factors. These stimulus-specific transcription factors can induce gene expression programs resulting in reversible functional states.

Figure 1-3: SRF Regulated by Rho-MRTF and MAPK-TCF Pathways

SRF can partner with two cofactors: MRTF (MAL) activated by the Rho-actin pathway, and TCF activated by the MAPK pathway. Monomeric actin can bind to MRTF (MAL) and sequester MRTF (MAL) in the cytoplasm. Activation of the Rho family GTPases (RhoA, Rac, and Cdc42) induces actin polymerization and dissociates monomeric actin from MRTF (MAL). MRTF can translocate into the nucleus and complex with SRF to activate transcription. The MAPK pathway can phosphorylate and activate TCF. TCF binds to the Ets site near SRF site. TCF can bind cooperatively with SRF and form a ternary complex to activate transcription.

Figure 1-1: TLR4 Signaling Pathways

(Brubaker et al., 2015; License#: 4082740867171)




 Brubaker SW, et al. 2015. Annu. Rev. Immunol. 33:257–90

Figure 1-2: Cell-Specific and Stimulus-Specific Regulation of Gene Transcription

(Ostuni and Natoli, 2013; License#: 4082741209330)

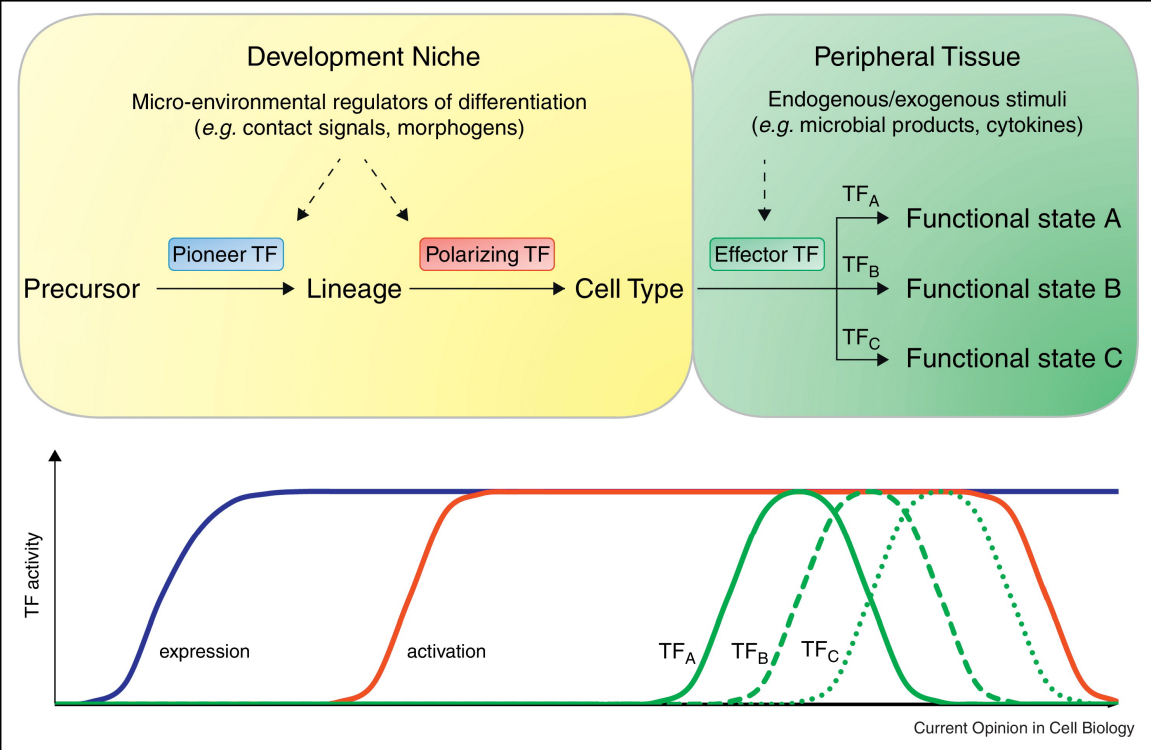
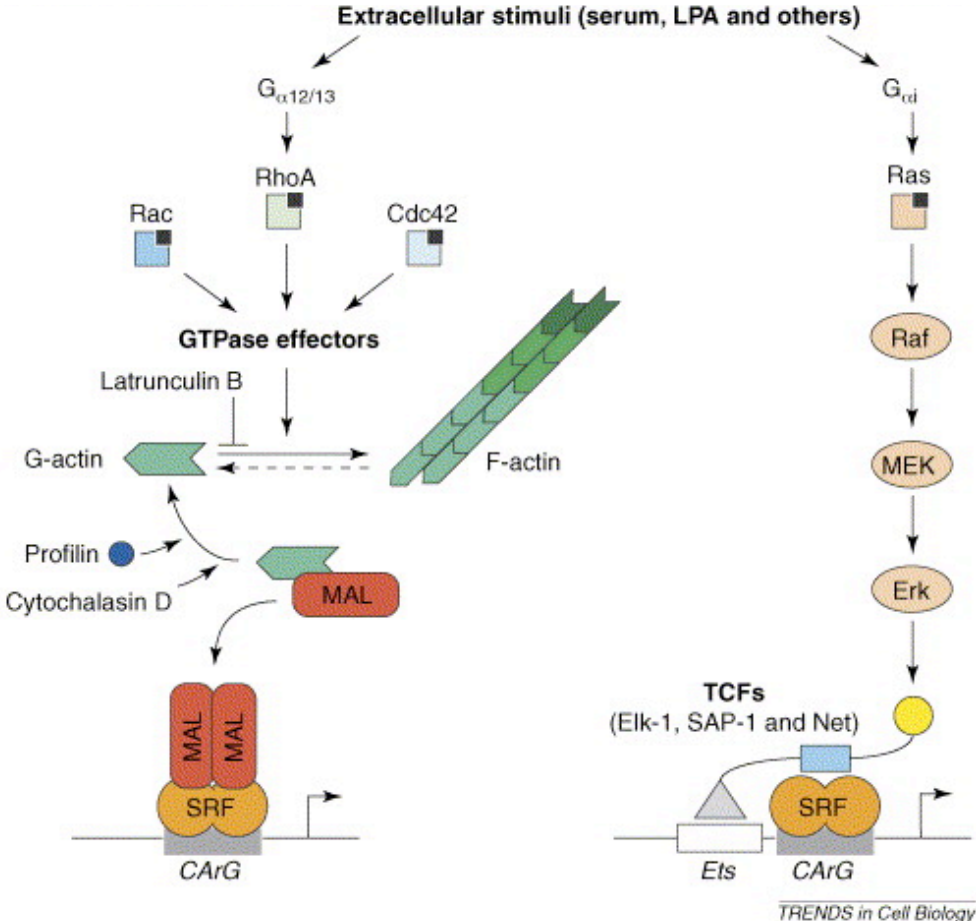


Figure 1-3: SRF Regulated by Rho-MRTF and MAPK-TCF Pathways

(Posern and Treisman, 2006; License#: 4082741323052)



Works Cited

- Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN promoter. *Cell* *103*, 667–678.
- Akira, S., and Takeda, K. (2004). Toll-like receptor signalling. *Nat Rev Immunol* *4*, 499–511.
- Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating the differential response to pathogens. *Science* *326*, 257–263.
- Amit, I., Winter, D.R., and Jung, S. (2016). The role of the local environment and epigenetics in shaping macrophage identity and their effect on tissue homeostasis. *Nat Immunol* *17*, 18–25.
- Babu, M., Devi, T.D., Mäkinen, P., Kaikkonen, M., Lesch, H.P., Junttila, S., Laiho, A., Ghimire, B., Gyenesei, A., and Ylä-Herttuala, S. (2015). Differential promoter methylation of macrophage genes is associated with impaired vascular growth in ischemic muscles of hyperlipidemic and type 2 diabetic mice. *Circulation Research* *117*, 289–299.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* *21*, 381–395.
- Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. (2015). Impact of regulatory variation from RNA to protein. *Science* *347*, 664–667.
- Benoit, M., Desnues, B., and Mege, J.-L. (2008). Macrophage polarization in bacterial infections. *The Journal of Immunology* *181*, 3733–3739.
- Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* *150*, 279–290.
- Boone, D.L., Turer, E.E., Lee, E.G., Ahmad, R.-C., Wheeler, M.T., Tsui, C., Hurley, P., Chien, M., Chai, S., Hitotsumatsu, O., et al. (2004). The ubiquitin-modifying enzyme A20 is required for termination of Toll-like receptor responses. *Nat Immunol* *5*, 1052–1060.
- Bradley, M.N., Zhou, L., and Smale, S.T. (2003). C/EBP β regulation in lipopolysaccharide-stimulated macrophages. *Mol. Cell. Biol.* *23*, 4841–4858.
- Bröske, A.-M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., et al. (2009). DNA methylation protects

hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet* *41*, 1207–1215.

Brubaker, S.W., Bonham, K.S., Zanoni, I., and Kagan, J.C. (2015). Innate immune pattern recognition: a cell biological perspective. *Annual Review of Immunology* *33*, 257–290.

Bürckstümmer, T., Baumann, C., Blüml, S., Dixit, E., Dürnberger, G., Jahn, H., Planyavsky, M., Bilban, M., Colinge, J., Bennett, K.L., et al. (2009). An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat Immunol* *10*, 266–272.

Cai, X., Chiu, Y.-H., and Chen, Z.J. (2014). The cGAS-cGAMP-STING pathway of cytosolic dna sensing and signaling. *Molecular Cell* *54*, 289–296.

Chai, J., and Tarnawski, A.S. (2002). Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. *J. Physiol. Pharmacol.* *53*, 147–157.

Chen, L., Mu, Y., and Greene, W.C. (2002). Acetylation of RelA at discrete sites regulates distinct nuclear functions of NFκB. *EMBO J* *21*, 6539–6548.

Choi, J.K. (2010). Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biology* *11*, R70.

Cildir, G., Low, K.C., and Tergaonkar, V. (2016). Noncanonical NF-κB signaling in health and disease. *Trends in Molecular Medicine* *22*, 414–429.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. *Annual Review of Biochemistry* *78*, 273–304.

Costello, P.S., Nicolas, R.H., Watanabe, Y., Rosewell, I., and Treisman, R. (2004). Ternary complex factor SAP-1 is required for Erk-mediated thymocyte positive selection. *Nat Immunol* *5*, 289–298.

Decque, A., Joffre, O., Magalhaes, J.G., Cossec, J.-C., Blecher-Gonen, R., Lapaquette, P., Silvin, A., Manel, N., Joubert, P.-E., Seeler, J.-S., et al. (2015). Sumoylation coordinates the repression of inflammatory and anti-viral gene-expression programs during innate sensing. *Nat Immunol* *advance online publication*.

Deleault, K.M., Skinner, S.J., and Brooks, S.A. (2008). Tristetraprolin regulates TNF TNF-α mRNA stability via a proteasome dependent mechanism involving the combined action of the ERK and p38 pathways. *Molecular Immunology* *45*, 13–24.

Dragan, A.I., Hargreaves, V.V., Makeyeva, E.N., and Privalov, P.L. (2007).

Mechanisms of activation of interferon regulator factor 3: the role of C-terminal domain phosphorylation in IRF-3 dimerization and DNA binding. *Nucleic Acids Res* 35, 3525–3534.

Eberl, G. (2016). Immunity by equilibrium. *Nat Rev Immunol* 16, 524–532.

Esnault, C., Gualdrini, F., Horswell, S., Kelly, G., Stewart, A., East, P., Matthews, N., and Treisman, R. (2017). ERK-Induced activation of TCF family of SRF cofactors initiates a chromatin modification cascade associated with transcription. *Molecular Cell* 65, 1081–1095.e5.

Esnault, C., Stewart, A., Gualdrini, F., East, P., Horswell, S., Matthews, N., and Treisman, R. (2014). Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. *Genes Dev.* 28, 943–958.

Fowler, T., Sen, R., and Roy, A.L. (2011). Regulation of primary response genes. *Molecular Cell* 44, 348–360.

Gautier, E.L., Ivanov, S., Williams, J.W., Huang, S.C.-C., Marcelin, G., Fairfax, K., Wang, P.L., Francis, J.S., Leone, P., Wilson, D.B., et al. (2014). Gata6 regulates aspartoacylase expression in resident peritoneal macrophages and controls their survival. *Journal of Experimental Medicine* 211, 1525–1531.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.-L., et al. (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* 32, 317–328.

Gineitis, D., and Treisman, R. (2001). Differential usage of signal transduction pathways defines two types of serum response factor target gene. *J. Biol. Chem.* 276, 24531–24539.

Gjidoda, A., Tagore, M., McAndrew, M.J., Woods, A., and Floer, M. (2014). Nucleosomes are stably evicted from enhancers but not promoters upon induction of certain pro-inflammatory genes in mouse macrophages. *PLOS ONE* 9, e93971.

Glass, C.K., and Natoli, G. (2016). Molecular control of activation and priming in macrophages. *Nat Immunol* 17, 26–33.

Gosselin, D., Link, V.M., Romanoski, C.E., Fonseca, G.J., Eichenfield, D.Z., Spann, N.J., Stender, J.D., Chun, H.B., Garner, H., Geissmann, F., et al. (2014). Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* 159, 1327–1340.

Gregg, J., and Fraizer, G. (2011). Transcriptional regulation of EGR1 by EGF and

the ERK signaling pathway in prostate cancer cells. *Genes Cancer* 2, 900–909.

Hansson, G.K., and Libby, P. (2006). The immune response in atherosclerosis: a double-edged sword. *Nat Rev Immunol* 6, 508–519.

Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* 138, 129–145.

Hassler, M. (2001). The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. *The EMBO Journal* 20, 3018–3028.

Hato, T., and Dagher, P.C. (2014). How the innate immune system senses trouble and causes trouble. *CJASN CJN.04680514*.

Hayden, M.S., and Ghosh, S. (2012). NF- κ B, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* 26, 203–234.

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. *Molecular Cell* 38, 576–589.

Herrera, R.E., Shaw, P.E., and Nordheim, A. (1989). Occupation of the c-fos serum response element in vivo by a multi-protein complex is unaltered by growth factor induction. *Nature* 340, 68–70.

Herschman, H.R. (1991). Primary response genes induced by growth factors and tumor promoters. *Annual Review of Biochemistry* 60, 281–319.

Hoesel, B., and Schmid, J.A. (2013). The complexity of NF- κ B signaling in inflammation and cancer. *Molecular Cancer* 12, 86.

Hoffmann, A., and Baltimore, D. (2006). Circuitry of nuclear factor κ B signaling. *Immunological Reviews* 210, 171–186.

Hoving, J.C., Wilson, G.J., and Brown, G.D. (2014). Signalling C-Type lectin receptors, microbial recognition and immunity. *Cell Microbiol* 16, 185–194.

Huang, B., Yang, X.-D., Lamb, A., and Chen, L.-F. (2010). Posttranslational modifications of NF- κ B: Another layer of regulation for NF- κ B signaling pathway. *Cellular Signalling* 22, 1282–1290.

Ifrim, D.C., Quintin, J., Joosten, L.A.B., Jacobs, C., Jansen, T., Jacobs, L., Gow, N.A.R., Williams, D.L., Meer, J.W.M. van der, and Netea, M.G. (2014). Trained immunity or tolerance: opposing functional programs induced in human monocytes after engagement of various pattern recognition receptors. *Clin. Vaccine Immunol.*

21, 534–545.

Ishii, M., Wen, H., Corsa, C.A.S., Liu, T., Coelho, A.L., Allen, R.M., Carson, W.F., Cavassani, K.A., Li, X., Lukacs, N.W., et al. (2009). Epigenetic regulation of the alternatively activated macrophage phenotype. *Blood* 114, 3244–3254.

Iwasaki, A., and Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. *Nat Immunol* 16, 343–353.

Iyer, D., Belaguli, N., Flück, M., Rowan, B.G., Wei, L., Weigel, N.L., Booth, F.W., Epstein, H.F., Schwartz, R.J., and Balasubramanyam, A. (2003). Novel phosphorylation target in the serum response factor MADS box regulates α -actin transcription. *Biochemistry* 42, 7477–7486.

Iyer, D., Chang, D., Marx, J., Wei, L., Olson, E.N., Parmacek, M.S., Balasubramanyam, A., and Schwartz, R.J. (2006). Serum response factor MADS box serine -162 phosphorylation switches proliferation and myogenic gene programs. *PNAS* 103, 4516–4521.

Janeway, C.A. (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol* 54, 1–13.

Jovanovic, M., Rooney, M.S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E.H., Fields, A.P., Schwartz, S., Raychowdhury, R., et al. (2015). Dynamic profiling of the protein life cycle in response to pathogens. *Science* 347, 1259038.

K R Yamamoto, and Alberts, and B.M. (1976). Steroid receptors: elements for modulation of eukaryotic transcription. *Annual Review of Biochemistry* 45, 721–746.

Kawai, T., and Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34, 637–650.

Kearns, J.D., Basak, S., Werner, S.L., Huang, C.S., and Hoffmann, A. (2006). I κ B ϵ provides negative feedback to control NF- κ B oscillations, signaling dynamics, and inflammatory gene expression. *J Cell Biol* 173, 659–664.

Kellogg, R.A., and Tay, S. (2015). Noise facilitates transcriptional control under dynamic inputs. *Cell* 160, 381–392.

Kim, H.M., Park, B.S., Kim, J.-I., Kim, S.E., Lee, J., Oh, S.C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O.J., et al. (2007). Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist eritoran. *Cell* 130, 906–917.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread

transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.

Knöll, B., and Nordheim, A. (2009). Functional versatility of transcription factors in the nervous system: the SRF paradigm. *Trends in Neurosciences* 32, 432–442.

Koshiba, R., Yanai, H., Matsuda, A., Goto, A., Nakajima, A., Negishi, H., Nishio, J., Smale, S.T., and Taniguchi, T. (2013). Regulation of cooperative function of the *Il12b* enhancer and promoter by the interferon regulatory factors 3 and 5. *Biochemical and Biophysical Research Communications* 430, 95–100.

Kotas, M.E., and Medzhitov, R. (2015). Homeostasis, inflammation, and disease susceptibility. *Cell* 160, 816–827.

Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* 25, 1860–1872.

Krapp, C., Hotter, D., Gawanbacht, A., McLaren, P.J., Kluge, S.F., Stürzel, C.M., Mack, K., Reith, E., Engelhart, S., Ciuffi, A., et al. (2016). Guanylate binding protein (GBP) 5 is an interferon-inducible inhibitor of HIV-1 infectivity. *Cell Host & Microbe* 19, 504–514.

Kubota, T., Matsuoka, M., Chang, T.-H., Tailor, P., Sasaki, T., Tashiro, M., Kato, A., and Ozato, K. (2008). Virus infection triggers SUMOylation of IRF3 and IRF7, leading to the negative regulation of type I interferon gene expression. *J. Biol. Chem.* 283, 25660–25670.

Kukushkin, A.N., Abramova, M.V., Svetlikova, S.B., Darieva, Z.A., Pospelova, T.V., and Pospelov, V.A. (2002). Downregulation of *c-fos* gene transcription in cells transformed by E1A and cHa-ras oncogenes: a role of sustained activation of MAP/ERK kinase cascade and of inactive chromatin structure at *c-fos* promoter. *Oncogene* 21, 719–730.

Lai, W.S., Parker, J.S., Grissom, S.F., Stumpo, D.J., and Blackshear, P.J. (2006). Novel mRNA targets for tristetraprolin (TTP) identified by global analysis of stabilized transcripts in TTP-deficient fibroblasts. *Mol. Cell. Biol.* 26, 9196–9208.

Lampron, A., ElAli, A., and Rivest, S. (2013). Innate immunity in the CNS: redefining the relationship between the CNS and its environment. *Neuron* 78, 214–232.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Chromatin state dynamics during blood formation. *Science* 345, 943–949.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159, 1312–1326.

Lawrence, T. (2009). The nuclear factor NF- κ B pathway in inflammation. *Cold Spring Harb Perspect Biol* 1, a001651.

LeibundGut-Landmann, S., Waldburger, J.-M., Krawczyk, M., Otten, L.A., Suter, T., Fontana, A., Acha-Orbea, H., and Reith, W. (2004). Mini-review: specificity and expression of CIITA, the master regulator of MHC class II genes. *Eur. J. Immunol.* 34, 1513–1525.

Li, J.J., and Biggin, M.D. (2015). Statistics requantitates the central dogma. *Science* 347, 1066–1067.

Lin, R., Heylbroeck, C., Pitha, P.M., and Hiscott, J. (1998). Virus-dependent phosphorylation of the IRF-3 transcription factor regulates nuclear translocation, transactivation potential, and proteasome-mediated degradation. *Mol. Cell. Biol.* 18, 2986–2996.

Loo, Y.-M., and Gale, M. (2011). Immune signaling by RIG-I-like receptors. *Immunity* 34, 680–692.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotech* 25, 117–124.

Ludigs, K., Seguí-Estévez, Q., Lemeille, S., Ferrero, I., Rota, G., Chelbi, S., Mattmann, C., MacDonald, H.R., Reith, W., and Guarda, G. (2015). NLRC5 exclusively transactivates MHC Class I and related genes through a distinctive SXY module. *PLOS Genet* 11, e1005088.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* 13, 436–447.

Mantovani, A., and Locati, M. (2009). Orchestration of macrophage polarization. *Blood* 114, 3135–3136.

Martinez, F.O., and Gordon, S. (2014). The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep* 6.

Medjkane, S., Perez-Sanchez, C., Gaggioli, C., Sahai, E., and Treisman, R. (2009). Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nat Cell Biol* 11, 257–268.

Medzhitov, R., and Horng, T. (2009). Transcriptional control of the inflammatory response. *Nat Rev Immunol* 9, 692–703.

Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN β enhanceosome is required for synergistic activation of transcription. *Molecular Cell* 1, 277–287.

Miano, J.M. (2010). Role of serum response factor in the pathogenesis of disease. *Lab Invest* 90, 1274–1284.

Misra, R.P., Rivera, V.M., Wang, J.M., Fan, P.D., and Greenberg, M.E. (1991). The serum response factor is extensively modified by phosphorylation following its synthesis in serum-stimulated fibroblasts. *Mol. Cell. Biol.* 11, 4545–4554.

Mo, Y., Ho, W., Johnston, K., and Marmorstein, R. (2001). Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex1. *Journal of Molecular Biology* 314, 495–506.

Motta, V., Soares, F., Sun, T., and Philpott, D.J. (2015). NOD-Like receptors: versatile cytosolic sentinels. *Physiological Reviews* 95, 149–178.

Murai, K., and Treisman, R. (2002). Interaction of Serum Response Factor (SRF) with the Elk-1 B box inhibits RhoA-Actin signaling to SRF and potentiates transcriptional activation by Elk-1. *Mol. Cell. Biol.* 22, 7083–7092.

Nelson, D.E., Ihekwaba, A.E.C., Elliott, M., Johnson, J.R., Gibney, C.A., Foreman, B.E., Nelson, G., See, V., Horton, C.A., Spiller, D.G., et al. (2004). Oscillations in NF- κ B signaling control the dynamics of gene expression. *Science* 306, 704–708.

Nerlov, C., and Graf, T. (1998). PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.* 12, 2403–2412.

Netea, M.G., Joosten, L.A.B., Latz, E., Mills, K.H.G., Natoli, G., Stunnenberg, H.G., O'Neill, L.A.J., and Xavier, R.J. (2016). Trained immunity: A program of innate immune memory in health and disease. *Science* 352, aaf1098.

Oishi, Y., Spann, N.J., Link, V.M., Muse, E.D., Strid, T., Edillor, C., Kolar, M.J., Matsuzaka, T., Hayakawa, S., Tao, J., et al. (2017). SREBP1 contributes to resolution of pro-inflammatory TLR4 signaling by reprogramming fatty acid metabolism. *Cell Metabolism* 25, 412–427.

Okin, D., and Medzhitov, R. (2012). Evolution of inflammatory diseases. *Current Biology* 22, R733–R740.

Olson, E.N., and Nordheim, A. (2010). Linking actin dynamics and gene transcription to drive cellular motile functions. *Nat Rev Mol Cell Biol* 11, 353–365.

Ostuni, R., and Natoli, G. (2013). Lineages, cell types and functional states: a genomic view. *Current Opinion in Cell Biology* 25, 759–764.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. *Cell* 152, 157–171.

Ourthiague, D.R., Birnbaum, H., Ortenlöf, N., Vargas, J.D., Wollman, R., and Hoffmann, A. (2015). Limited specificity of IRF3 and ISGF3 in the transcriptional innate-immune response to double-stranded RNA. *J Leukoc Biol* 98, 119–128.

Pamukcu, B., Lip, G.Y.H., and Shantsila, E. (2011). The nuclear factor – kappa B pathway in atherosclerosis: A potential therapeutic target for atherothrombotic vascular disease. *Thrombosis Research* 128, 117–123.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon- β enhanceosome. *Cell* 129, 1111–1123.

Panne, D., McWhirter, S.M., Maniatis, T., and Harrison, S.C. (2007). Interferon regulatory factor 3 is regulated by a dual phosphorylation-dependent switch. *J. Biol. Chem.* 282, 22816–22822.

Parvatiyar, K., Barber, G.N., and Harhaj, E.W. (2010). TAX1BP1 and A20 inhibit antiviral signaling by targeting TBK1-IKKi kinases. *J. Biol. Chem.* 285, 14999–15009.

Patel, S., and Santani, D. (2009). Role of NF-kappa B in the pathogenesis of diabetes and its associated complications. *Pharmacol Rep* 61, 595–603.

Pellegrini, L., Tan, S., and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. *Nature* 376, 490–498.

Perkins, N.D. (2006). Post-translational modifications regulating the activity and function of the nuclear factor kappa B pathway. *Oncogene* 25, 6717–6730.

Philibert, R.A., Sears, R.A., Powers, L.S., Nash, E., Bair, T., Gerke, A.K., Hassan, I., Thomas, C.P., Gross, T.J., and Monick, M.M. (2012). Coordinated DNA methylation and gene expression changes in smoker alveolar macrophages: specific effects on VEGF receptor 1 expression. *J Leukoc Biol* 92, 621–631.

Posern, G., and Treisman, R. (2006). Actin' together: serum response factor, its cofactors and the link to signal transduction. *Trends in Cell Biology* 16, 588–596.

Rabani, M., Raychowdhury, R., Jovanovic, M., Rooney, M., Stumpo, D.J., Pauli, A., Hacohen, N., Schier, A.F., Blackshear, P.J., Friedman, N., et al. (2014). High-resolution sequencing and modeling identifies distinct dynamic rna regulatory

strategies. *Cell* 159, 1698–1710.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by cpg islands and nucleosome remodeling. *Cell* 138, 114–128.

Ramirez-Carrozzi, V.R., Nazarian, A.A., Li, C.C., Gore, S.L., Sridharan, R., Imbalzano, A.N., and Smale, S.T. (2006). Selective and antagonistic functions of SWI/SNF and Mi-2 β nucleosome remodeling complexes during an inflammatory response. *Genes Dev.* 20, 282–296.

Ramirez, S., Ali, S.A.S., Robin, P., Trouche, D., and Harel-Bellan, A. (1997). The CREB-binding protein (CBP) cooperates with the serum response factor for transactivation of the c-fos serum response element. *J. Biol. Chem.* 272, 31016–31021.

Raza, S., Barnett, M.W., Barnett-Itzhaki, Z., Amit, I., Hume, D.A., and Freeman, T.C. (2014). Analysis of the transcriptional networks underpinning the activation of murine macrophages by inflammatory mediators. *J Leukoc Biol* 96, 167–183.

Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlén, S.-E., Greco, D., Söderhäll, C., Scheynius, A., and Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLOS ONE* 7, e41361.

Reményi, A., Schöler, H.R., and Wilmanns, M. (2004). Combinatorial control of gene expression. *Nat Struct Mol Biol* 11, 812–815.

Rönnerblad, M., Andersson, R., Olofsson, T., Douagi, I., Karimi, M., Lehmann, S., Hoof, I., Hoon, M. de, Itoh, M., Nagao-Sato, S., et al. (2014). Analysis of the DNA methylome and transcriptome in granulopoiesis reveals timed changes and dynamic enhancer methylation. *Blood* 123, e79–e89.

Rosas, M., Davies, L.C., Giles, P.J., Liao, C.-T., Kharfan, B., Stone, T.C., O'Donnell, V.B., Fraser, D.J., Jones, S.A., and Taylor, P.R. (2014). The transcription factor gata6 links tissue macrophage phenotype and proliferative renewal. *Science* 344, 645–648.

Ryo, A., Suizu, F., Yoshida, Y., Perrem, K., Liou, Y.-C., Wulf, G., Rottapel, R., Yamaoka, S., and Lu, K.P. (2003). Regulation of NF- κ B signaling by pin1-dependent prolyl isomerization and ubiquitin-mediated proteolysis of p65/RelA. *Molecular Cell* 12, 1413–1426.

Saitoh, T., Tun-Kyi, A., Ryo, A., Yamamoto, M., Finn, G., Fujita, T., Akira, S., Yamamoto, N., Lu, K.P., and Yamaoka, S. (2006). Negative regulation of interferon-

regulatory factor 3–dependent innate antiviral response by the prolyl isomerase Pin1. *Nat Immunol* 7, 598–605.

Sánchez-Martín, L., Estecha, A., Samaniego, R., Sánchez-Ramón, S., Vega, M.Á., and Sánchez-Mateos, P. (2011). The chemokine CXCL12 regulates monocyte-macrophage differentiation and RUNX3 expression. *Blood* 117, 88–97.

Sanjabi, S., Williams, K.J., Sacconi, S., Zhou, L., Hoffmann, A., Ghosh, G., Gerondakis, S., Natoli, G., and Smale, S.T. (2005). A c-Rel subdomain responsible for enhanced DNA-binding affinity and selective gene activation. *Genes Dev.* 19, 2138–2151.

Shembade, N., Ma, A., and Harhaj, E.W. (2010). Inhibition of NF- κ B Signaling by A20 Through Disruption of Ubiquitin Enzyme Complexes. *Science* 327, 1135–1139.

Shi, H.-X., Yang, K., Liu, X., Liu, X.-Y., Wei, B., Shan, Y.-F., Zhu, L.-H., and Wang, C. (2010). Positive regulation of interferon regulatory factor 3 activation by Herc5 via ISG15 modification. *Mol. Cell. Biol.* 30, 2424–2436.

Shih, V.F.-S., Tsui, R., Caldwell, A., and Hoffmann, A. (2011). A single NF κ B system for both canonical and non-canonical signaling. *Cell Res* 21, 86–102.

Siggers, T., Chang, A.B., Teixeira, A., Wong, D., Williams, K.J., Ahmed, B., Ragoussis, J., Udalova, I.A., Smale, S.T., and Bulyk, M.L. (2012). Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF- κ B family DNA binding. *Nat Immunol* 13, 95–102.

Smith, A.M., Qualls, J.E., O'Brien, K., Balouzian, L., Johnson, P.F., Schultz-Cherry, S., Smale, S.T., and Murray, P.J. (2011). A distal enhancer in *il12b* is the target of transcriptional repression by the STAT3 pathway and requires the basic leucine zipper (B-ZIP) protein NFIL3. *J. Biol. Chem.* 286, 23582–23590.

Sullivan, A.L., Benner, C., Heinz, S., Huang, W., Xie, L., Miano, J.M., and Glass, C.K. (2011). Serum response factor utilizes distinct promoter- and enhancer-based mechanisms to regulate cytoskeletal gene expression in macrophages. *Mol. Cell. Biol.* 31, 861–875.

Sun, Q., Chen, G., Streb, J.W., Long, X., Yang, Y., Stoeckert, C.J., and Miano, J.M. (2006). Defining the mammalian CARome. *Genome Res.* 16, 197–207.

Sun, S.-C. (2011). Non-canonical NF- κ B signaling pathway. *Cell Res* 21, 71–85.

Tagore, M., McAndrew, M.J., Gjidoda, A., and Floer, M. (2015). The lineage-specific transcription factor PU.1 prevents polycomb-mediated heterochromatin formation at macrophage-specific genes. *Mol. Cell. Biol.* 35, 2610–2625.

Takaoka, A., Wang, Z., Choi, M.K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi,

M., Kodama, T., Honda, K., et al. (2007). DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* 448, 501–505.

Takaoka, A., Yanai, H., Kondo, S., Duncan, G., Negishi, H., Mizutani, T., Kano, S., Honda, K., Ohba, Y., Mak, T.W., et al. (2005). Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature* 434, 243–249.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell* 60, 909–920.

Teng, M.W.L., Bowman, E.P., McElwee, J.J., Smyth, M.J., Casanova, J.-L., Cooper, A.M., and Cua, D.J. (2015). IL-12 and IL-23 cytokines: from discovery to targeted therapies for immune-mediated inflammatory diseases. *Nat Med* 21, 719–729.

Thompson, M.R., Kaminski, J.J., Kurt-Jones, E.A., and Fitzgerald, K.A. (2011). Pattern recognition receptors and the innate immune response to viral infection. *viruses* 3, 920–940.

Tong, A.-J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. *Cell* 165, 165–179.

Treisman, R., Marais, R., and Wynne, J. (1992). Spatial flexibility in ternary complexes between SRF and its accessory proteins. *EMBO J* 11, 4631–4640.

Trinchieri, G. (1995). Interleukin-12: A proinflammatory cytokine with immunoregulatory functions that bridge innate resistance and antigen-specific adaptive immunity. *Annual Review of Immunology* 13, 251–276.

Trinchieri, G. (2003). Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nat Rev Immunol* 3, 133–146.

Tullai, J.W., Schaffer, M.E., Mullenbrock, S., Sholder, G., Kasif, S., and Cooper, G.M. (2007). Immediate-early and delayed primary response genes are distinct in function and genomic architecture. *J Biol Chem* 282, 23981–23995.

Val, S.D., Chi, N.C., Meadows, S.M., Minovitsky, S., Anderson, J.P., Harris, I.S., Ehlers, M.L., Agarwal, P., Visel, A., Xu, S.-M., et al. (2008). Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. *Cell* 135, 1053–1064.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* 16, 178–189.

Vignali, D.A.A., and Kuchroo, V.K. (2012). IL-12 family cytokines: immunological playmakers. *Nat Immunol* 13, 722–728.

Vogel, C., Abreu, R. de S., Ko, D., Le, S.-Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., and Penalva, L.O. (2010). Sequence signatures and mRNA concentration can explain two - thirds of protein abundance variation in a human cell line. *Molecular Systems Biology* 6, 400.

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

Wallner, S., Schröder, C., Leitão, E., Berulava, T., Haak, C., Beißer, D., Rahmann, S., Richter, A.S., Manke, T., Bönisch, U., et al. (2016). Epigenetic dynamics of monocyte-to-macrophage differentiation. *Epigenetics & Chromatin* 9, 33.

Wang, V.Y.-F., Huang, W., Asagiri, M., Spann, N., Hoffmann, A., Glass, C., and Ghosh, G. (2012). The transcriptional specificity of NF- κ B dimers is coded within the kb dna response elements. *Cell Reports* 2, 824–839.

Wang, Z., Wang, D.-Z., Hockemeyer, D., McAnally, J., Nordheim, A., and Olson, E.N. (2004). Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression. *Nature* 428, 185–189.

Wang, Z., Wang, D.-Z., Pipes, G.C.T., and Olson, E.N. (2003). Myocardin is a master regulator of smooth muscle gene expression. *PNAS* 100, 7129–7134.

Warner, N., and Núñez, G. (2013). MyD88: A critical adaptor protein in innate immunity signal transduction. *The Journal of Immunology* 190, 3–4.

Weinmann, A.S., Mitchell, D.M., Sanjabi, S., Bradley, M.N., Hoffmann, A., Liou, H.-C., and Smale, S.T. (2001). Nucleosome remodeling at the IL-12 p40 promoter is a TLR-dependent, Rel-independent event. *Nat Immunol* 2, 51–57.

Weinmann, A.S., Plevy, S.E., and Smale, S.T. (1999). Rapid and selective remodeling of a positioned nucleosome during the induction of IL-12 p40 transcription. *Immunity* 11, 665–675.

Yamamoto, H., Kihara-Negishi, F., Yamada, T., Hashimoto, Y., and Oikawa, T. (1999). Physical and functional interactions between the transcription factor PU.1 and the coactivator CBP. *Oncogene* 18, 1495–1501.

Yamamoto, M., and Takeda, K. (2010). Current views of Toll-like receptor signaling pathways. *Gastroenterology Research and Practice* 2010, e240365.

Yang, X.-D., Huang, B., Li, M., Lamb, A., Kelleher, N.L., and Chen, L.-F. (2009). Negative regulation of NF- κ B action by Set9-mediated lysine methylation of the RelA subunit. *EMBO J* 28, 1055–1066.

Yoshida, K., Maekawa, T., Zhu, Y., Renard-Guillet, C., Chatton, B., Inoue, K., Uchiyama, T., Ishibashi, K., Yamada, T., Ohno, N., et al. (2015). The transcription factor ATF7 mediates lipopolysaccharide-induced epigenetic changes in macrophages involved in innate immunological memory. *Nat Immunol* 16, 1034–1043.

Zaromytidou, A.-I., Miralles, F., and Treisman, R. (2006). MAL and ternary complex factor use different mechanisms to contact a common surface on the serum response factor DNA-binding domain. *Mol. Cell. Biol.* 26, 4134–4148.

Zhang, Q., Lenardo, M.J., and Baltimore, D. (2017). 30 years of NF- κ B: a blossoming of relevance to human pathobiology. *Cell* 168, 37–57.

Zhong, H., May, M.J., Jimi, E., and Ghosh, S. (2002). The phosphorylation status of nuclear NF- κ B determines its association with CBP/p300 or HDAC-1. *Molecular Cell* 9, 625–636.

Zhou, L., Nazarian, A.A., Xu, J., Tantin, D., Corcoran, L.M., and Smale, S.T. (2007). An inducible enhancer required for *Il12b* promoter activity in an insulated chromatin environment. *Mol. Cell. Biol.* 27, 2698–2712.

Zhu, J., He, F., Hu, S., and Yu, J. (2008). On the nature of human housekeeping genes. *Trends in Genetics* 24, 481–484.

CHAPTER 2

A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation

A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation

Ann-Jay Tong,^{1,3} Xin Liu,^{1,3} Brandon J. Thomas,^{1,3} Michelle M. Lissner,¹ Mairead R. Baker,² Madhavi D. Senagolage,² Amanda L. Allred,² Grant D. Barish,² and Stephen T. Smale^{1,*}

¹Department of Microbiology, Immunology, and Molecular Genetics, and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

²Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

³Co-first author

*Correspondence: smale@mednet.ucla.edu

<http://dx.doi.org/10.1016/j.cell.2016.01.020>

SUMMARY

Much has been learned about transcriptional cascades and networks from large-scale systems analyses of high-throughput datasets. However, analysis methods that optimize statistical power through simultaneous evaluation of thousands of ChIP-seq peaks or differentially expressed genes possess substantial limitations in their ability to uncover mechanistic principles of transcriptional control. By examining nascent transcript RNA-seq, ChIP-seq, and binding motif datasets from lipid A-stimulated macrophages with increased attention to the quantitative distribution of signals, we identified unexpected relationships between the *in vivo* binding properties of inducible transcription factors, motif strength, and transcription. Furthermore, rather than emphasizing common features of large clusters of co-regulated genes, our results highlight the extent to which unique mechanisms regulate individual genes with key biological functions. Our findings demonstrate the mechanistic value of stringent interrogation of well-defined sets of genes as a complement to broader systems analyses of transcriptional cascades and networks.

INTRODUCTION

The molecular biology revolution of the 1970s was followed by a 20-year period during which gene regulation was studied at the level of individual model genes. Near the turn of the century, the emergence of DNA microarrays and whole-genome sequences opened avenues toward the study of gene regulation at a global scale, making it possible to identify genes and networks that characterize a cell type, environmental response, or disease state. More recently, RNA sequencing (RNA-seq) has emerged as a method that allows global transcript levels to be evaluated with greater accuracy (Marioni et al., 2008). RNA-seq also provides an opportunity to monitor nascent transcripts in addition to mRNA (e.g., Bhatt et al., 2012; Core et al., 2008;

Rabani et al., 2011). For studies of stimulus-induced transcription, nascent transcript levels provide more accurate information about the kinetics with which gene transcription is activated, and they allow transcription to be studied independently of mRNA stability.

Transcriptional cascades induced by inflammatory stimuli in cells of the mouse innate immune system have been especially well studied at a global scale, with most studies focusing on cells stimulated with lipopolysaccharide (LPS) or lipid A. LPS and lipid A engage Toll-like receptor 4 (TLR4), which then activates common signaling pathways via the MyD88 and TRIF adaptors. The TLR4-induced cascade has been monitored by DNA microarray, RNA-seq, and nascent transcript RNA-seq (e.g., Amit et al., 2009; Bhatt et al., 2012; Ramsey et al., 2008). Binding sites for several transcription factors are enriched within the promoters of defined clusters of co-regulated genes, and distinct subsets of promoters contain features of either active or inactive chromatin prior to cell stimulation (Hargreaves et al., 2009; Ramirez-Carrozzi et al., 2009). Moreover, thousands of inducible enhancers have been defined, with some enhancers poised for activation and others lacking chromatin marks prior to stimulation (Ghisletti et al., 2010; Heinz et al., 2010; Ostuni et al., 2013). Gene expression profiles have been further integrated with chromatin immunoprecipitation sequencing (ChIP-seq) datasets and small interfering RNA (siRNA) knockdown experiments for transcription factors and chromatin regulators (e.g., Amit et al., 2009; Garber et al., 2012).

Although conventional systems analyses have provided considerable insight into the logic underlying the transcriptional response to a stimulus, the results are often limited to statistical trends and lack the precision needed to fully uncover molecular mechanisms. Moreover, for most systems analyses, all genes that are induced or differentially expressed by a magnitude exceeding a low threshold—often 2-fold—are considered equally. This approach enhances statistical power and provides an opportunity to simultaneously examine an entire “system.” However, the results tend to be strongly biased toward genes that are differentially expressed by small magnitudes; these genes are far more prevalent—and may be regulated by different mechanisms—than genes differentially expressed by large magnitudes.

Here, we describe an analysis of lipid A-induced transcription using gene-centric approaches that place greater emphasis on

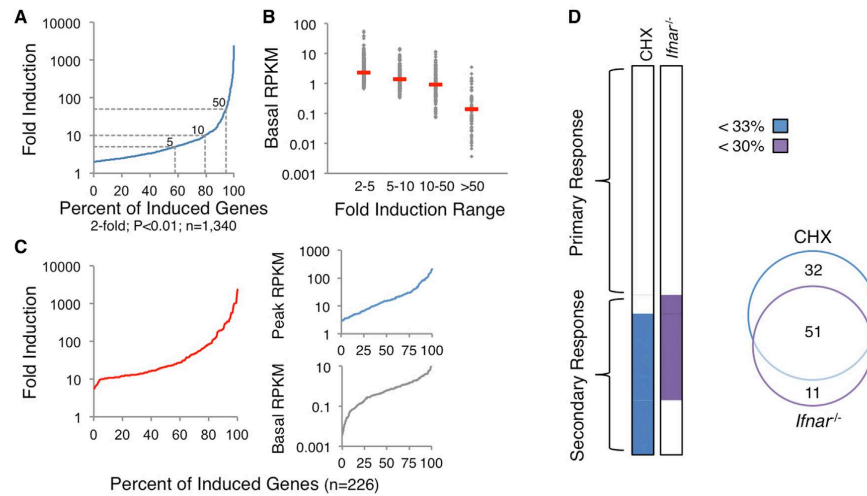


Figure 1. Properties of the Lipid A-Induced Transcriptional Cascade
 Chromatin-associated transcripts from BMDMs stimulated with lipid A were analyzed by RNA-seq.
 (A) The distribution of maximum fold induction values over the 2 hr stimulation period is shown for the 1,340 significantly induced (2-fold, $p < 0.01$) and expressed (three RPKM) genes. With multiple hypothesis testing, two weakly induced genes exhibited q values > 0.01 . The dashed gray lines represent 5-, 10-, and 50-fold induction thresholds.
 (B) The 1,340 induced genes were grouped into bins, with basal RPKMs shown for each bin and red dashes indicating median RPKMs.
 (C) The distributions of maximum fold inductions (left), peak RPKMs (top right), and basal RPKMs (bottom right) are shown for the 226 genes selected for analysis.
 (D) The 226 genes were separated into PRG and SRG groups on the basis of their expression in CHX-treated and *Ifnar*^{-/-} BMDMs. Genes were classified as SRGs if they were expressed $< 33\%$ in CHX or $< 30\%$ in *Ifnar*^{-/-} samples. The Venn diagram indicates the number of genes affected by CHX treatment, the absence of IFNAR, or both.

quantitative aspects of nascent transcript RNA-seq, ChIP-seq, and binding motif datasets. In addition to providing insight into a number of unanswered mechanistic questions, these approaches allowed us to move beyond the identification of common features of large clusters of co-regulated genes and toward an appreciation of the unique molecular mechanisms used to regulate individual genes within the inflammatory cascade.

RESULTS

Basic Properties of the Transcriptional Cascade

We first performed RNA-seq with mouse bone-marrow-derived macrophages (BMDMs) treated with lipid A for 0, 15, 30, 60, and 120 min. To separate transcription from mRNA stability, we analyzed nascent, chromatin-associated transcripts. 3,863 (14.1%) of the 27,384 annotated Refseq genes (prior to removal of duplicate isoforms) reached an expression level of at least three RPKM in at least one time point. We used a high expression threshold because our subsequent analysis emphasized induction magnitudes, which can be quantified most accurately when both basal and induced transcript levels can be measured with confidence.

Of the 3,863 expressed genes, 1,340 (34.7%) were induced by at least 2-fold ($p < 0.01$) (Figure 1A). Importantly, however, 79.5% of these genes were induced less than 10-fold (Figure 1A). If all

genes induced by 2-fold or greater were evaluated together, the analysis would be dominated by weakly induced genes. Notably, most induced genes encoding key cytokines, chemokines, and transcription factors were induced by > 10 -fold (data not shown). We therefore focused on the potentially induced genes, with the resulting insights then examined in the context of the weakly induced genes (see below). Notably, the basal transcript levels of the weakly induced genes were generally higher than those of the strongly induced genes (Figure 1B).

With the above considerations in mind, we focused on 226 genes, 215 of which were induced ($p < 0.01$) > 10 -fold during the 2-hr induction period. The remaining 11 genes were transiently induced by 5- to 10-fold at the 15-min time point; these genes were added to capture a larger number of genes that are rapidly downregulated after their early induction. Although the analysis focuses on only 226 genes, their basal and peak transcript levels were distributed over more than two orders of magnitude (Figure 1C).

Separation of Primary and Secondary Response Genes

We next separated primary response genes (PRGs) and secondary response genes (SRGs) by performing RNA-seq with nascent transcripts from BMDMs stimulated with lipid A in the presence of cycloheximide (CHX). This analysis revealed 83 genes that were expressed at a level in CHX-treated cells that was $< 33\%$

of the expression level in untreated cells (Figure 1D). These 83 genes were included in the SRG group (Figure 1D).

Interferon- β (IFN- β) expression is induced by lipid A and activates a type I IFN gene program. RNA-seq analysis of nascent transcripts from type I IFN receptor (IFNAR)-deficient (*Ifnar*^{-/-}) BMDMs stimulated with lipid A revealed 62 genes that were expressed <30% of wild-type (WT) (Figure 1D). Interestingly, 11 of these IFNAR-dependent genes were classified as PRGs in the CHX analysis because they exhibited expression levels in the presence of CHX that placed them just above the threshold used for SRG classification. Nevertheless, an analysis of their induction kinetics revealed greater similarity to the other IFNAR-dependent SRGs than to the PRGs (data not shown; see Figure S1). Because of their strong IFNAR dependence and kinetic profiles, these 11 genes were added to the SRG category (Figure 1D). Thus, 132 and 94 genes, respectively, were defined as PRGs and SRGs for the current analysis. Because some genes possess both primary and secondary response components (data not shown), the classification assignments will need to be re-evaluated as our knowledge increases.

Separation of IFNAR-Dependent and -Independent SRGs

As described above, a central feature of the response to lipid A is the activation of type I IFN. Therefore, we separated SRGs into IFNAR-dependent and -independent groups. Forty-two of the 94 SRGs were expressed <10% of WT in *Ifnar*^{-/-} BMDMs, with an additional 22 expressed between 10% and 33% (Figures 2A and 2B). Kinetic analyses revealed that 41 of the 42 genes expressed <10% of WT failed to reach an expression level in WT cells corresponding to 10% of the maximum level until the 120-min time point (Figure 2C), indicating that a robust transcriptional response to IFNAR signaling begins between 60 and 120 min post-stimulation. In contrast, 22 of the 23 SRGs that were largely unaltered in the *Ifnar*^{-/-} cells (expression level >50% of WT) reached an expression level in WT cells corresponding to 10% of their maximum within 60 min (Figure 2C). Thus, the CHX-sensitive events needed for activation of IFNAR-independent SRGs generally occur more rapidly than the autocrine/paracrine loop that activates IFNAR-dependent genes.

To separate IFNAR-dependent and -independent genes more carefully, we further examined the RNA-seq datasets from lipid A-stimulated *Ifnar*^{-/-} BMDMs, as well as RNA-seq datasets from WT BMDMs stimulated with Pam3CSK4 (PAM), a TLR2 ligand that does not induce IFNAR signaling (Toshchakov et al., 2002). Twenty-nine SRGs remained strongly induced in these datasets (Figure 2D, top).

Interestingly, although these 29 SRGs were strongly induced in the absence of IFNAR signaling, a subset, including the critical T cell polarizing cytokines *Il12b*, *Il6*, *Lif*, and *Il27* (Metcalfe, 2011; Shih et al., 2014), were induced much less potently by PAM than by lipid A (Figure 2D, bottom). In fact, *Il12b*, *Il6*, *Lif*, and *Il27* exhibited greater differential induction by TLR4 versus TLR2 ligands than any other PRG or SRG (Figure 2E). This finding suggests that the TRIF pathway activated by lipid A but not by PAM may be important for the activation of these genes, but not due to its role in activating IFNAR signaling. Consistent with this possi-

bility, a direct comparison of WT to *Trif*^{-/-} BMDMs revealed strong TRIF dependence of these genes (Figure 2D, bottom). Together, the data suggest that lipid A induces the expression of key T cell polarizing cytokines (*Il12b*, *Il6*, *Lif*, and *Il27*) much more potently than does PAM because the TRIF pathway strongly promotes the expression of these genes in an IFNAR-independent manner.

To better understand the significance of the above regulatory strategies, we performed gene ontology analysis with the 132 PRGs, 65 IFNAR-dependent SRGs, and 29 IFNAR-independent SRGs (Figure 2F). The PRG analysis suggested broad roles in regulating inflammation and the functions of blood cells. As expected, the IFNAR-dependent SRGs were implicated in anti-viral responses. Most interestingly, the small group of IFNAR-independent SRGs exhibited highly significant enrichment for genes that regulate T cell proliferation, differentiation, and activation. Specifically, 14 of the 29 IFNAR-independent SRGs are involved in the regulation of T cell responses (Figure 2G). Eleven of these 14 genes are among the 13 IFNAR-independent SRGs that are most potently induced by lipid A. Thus, these results reveal common regulatory features of a prominent group of genes that helps bridge the innate and adaptive immune systems. Nevertheless, a careful examination reveals that the induction kinetics for each of these genes is unique (Figure S1), suggesting that gene-specific regulatory events are superimposed on top of their common characteristics of potent and rapid CHX-sensitive yet IFNAR-independent induction.

Initial Analysis of PRGs

Shifting our attention to the 132 PRGs, we first examined their expression kinetics in greater detail by nascent transcript RNA-seq from lipid A-stimulated BMDMs collected every 5 min during the first hour of activation, with an additional 120-min time point. We also performed nascent transcript RNA-seq with BMDMs from *Myd88*^{-/-}, *Trif*^{-/-}, *Myd88*^{-/-}*Trif*^{-/-}, and *Irf3*^{-/-} mice, and with WT BMDMs stimulated with lipid A in the presence of ERK and p38 MAPK inhibitors; the two inhibitors were analyzed together because little effect was observed in pilot experiments with each inhibitor alone (data not shown). The results consider the maximum induced RPKM in WT cells for each gene to be 100% and the basal RPKM in unstimulated WT cells to be 0%; the maximum induced RPKM observed in each mutant strain for each gene is then displayed as a percentage of the maximum WT RPKM.

Figure 3A (see also Figure S2) shows that each perturbation resulted in a continuum of effects. For this study, genes expressed <33% of WT were considered to be dependent on the factor that was absent. By combining these datasets with k-means cluster analysis of expression kinetics, an initial classification of the 132 PRGs emerged (Figure 3D; see Figure S2 for gene names). Cluster 1 includes nine genes that exhibited reduced expression (<33% of WT) in both *Trif*^{-/-} and *Irf3*^{-/-} BMDMs. Clusters 2–5 include 28 genes that exhibited reduced expression in *Trif*^{-/-} but not in *Irf3*^{-/-} BMDMs (Figures 3B and 3D); these genes were then subdivided by k-means clustering on the basis of their expression kinetics. Clusters 6–9 include 38 genes that exhibited reduced expression in WT BMDMs treated with MAPK inhibitors, but without strongly reduced expression in *Trif*^{-/-} BMDMs; as

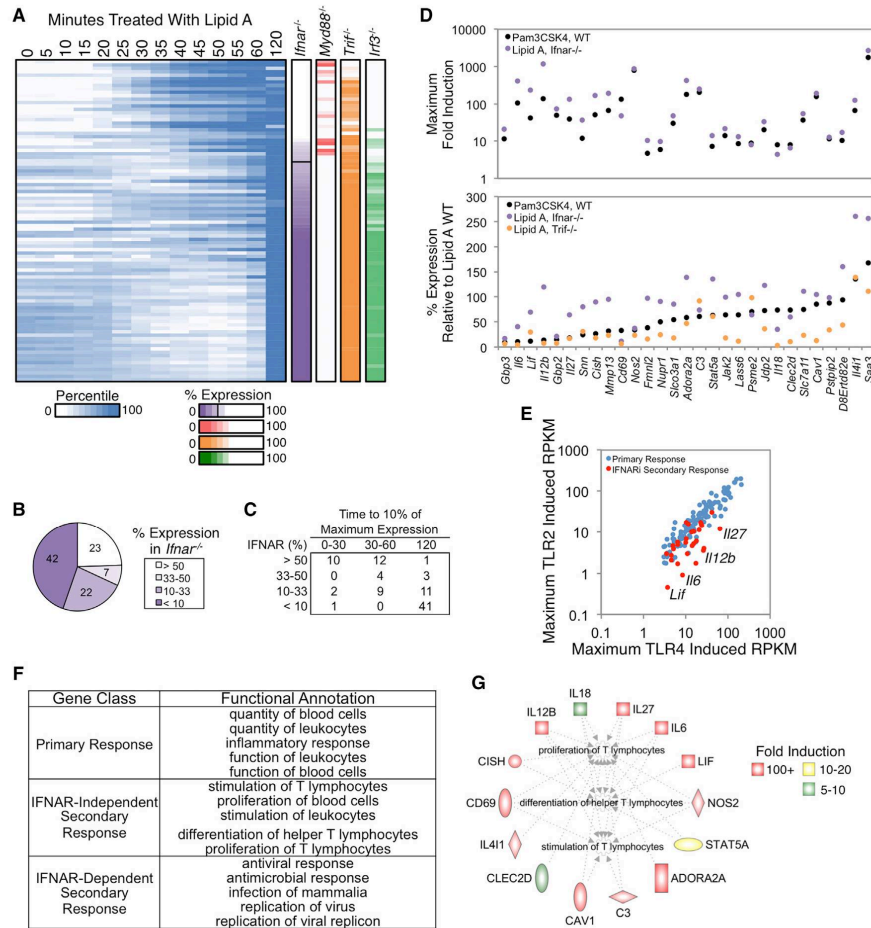


Figure 2. Analysis of IFNAR-Independent and -Dependent SRGs

(A) Activation kinetics are shown for SRGs from BMDMs stimulated at 5-min intervals from 0–60 min, with an additional 120-min time point. Shades of blue indicate percentile values. Genes were sorted on their maximum percent expression in *Ifnar*^{-/-} BMDMs relative to WT BMDMs (purple column). The maximum percent expressions in *Myd88*^{-/-}, *Trif*^{-/-}, and *Ifi3*^{-/-} BMDMs are shown to the right. See also Figure S1.

(B) The distribution of genes in IFNAR-dependence bins based on their expression in *Ifnar*^{-/-} BMDMs is shown.

(C) The time point at which each SRG in the IFN-dependence bins reached 10% of its maximum expression is indicated.

(D) The maximum fold induction of the 29 IFNAR-independent genes in PAM-stimulated (black) and lipid A-stimulated *Ifnar*^{-/-} (purple) BMDMs is shown (top), along with the percent expression of these genes in PAM-stimulated (black), lipid A-stimulated *Ifnar*^{-/-} (purple), and lipid A-stimulated *Trif*^{-/-} (orange) BMDMs relative to WT BMDMs stimulated with lipid A (bottom). IFNAR-independent genes were defined as those induced >10-fold and expressed >3 RPKM in the absence of IFNAR signaling, or expressed at greater than 50% of WT in *Ifnar*^{-/-} BMDMs stimulated with lipid A or WT BMDMs stimulated with PAM.

(E) A scatterplot comparing the maximum RPKMs in PAM-stimulated BMDMs (y axis) and lipid A-stimulated BMDMs (x axis) for PRGs (blue) and the IFNAR-independent SRGs (red) is shown.

(F) Ingenuity Pathway Analysis was used to identify the top functional annotations for PRGs and the IFNAR-dependent and -independent SRGs.

(G) The IFNAR-independent genes involved in the proliferation, differentiation, and activation of T lymphocytes (Ingenuity Pathway Analysis) are colored based on their fold induction in *Ifnar*^{-/-} BMDMs.

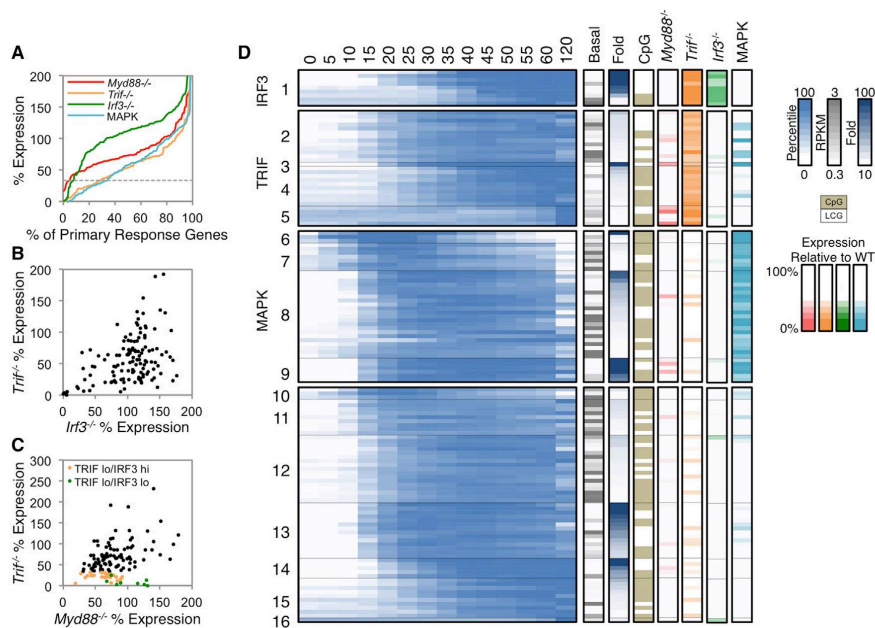


Figure 3. Properties of PRGs

(A) The distribution of the maximum percent expressions in *Myd88*^{-/-} (red), *Trif*^{-/-} (orange), *Irf3*^{-/-} (green), and MAP kinase inhibitor-treated (light blue) BMDMs stimulated with lipid A are shown for the 132 PRGs. The horizontal dashed gray line indicates the 33% expression threshold. (B and C) The percent expression of each PRG is shown in *Trif*^{-/-} versus *Irf3*^{-/-} cells (B) or in *Trif*^{-/-} versus *Myd88*^{-/-} cells (C). TRIF lo (<33% relative to WT) IRF3 hi (>33% relative to WT) genes are in orange, and TRIF lo (<33% relative to WT) IRF3 lo (<33% relative to WT) genes are in green. (D) Activation kinetics are shown (log₂-normalized and mean-centered RPKMs) for the PRGs in BMDMs stimulated for 5-min intervals between 0 and 60 min, and for 120 min. The PRGs were broadly classified based on their expression in *Myd88*^{-/-} (red), *Trif*^{-/-} (orange), *Irf3*^{-/-} (green), and MAP kinase inhibitor-treated (light blue) BMDMs with the following order: IRF3-dependent (cluster 1; <33% in both *Trif*^{-/-} and *Irf3*^{-/-}), TRIF-dependent (clusters 2–5; <33% in *Trif*^{-/-} only), and MAPK-dependent (clusters 6–9; <33% in MAPK inhibitor-treated samples). The remaining PRGs were not dependent on any perturbation examined (clusters 10–16; >33% in all perturbed datasets). The genes in each class were subclustered (k-means) on their expression kinetics. The properties of each gene are shown to the right of the heatmap: basal expression value (gray), fold induction magnitude (blue), promoter CpG-island (beige), and the maximum percent expression in *Myd88*^{-/-} (red), *Trif*^{-/-} (orange), *Irf3*^{-/-} (green), and MAPK inhibitor-treated (light blue) BMDMs. See also Figures S2 and S3.

above, the genes were subdivided by k-means clustering (Figure 3D). Finally, clusters 10–16 include the remaining 57 genes that did not exhibit reduced expression in the presence of MAPK inhibitors or in *Trif*^{-/-} or *Irf3*^{-/-} cells; these genes were divided into seven kinetic clusters. It is noteworthy that only five of the 132 PRGs exhibited reduced expression in *Myd88*^{-/-} cells (Figures 3C and 3D). No genes were induced in *Myd88*^{-/-} *Trif*^{-/-} mutant cells (data not shown).

In addition to the degree of dependence of each PRG on MyD88, TRIF, IRF3, and MAPKs, Figure 3D indicates basal transcript and fold-induction values. Furthermore, Figure 3D indicates which genes contain CpG-island or low CpG (LCG) promoters. As shown previously (Bhatt et al., 2012), all early transiently induced genes (e.g., clusters 6 and 10) contain CpG-island promoters and a high percentage of the most potently induced genes contain LCG promoters (e.g., clusters 1 and

14), whereas the two promoter types are distributed fairly randomly among the other clusters.

Initial Transcription Factor Binding Motif and ChIP-Seq Analyses

To extend the above foundation, we evaluated the over-representation of transcription factor binding motifs within the promoters of the PRGs within each of the 16 clusters in Figure 3D. This analysis (Figure S3) provided insight into transcription factors that may regulate each cluster. However, toward the goal of elucidating molecular mechanisms, these statistical enrichments were unsatisfying. For example, although nuclear factor κ B (NF- κ B) motifs are enriched in the promoters of genes in several clusters, a closer analysis revealed considerable heterogeneity within each cluster, with only a subset of promoters in a cluster generally containing a strong NF- κ B motif (data not

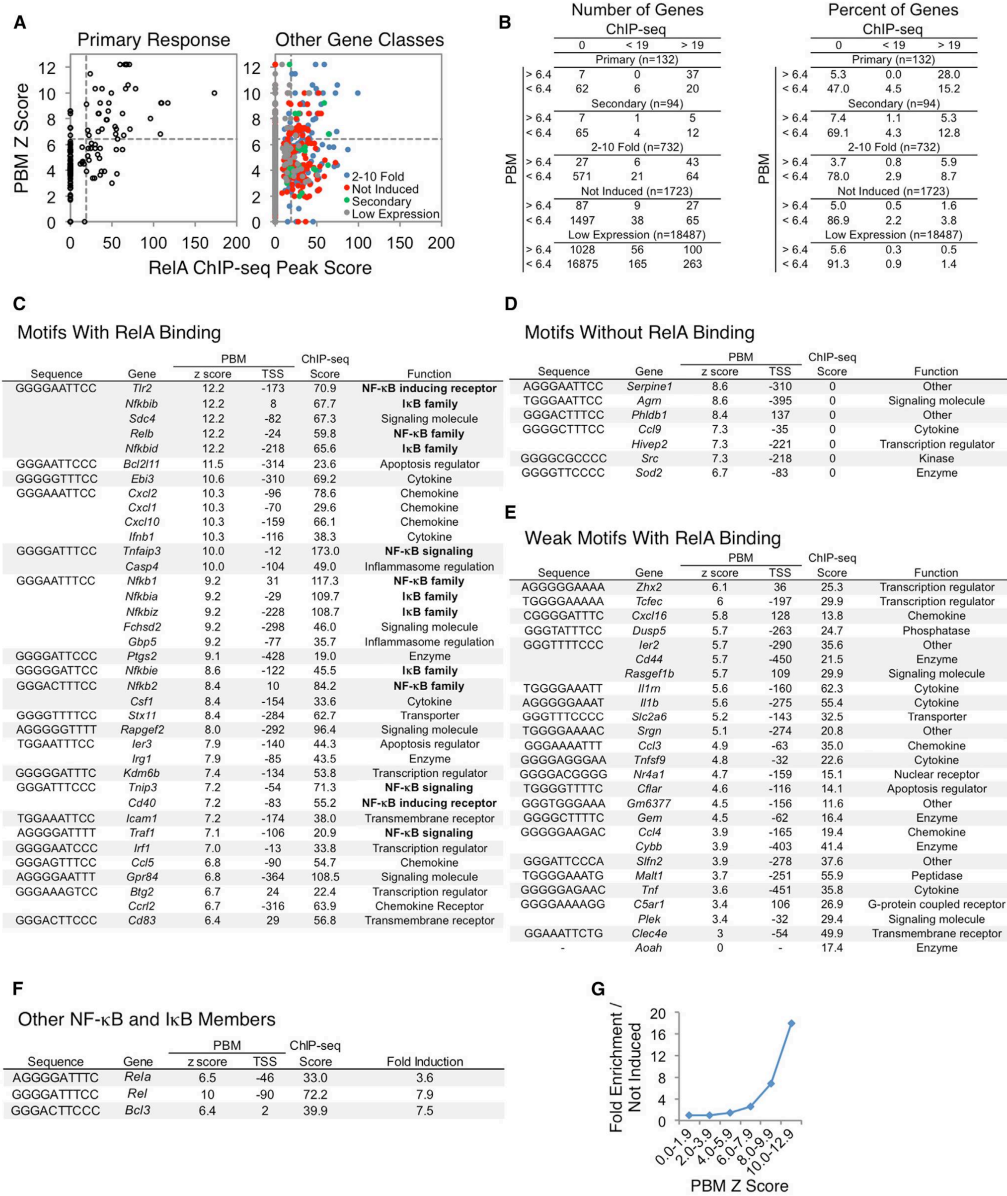


Figure 4. NF-κB Interactions at the Promoters of Defined Gene Classes

(A) PBM Z scores of p50:RelA (y axis) and RelA ChIP-seq peak scores (x axis) in the promoters of the PRGs (left) and all remaining genes in the genome (right) were plotted. The remaining genes were assigned to 2- to 10-fold induced (blue), not induced (red), SRC (green), or low expression (gray) categories. The horizontal dashed line indicates the PBM Z score threshold (6.4), and the vertical dashed line indicates the ChIP-seq peak score threshold (19).

(legend continued on next page)

shown). Imprecise correlations were also apparent when examining ChIP-seq datasets for NF- κ B and other transcription factors (data not shown). Therefore, additional strategies are needed to move beyond statistical enrichments toward more meaningful mechanistic insights.

Quantitative Analysis of NF- κ B's Contribution to the Transcriptional Cascade

We next focused on NF- κ B. Prior studies showed that a large percentage of ChIP-chip and ChIP-seq peaks for NF- κ B family members do not coincide with strong binding motifs (Lim et al., 2007; Martone et al., 2003; Zhao et al., 2014), raising questions about NF- κ B's DNA recruitment and transcriptional activation mechanisms. However, when focusing attention on the promoters of our well-defined set of strongly induced PRGs, a different relationship between NF- κ B binding and motifs emerged.

Specifically, Figure 4A examines NF- κ B ChIP-seq peak scores versus motif scores for the promoters (−500 to +150 relative to the transcription start site [TSS]) of each of the 132 PRGs. The NF- κ B motif scores were derived from protein binding microarray (PBM) results obtained with a recombinant RelA:p50 heterodimer, the most abundant NF- κ B dimer involved in TLR4-induced transcription (Siggers et al., 2012). RelA ChIP-seq experiments were performed with BMDMs stimulated with lipid A for 0, 15, 30, 60, and 120 min (followed by stringent peak-calling and a focus on peaks observed in multiple biological replicates). This analysis revealed 8,458 total peaks, with 942 promoter peaks.

When focusing on the promoters of the 132 strongly induced PRGs, a motif Z score threshold readily emerged that resulted in a high probability of a strong ChIP-seq peak; 37 of 44 promoters (84%) containing an NF- κ B motif exceeding a Z score of 6.4 supported strong RelA binding (ChIP-seq peak >19), whereas only 20 of 88 promoters (23%) whose strongest NF- κ B motif was below this motif threshold supported strong binding (Figures 4A, left, and 4B, left). These results suggest that, although a high percentage of NF- κ B genomic interactions do not coincide with strong binding motifs (see Figure 4A, right), most interactions observed at the promoters of a well-defined set of PRGs are associated with strong motifs. Thus, NF- κ B function may often require binding to a near-consensus motif. The results also suggest that, at the promoters of this well-defined set of genes, a surprisingly strict motif strength threshold exists, in which promoter motifs exceeding this threshold almost always support strong *in vivo* binding (see below). This *in vivo* threshold contrasts with the continuum of binding affinities observed *in vitro* (Siggers et al., 2012).

To evaluate the significance of these findings, we examined promoters for all other annotated genes separated into five

groups: the 132 strongly induced PRGs, the 94 strongly induced SRGs, 732 genes induced between 2- and 10-fold, 1,732 genes that were expressed at a nascent transcript level more than three RPKMs but without induction, and the remaining 18,487 annotated genes. Promoters within each group were separated into six classes on the basis of their ChIP-seq peak scores and motif scores, including three ChIP-seq categories (no binding, peak strength <19, and peak strength >19) combined with two motif categories (Z score <6.4 and >6.4) (Figure 4B).

An examination of the ChIP-seq/motif categories for the five groups of annotated genes revealed extensive enrichment of genes whose promoters combined strong ChIP-seq peaks and NF- κ B motifs among the strongly induced PRG class. Specifically, whereas 28% (37/132) of the strongly induced PRGs combined strong ChIP-seq peaks and motifs, only 1.6% (27/1,723) of expressed but uninduced genes combined strong peaks and motifs. Importantly, little or no enrichment of strongly induced PRGs was observed in four of the other ChIP-seq/motif categories (weak peak/strong motif, weak peak/weak motif, no peak/strong motif, no peak/weak motif). Substantial but lesser enrichment in the PRG class was observed for only one other ChIP-seq/motif category: those that combined a strong ChIP-seq peak with a weak motif (15.2% of strongly induced PRGs versus 3.8% of expressed uninduced genes).

The strong enrichment of promoters that combine strong ChIP-seq peaks and motifs in the group of 132 PRGs suggests that most or all of the 37 PRGs possessing these properties may be directly activated by RelA-containing dimers via direct promoter binding. Furthermore, the ability to define a motif Z-score threshold above which 84% of promoters supported strong NF- κ B binding suggests that a single strong NF- κ B motif is usually sufficient to support strong binding. Notably, although several of the 37 promoters contain two or more near-consensus NF- κ B binding motifs, a strong correlation was not found between the number of NF- κ B motifs and either the strength of the RelA ChIP-seq peak or the magnitude of transcriptional induction (data not shown). It is also noteworthy that ChIP-seq experiments examining the NF- κ B p50 subunit revealed strong peaks at all 37 promoters that contain RelA peaks and motifs (data not shown), suggesting that the promoters are typically bound by RelA:p50 heterodimers.

The substantial but lesser enrichment of promoters with strong NF- κ B peaks (score >19) and weak binding motifs (Z score <6.4) among the strongly induced PRGs is also of interest. In these promoters, NF- κ B may bind directly to weak motifs. Alternatively, NF- κ B may be recruited by other transcription factors, or the NF- κ B ChIP-seq signal could be due to looping of an NF- κ B-bound enhancer to the promoter. Although the significance of these interactions remains unknown, our ability to classify these promoters and distinguish them from the more

(B) Tables are shown indicating the distribution of genes from (A) for both numbers (left) and percentages (right) of genes.

(C–F) Tables are shown indicating the best matching κ B motif in each promoter (column 1), the gene name (column 2), the PBM p50:RelA Z score (column 3), the position of the motif relative to the TSS (column 4), the RelA ChIP-seq peak score (column 5), and either the function or fold induction (column 6). This information is included for the PRGs with (C) strong κ B motifs and strong RelA binding, (D) strong κ B motifs that do not support RelA binding, (E) weak κ B motifs and strong RelA binding, and (F) other NF- κ B and I κ B family members.

(G) A line graph is shown indicating the p50:RelA motif Z score enrichment in the promoters of the PRGs relative to the promoters of uninduced genes. See also Figure S5.

prevalent promoters that combined strong ChIP-seq peaks and motifs will facilitate future studies of their regulation.

An examination of the 732 genes induced by 2- to 10-fold provides additional insights. A higher percentage of genes in this weakly induced class (5.9%) contain strong NF- κ B peaks and motifs than in the class of genes that is expressed but not induced (1.6%). This enrichment suggests that a subset of weakly induced genes is regulated by NF- κ B binding to strong motifs. However, a much smaller percentage of genes in this 2- to 10-fold induced class (5.9%) combine strong NF- κ B peaks and motifs than in the strongly induced PRG class (28%), suggesting that a much smaller fraction of the weakly induced genes is regulated by NF- κ B promoter binding.

Examination of NF- κ B-Regulated Genes

A major goal of this study was to elucidate the logic through which the lipid A-induced transcriptional cascade is regulated. The identities of the 37 strongly induced PRGs that combine strong ChIP-seq peaks and strong motifs provide compelling evidence of an underlying logic; specifically, more than a third (13 of 37; Figure 4C) encode NF- κ B or I κ B family members or key regulators of NF- κ B activation, including three NF- κ B family members (*Nfkb1*, *Nfkb2*, and *RelB*), five I κ B family members (*Nfkbia*, *Nfkbib*, *Nfkbid*, *Nfkbie*, and *Nfkbiz*), two NF- κ B-inducing receptors (*Tlr2* and *Cd40*), and three regulators of NF- κ B signaling (*Tnfrsf3*, *Tnfrsf3*, and *Traf1*). Strikingly, these 13 genes include all of the NF- κ B/I κ B family members and direct regulators of NF- κ B signaling found among the 132 PRGs. Notably, the promoters of genes encoding the two NF- κ B family members and one I κ B family member missing from this list also combine a strong RelA ChIP-seq peak with a strong NF- κ B motif (Figure 4F); these genes were not among the 132 PRGs because they were only weakly induced.

The 37 PRGs in Figure 4C contain only 21 distinct motifs, which adhere to one of two motif definitions: (G/T)GG(G/A)(N)(A/T)(T/G)(T/C)CC (17 motifs) or (G/A)GGGG(G/A)(T/A)TT(T/C) (four motifs). The finding that a high level of similarity to the optimal NF- κ B consensus is usually associated with NF- κ B binding in the RelA ChIP-seq experiments was initially surprising. However, support for the significance of this finding emerged from an examination of binding motif enrichment at the 132 PRGs in comparison to the 1,723 expressed but uninduced genes, without any consideration of ChIP-seq data. Specifically, motifs with Z scores above 8.0 were strongly enriched among the promoters of the 132 PRGs. Motifs with Z scores between 6.0 and 7.9 were weakly enriched, but no enrichment was observed with motifs with Z scores below 6.0 (Figure 4G).

One remaining question is the reason seven promoters with motifs exceeding the threshold of 6.4 did not support RelA binding (Figures 4A and 4B). The motifs in three of these promoters possess very high Z scores (8.4–8.6, Figure 4D). However, two of these motifs are at a distance upstream of their TSS (–310 and –395) that exceeds the distance observed in all but five of the 37 promoters that support NF- κ B binding (Figure 4D). We hypothesize that these two motifs do not support binding in vivo because they are occluded by nucleosomes. The third strong motif is located farther downstream of the TSS (+137) than the motifs found in any of the promoters that support strong

NF- κ B binding, suggesting that this motif may also be masked by a nucleosome.

The three remaining motifs possess Z scores between 6.7 and 7.4 (Figure 4D). We speculate that their Z scores may be defined imperfectly due to limitations of the PBM method. One of these motifs is found in two different promoters, neither of which supports binding, and the other two do not conform to the motif definitions derived from the 21 motifs that support binding (see above). Detailed affinity measurements will be needed to better understand why a few motifs fail to support NF- κ B binding, but this quantitative analysis reveals a remarkably strong ability to predict NF- κ B promoter binding in vivo on the basis of in vitro motif strength, as well as a motif strength threshold below which the probability of in vivo binding is greatly diminished.

Kinetic and Functional Analysis of Putative NF- κ B Targets

To test the prediction that the 37 PRGs described above are regulated by NF- κ B, we examined their activation kinetics and RelA dependence. The initial upregulation of most of the genes occurred 10–20 min post-stimulation, as is evident from the third panel in Figure 5A, in which the fold increase in RPKM relative to the preceding time point is highlighted. Although most of these genes are initially upregulated at the same time, their overall expression kinetics are diverse (Figure 5A, second panel; see also Figure S4), implicating other factors in their regulation. Consistent with this suggestion, the NF- κ B target genes that depend on MAPK signaling were, on average, induced slightly earlier than the other putative target genes (Figure 5A, cluster 2; Figure 5C).

To examine RelA dependence, we compared WT and *Rela*^{–/–} fetal liver-derived macrophages by RNA-seq. Most of the putative NF- κ B targets exhibited RelA dependence (Figure 5A, *Rela*^{–/–} column), although the degree of dependence varied considerably, possibly due to redundancy between RelA and other NF- κ B family members.

We next asked whether the activation kinetics and RelA dependence observed in Figure 5A are unique to genes whose promoters contain strong ChIP-seq peaks and motifs. Interestingly, several other PRGs exhibited similar activation kinetics and/or degrees of RelA dependence (Figures 5B and 5D). A subset of these genes contains RelA ChIP-seq peaks in their promoters without strong motifs, but most lack RelA peaks (Figure 5B, right). We speculate that NF- κ B directly regulates these genes by binding to distant enhancers. Consistent with this possibility, strong RelA ChIP-seq non-promoter peaks (peak score >19) were found in the vicinity of many of the PRGs (Figure S5). Thus, NF- κ B may regulate strongly induced PRGs through either promoter or enhancer binding, with an underlying logic suggested by the fact that promoter binding characterizes genes encoding NF- κ B/I κ B family members and other NF- κ B regulators.

Gene-Specific Regulation of IRF3-Dependent Genes

Although most studies emphasize large clusters of co-regulated genes, the above data suggest that, when induction magnitudes are considered, the unique features of individual genes and small clusters of genes begin to emerge. This concept is further exemplified by an examination of PRGs dependent on the

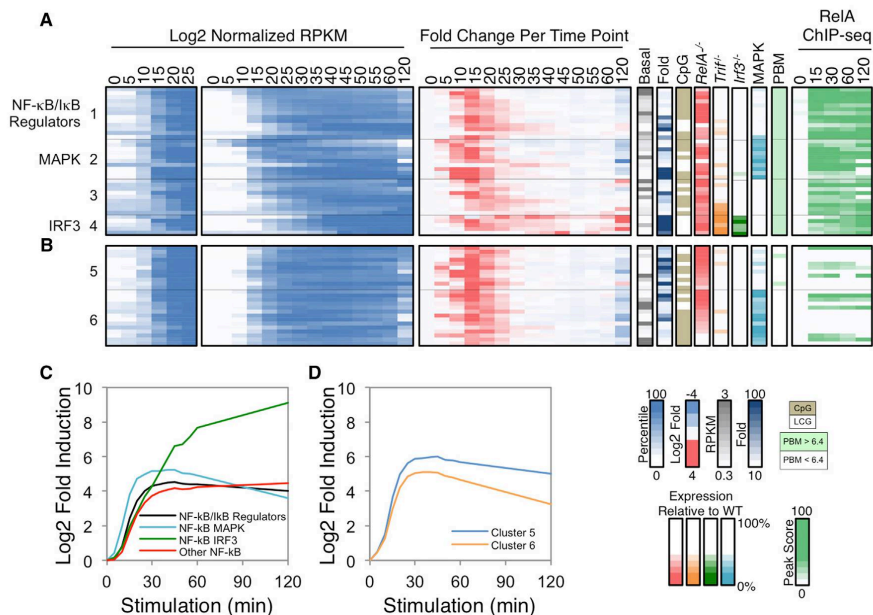


Figure 5. Kinetic and Functional Analysis of Putative NF- κ B Target Genes

(A) The 37 PRGs containing strong NF- κ B promoter motifs and RelA ChIP-seq promoter peaks were grouped into four categories; those that encode NF- κ B family members and regulators (group 1), those that exhibit either MAPK or IRF3 dependence (groups 2 and 4), and the remaining genes (group 3). Normalized expression values from 0 to 25 min (left panel) and 0 to 120 min (second panel) and the fold change relative to the previous time point (third panel) are shown. To the right of the heatmaps, the basal expression values, fold induction magnitudes, promoter CpG contents, and expression values in *RelA*^{-/-}, *Trif*^{-/-}, *Irf3*^{-/-}, and MAPK-inhibited BMDMs are shown. The presence of a p50:RelA motif based on PBM datasets and the RelA ChIP-seq binding peak scores are indicated in the far-right panels. See also Figure S4.

(B) Examples of PRGs that exhibited similar activation kinetics and/or RelA dependence to the 37 genes with strong NF- κ B motifs and ChIP-seq peaks are shown. See also Figure S4.

(C) The average activation kinetics of the NF- κ B subgroups is shown as log₂ fold inductions relative to basal during the 120 min lipid A treatment period.

(D) The average activation kinetics of the two additional clusters from Figure 5B (clusters 5 and 6) are shown.

transcription factor, IRF3. As shown in Figure 3, only nine strongly induced PRGs exhibited expression levels in both *Irf3*^{-/-} and *Trif*^{-/-} BMDMs that fail to reach 33% of the level observed in WT. Five of these genes are within the group of 37 PRGs containing strong NF- κ B ChIP-seq peaks and motifs in their promoters (Figure 5A). One notable difference between the five genes containing NF- κ B motifs and the four lacking NF- κ B motifs is that the induction magnitude of the former group is much higher than that of the latter, with average induction magnitudes of 643- and 40-fold, respectively (Figures 6A and 6B).

An examination of the five genes exhibiting both NF- κ B promoter binding and IRF3 dependence reveals the extent to which genes have evolved unique regulatory strategies. Within this group, the expression kinetics of *Ccl5* and *Ifnb1* are each unique, whereas *Cxcl10*, *Gbp5*, and *Irg1* are similar (Figure 6A). These latter three genes were initially induced 10–15 min post-stimulation along with most NF- κ B-dependent genes. Consistent with

the hypothesis that NF- κ B contributes to this early induction, RelA ChIP-seq peaks were observed at these genes by 15 min post-stimulation (Figure 6A, right), and their induction at early time points was unaltered in *Irf3*^{-/-} macrophages (data not shown). IRF3 dependence was observed only at later times, consistent with prior knowledge that IRF3 activation is relatively slow (Kagan et al., 2008).

Interestingly, *Ccl5* is unique in that RelA binding was not observed until the 30-min time point; at all other PRGs bound by RelA, RelA binding was readily detected at the 15-min time point (Figures 5A and 5B, right; Figure 6A, right). The delay in RelA binding correlates with the delayed *Ccl5* activation. Thus, RelA binding to this promoter requires an additional event that is unique among PRGs.

Ifnb1 regulation also appears unique. *Ifnb1* induction was not observed until the 35-min time point, but RelA binding was observed by 15 min (Figure 6A, right). This early binding is consistent with evidence that the promoter lacks a nucleosome

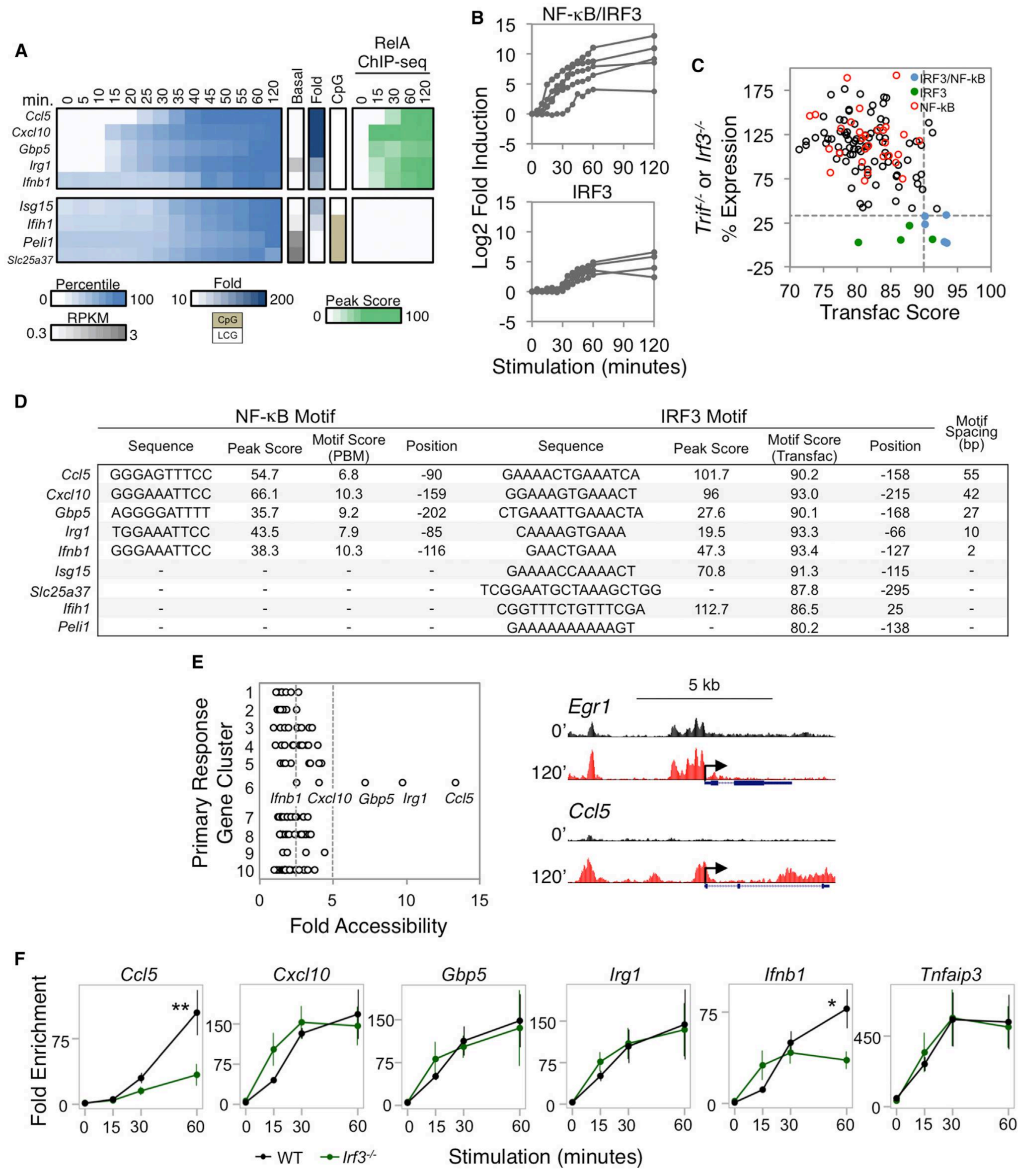


Figure 6. Analysis of IRF3 Target Genes

(A) PRGs exhibiting IRF3 dependence (<33% expression in both *Irf3*^{-/-} and *Trif*^{-/-} macrophages) were separated based on the presence or absence of strong NF-κB promoter motifs and RelA ChIP-seq peaks. Colors indicate the percentile of the relative expression. Also shown are the basal RPKM, fold induction magnitude, and promoter CpG content. The rightmost heatmap indicates the RelA ChIP-seq binding peak scores.

(B) The fold induction for each IRF3-dependent gene is shown over the 2-hr time period, grouped based on their additional requirement for NF-κB.

(legend continued on next page)

in unstimulated cells (Agalioti et al., 2000). Nevertheless, the delayed induction is consistent with evidence that activation is strongly dependent on IRF3 (Panne et al., 2007).

Figure 6C shows the distribution of promoter IRF motif scores relative to functional dependence on IRF3. IRF motifs with Transfac Position Weight Matrix (PWM) scores of 90 or greater accompany the strong NF- κ B motifs in all five promoters (Figures 6C and 6D). The distances between the IRF3 and NF- κ B motifs range from 2 to 55 bp (Figure 6D). Notably, of the four IRF3-dependent genes that do not contain NF- κ B promoter motifs, only one (*Isg15*) contains an IRF3 motif of similar strength to those found in the genes with strong NF- κ B motifs (Figures 6C and 6D).

The above results support a hypothesis in which multiple distinct mechanisms regulate the nine IRF3-dependent genes. To examine this hypothesis further, three additional experiments were performed. First, IRF3 ChIP-seq experiments revealed that strong IRF3 peaks (>19) coincide with strong IRF motifs (>90) at the promoters of only six of the 132 primary response genes, including the five NF- κ B/IRF3 genes described above and the IRF3-dependent *Isg15* gene that lacks NF- κ B binding (see peak scores in Figure 6D; a detailed analysis of the IRF3 ChIP-seq data will be presented elsewhere). An IRF3 ChIP-seq peak was also observed in the promoter of one of the IRF3-dependent genes that lacks a strong IRF motif (*Irh1*; Figure 6D).

Second, ATAC-seq experiments revealed weak increases in chromatin accessibility upon lipid A stimulation at PRGs in many different classes (Figure 6E). However, the largest increase was observed at the *Ccl5* promoter, with large increases also observed at the *Irg1* and *Gbp5* promoters (Figure 6E). The large increase at the *Ccl5* promoter is consistent with the hypothesis that a nucleosome remodeling requirement is responsible for the delayed binding of RelA. Furthermore, the absence of an inducible ATAC-seq signal at the *Irh1* promoter is consistent with prior evidence that the promoter is nucleosome-free prior to stimulation. However, the strong increases in ATAC-seq signal at the *Irg1* and *Gbp5* promoters were surprising, given the rapid RelA binding and induction of these genes.

The third experiment performed was ChIP-qPCR examining RelA binding in *Irh3*^{-/-} macrophages. This experiment revealed strong IRF3 dependence of RelA binding to the *Ccl5* promoter (Figure 6F), consistent with our evidence from nuclease accessibility experiments that IRF3 is important for nucleosome remodeling

at this promoter (Ramirez-Carrozzi et al., 2009). At the *Irh1* promoter, the initial binding of RelA was not dependent on IRF3 (Figure 6F); however, the increase in RelA binding at later time points exhibited IRF3 dependence, consistent with the notion that IRF3 stabilizes RelA binding while promoting synergistic transcriptional activation (Agalioti et al., 2000; Panne et al., 2007). Finally, although potently induced ATAC-seq signals were observed at the *Irg1* and *Gbp5* promoters, RelA binding to these promoters was not IRF3 dependent (Figure 6F). Thus, the nucleosome remodeling observed at these promoters by ATAC-seq is likely to be dictated by NF- κ B itself or by other rapidly induced factors.

Together, these results support a model in which the mechanisms by which NF- κ B and IRF3 regulate the *Ccl5* and *Irh1* genes are unique, with these two transcription factors contributing to *Irg1* and *Gbp5* activation (and possibly *Cxcl10* activation) by a third distinct mechanism. To determine whether these mechanisms appear to be unique only because we focused on a stringently defined group of PRGs, we asked whether any additional annotated promoters throughout the genome could be identified that possess the basic DNA properties of the five NF- κ B/IRF3 genes (i.e., a strong RelA ChIP-seq peak [>19], a strong NF- κ B motif [Z score >6.4], a strong IRF3 motif [Transfac score ≥ 90], and a distance between the NF- κ B and IRF3 motifs of less than 100 bp [see Figure 6D]). Strikingly, only six additional promoters from among the 21,168 annotated promoters share these properties (data not shown).

Together, these results reveal the extent to which a quantitative, gene-centric analysis can begin to move toward an understanding of the unique molecular mechanisms used to regulate key genes in the transcriptional cascade. Although previous ChIP-seq studies led to the hypothesis that IRF3 and NF- κ B cooperatively activate hundreds of genes (Freaney et al., 2013), the results presented here demonstrate that only five PRGs induced greater than 10-fold by lipid A combine strong NF- κ B promoter binding, strong IRF3 dependence, a strong IRF3 promoter motif, and strong IRF3 binding, yet with at least three distinct modes of collaboration between NF- κ B and IRF3 among these five genes. Although IRF3 can also bind many enhancers (Freaney et al., 2013), these interactions may have more subtle modulatory functions in lipid A-stimulated macrophages or may represent opportunistic binding events that lack functional consequences.

(C) For each PRG, the higher maximum percent expression from either *Trif*^{-/-} or *Irh3*^{-/-} BMDMs (y axis) was assessed against the best scoring IRF3 motif (x axis) within the promoter based on the IRF Transfac PWM. The five IRF3/NF- κ B genes are shown in blue, and the four IRF3 genes are shown in green. The PRGs containing strong NF- κ B promoter motifs and RelA ChIP-seq peaks are shown in red. The horizontal dashed line indicates the expression threshold (33%), and the vertical dashed line indicates the Transfac threshold (90).

(D) For each IRF3-dependent gene, the IRF3 and RelA:p50 binding sites (for the IRF3/NF- κ B groups of genes) were identified. The spacing between the NF- κ B and IRF3 motifs is indicated at the right. The strengths of the κ B motifs are represented by PWM Z scores, and the strengths of the IRF motifs are represented by PWM Transfac scores. For the four genes lacking NF- κ B motifs, the best IRF promoter motif is shown.

(E) Left: the fold increase in ATAC-seq RPM at gene promoters (x axis) is shown according to the PRG clusters 1–10 (y axis) where the cluster designations denote 1, SRP; 2, MAPK; 3, MAPK/NF- κ B; 4, NF- κ B/I κ B regulator; 5, NF- κ B/other; 6, NF- κ B/IRF3; 7, NF- κ B/enhancer; 8, TRIF; 9, IRF3; 10, unknown (see also Figure S6). The vertical dashed lines indicate the 2.5- and 5-fold cutoffs. Right: UCSC Genome Browser tracks of chromatin accessibility in resting and 120 min stimulated BMDMs at the promoters of two genes from different gene clusters are shown.

(F) RelA ChIP-qPCR was performed using WT and *Irh3*^{-/-} BMDMs stimulated with lipid A. The relative enrichment of RelA binding was normalized to a negative control region. The RelA binding kinetics at the promoters of the five NF- κ B/IRF3 genes were compared to the *Tnfr1* promoter as a control (far right). The data shown represent an average of three biological replicates. Error bars indicate the SE. **p < 0.01; *p < 0.05.

Regulation of Transiently Transcribed Genes by Serum Response Factor

The most distinctive cluster of genes in Figure 3 is arguably the MAPK-dependent cluster 6, which contains genes that exhibit rapid yet transient upregulation within 5 min of lipid A stimulation. This cluster contains only three genes, *Egr1*, *Fos*, and *Nr4a1*, yet the initial motif analysis (Figure S3) suggests enrichment of promoter binding sites for serum response factor (SRF). We therefore examined SRF binding by ChIP-seq in BMDMs stimulated with lipid A for 0, 15, 30, 60, and 120 min. SRF peaks remained unchanged through the time course, consistent with knowledge that SRF binds its targets constitutively, with inducible activity due to the induction of co-regulatory ternary complex factors (TCFs, Treisman, 1994).

The SRF ChIP-seq datasets yielded the strongest peaks we have detected and the greatest specificity of binding, with only a small number of strong peaks and very little background. A simultaneous examination of ChIP-seq peaks and Transfac PWM-defined motifs revealed that only seven of the 132 PRGs contain promoters with strong ChIP-seq peaks (peak score >10); all seven promoters contain strong motifs (Transfac score >90) (Figure 7A). No strong ChIP-seq peaks were observed at these promoters in the absence of a strong motif, and only two promoters contained a strong motif without a strong ChIP-seq peak; both of these motifs are far from their TSS (–306 and –331), suggesting that they may be occluded by nucleosomes. Thus, to even a greater extent than observed with NF- κ B, strong binding of SRF correlated closely with motif strength, leading to a motif threshold that may be both necessary and sufficient for SRF binding in the context of a well-defined set of promoters.

Surprisingly, only 39 additional promoters within the remaining 21,036 annotated genes reached the same peak and motif thresholds achieved by the seven binding events at the primary response genes (Figures 7A and 7B). Instead, the vast majority of binding events at other gene classes coincided with weak motifs (Figures 7A and 7B).

A closer examination of the seven genes that combine strong SRF ChIP-seq peaks and motifs supports the hypothesis that at least six are functional targets of SRF. This group of seven genes includes the three found in cluster 6 of Figure 3A (*Egr1*, *Fos*, and *Nr4a1*) and four additional genes (*Egr2*, *Dusp5*, *Zfp36*, and *Rnd3*). All but *Rnd3* were initially upregulated during the first 5 min of lipid A stimulation (Figure 7C, third panel), and all but *Rnd3* exhibited MAPK dependence. MAPKs are responsible for activation of the TCFs (Treisman, 1994). The fact that *Rnd3* exhibited different properties suggests that this gene may instead require a second class of SRF co-activator proteins that are not activated by MAPKs (Posem and Treisman, 2006).

An examination of the expression kinetics of the seven genes explains why only three were placed in the same kinetic cluster in Figure 3A: these three genes exhibited relatively uniform induction and repression kinetics, whereas *Egr2*, *Dusp5*, and *Zfp36*, although initially induced by 5 min, were either further upregulated at later time points or were upregulated less potently and downregulated more slowly, presumably due to the contributions of other factors.

Last, an analysis of the 132 PRGs led to the identification of only two additional genes that exhibit similarly rapid induction ki-

netics as the six genes discussed above: *Btg2* and *Ier2* (Figure 7D). These two genes lack SRF ChIP-seq peaks and motifs in their promoters but instead contained strong promoter NF- κ B ChIP-seq peaks. This finding raises the question of how these two genes achieve induction kinetics similar to those of the genes whose promoters are directly bound by SRF. Interestingly, both of these genes contain strong SRF ChIP-seq peaks at upstream regions (Figure 7E); in both instances, the SRF peaks coincide with CpG islands and are conserved through evolution (data not shown). The SRF peaks at the *Btg2* and *Ier2* loci are 10 and 1 kb upstream of their TSSs, respectively. Remarkably, only three other PRGs contained SRF ChIP-seq peaks within the region 10 kb upstream of their TSS, indicating that this property is rare. These results suggest that SRF contributes to the early transient induction of these genes by cooperating with NF- κ B bound to the promoters.

DISCUSSION

Broad systems analyses of gene expression cascades and networks continue to provide important biological and mechanistic insights. However, the focus of most conventional studies on large numbers of genes or ChIP-seq peaks meeting low-stringency criteria, for the purpose of optimizing statistical power, possesses significant limitations. The results described here demonstrate that, toward the goal of a mechanistic understanding of transcriptional control at a genome-wide level, it is not only possible, but often preferable, to use more stringent and quantitative approaches to examine RNA-seq, ChIP-seq, and binding motif datasets.

This approach allowed us to obtain evidence that a single NF- κ B or SRF motif that reaches a defined threshold consistently supports factor binding and function *in vivo*. Moreover, we obtained evidence of an underlying logic through which NF- κ B may regulate distinct sets of genes by binding to promoters versus enhancers. We speculate that promoter binding may be compatible with transcriptional induction in response to any stimulus that induces NF- κ B activity, whereas enhancer binding may often be preferred at genes that require cell-type-specific induction. Of greatest interest, our results reveal that, even when two inducible factors (e.g., NF- κ B and IRF3) act in concert to regulate a small cluster of only five genes, individual genes within the cluster are regulated by unique mechanisms. Overall, the results of this analysis provide a wealth of mechanistic insights that are accessible to future exploration.

Figures S6 and S7 display summaries of the results obtained in this study. Notably, although the study derived great benefit from detailed kinetic analyses of chromatin-associated transcripts, diverse overall expression kinetics are observed within each cluster. This observation supports the long-standing view that multiple transcription factors act in concert to shape the expression pattern of each gene.

The findings described here contrast with ChIP-seq studies that implicate key transcription factors in the regulation of hundreds or thousands of genes. It is important to emphasize that our study focused on the properties of the limited number of potentially induced genes for which NF- κ B, IRF3, and SRF appear to be major regulators, but they do not rule out the possibility that

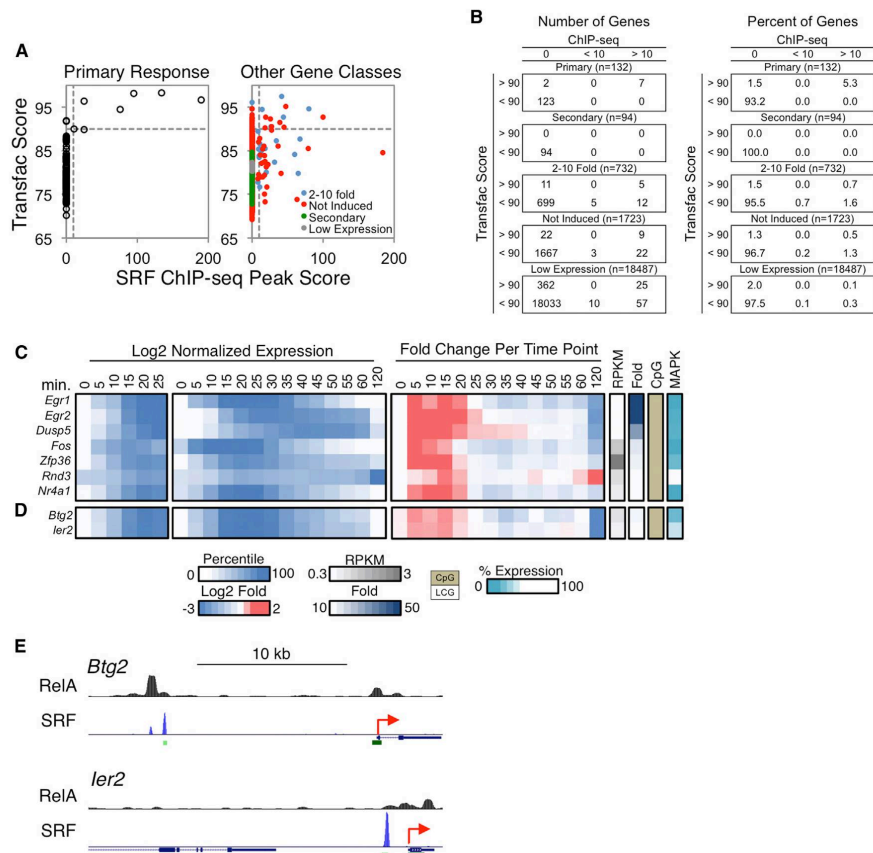


Figure 7. Analysis of SRF Target Genes

(A) Scatterplots comparing the Transfac PWM scores of SRF binding motifs (y axis) versus the SRF ChIP-seq peak scores (x axis) in the promoters (-500 to +150) of the PRGs (left) and all remaining genes in the genome (right) are shown. The genes in the latter graph were divided into categories as in Figure 4A. The horizontal and vertical dashed lines indicate the SRF motif (90) and ChIP-seq peak (10) thresholds.

(B) Tables are shown indicating the distribution of genes from (A), for both numbers (left) and percentages (right) of genes.

(C) Log2 normalized expression values from 0 to 25 min (first panel), 0 to 120 min (second panel), and the fold induction relative to the expression level at the previous time point (third panel) are shown for the seven putative SRF target genes. To the right are columns indicating the basal expression level, fold induction magnitude, promoter CpG content, and MAPK dependence for each gene.

(D) Two genes that exhibited similar activation kinetics as the putative SRF target genes are shown, with the same layout as in Figure 7C.

(E) The two genes from (D) were examined on UCSC Genome Browser to identify distal SRF binding peaks. RelA binding peaks were also examined for these genes. The TSSs of the genes are indicated as red arrows, and the green rectangles indicate CpG islands.

See also Figures S6 and S7.

these same factors play more subtle roles in the regulation of hundreds of additional genes. For example, although only a small fraction of genes induced by 2- to 10-fold contain promoter binding sites for NF- κ B, this factor may contribute to the induction of a large fraction of weakly induced genes by binding to distant enhancers. Nevertheless, the current results document a clear distinction between strongly induced and weakly induced genes

with respect to the prevalence of NF- κ B promoter binding and promoter motifs, providing a framework for studies to elucidate the diverse mechanisms by which NF- κ B contributes to the lipid A response. Similarly, the ability to identify consistent properties of genes that appear to be regulated by SRF, IRF3, and IRF3/NF- κ B provides a step toward a precise understanding of the broad mechanisms regulating the transcriptional cascade. In addition

to exploring the insights obtained in greater depth, an important goal for the future will be to continue building on this framework by using stringent approaches to examine additional signaling pathways, transcription factors, and chromatin regulators, while extending the analysis to enhancers, weakly induced genes, and gene expression cascades induced by diverse stimuli in diverse cell types and physiological settings.

EXPERIMENTAL PROCEDURES

Cell Culture

BMDMs were prepared from 6-week-old C57BL/6, *Myd88*^{-/-}, *Trif*^{-/-}, *Irf3*^{-/-}, or *Ifnar*^{-/-} male mice. Fetal liver macrophages were from D14.5 C57BL/6 or *RelA*^{-/-} embryos. Macrophages were activated on day 6 with 100 ng/ml lipid A (Sigma) or Pam3CSK4 (InvivoGen). When indicated, cells were preincubated for 15 min with 10 mg/ml CHX or 1 hr with 10 μ M PD0325901 (Sigma) and 1 μ M BIRB0796 (AXON Medchem). The use of mice for this study was specifically approved by the UCLA Chancellor's Animal Research Committee.

RNA-Seq

Total RNA and chromatin-associated RNA were prepared as described (Shatt et al., 2012). Strand-specific libraries were generated from 60 ng chromatin RNA or 400 ng total RNA using the TruSeq RNA Sample Preparation Kit v2 (Illumina) and the dUTP second strand cDNA method (Levin et al., 2010). cDNA libraries were single-end sequenced (50bp) on an Illumina HiSeq 2000.

Reads were aligned to the mouse genome (NCBI37/mm9) with TopHat v1.3.3 and allowed one alignment with up to two mismatches per read. Chromatin RNA RPKM values were calculated by dividing all mapped reads within the transcription unit by the length of the entire locus. mRNA RPKM values were calculated by dividing mapped exonic reads by the length of the spliced product.

All RPKMs represent an average from two or three biological replicates. A gene was included in the analysis if it met all of the following criteria: the maximum RPKM reached 3 at any time point, the gene was induced at least 10-fold, and the induced expression was significantly different from the basal ($p < 0.01$) as determined by the edgeR package in R Bioconductor (Robinson et al., 2010). Additionally, a gene was included if its induction reached 5-fold at the 15-min time point. p values were adjusted using the Benjamini-Hochberg procedure of multiple hypothesis testing (Benjamini and Hochberg, 1995).

To determine the impact of a perturbation, the basal RPKM in WT samples was set at 0% and the maximum RPKM at 100% for each gene. The maximum RPKMs in the mutant samples were converted to percent expression using this scale. For the *RelA*^{-/-} analysis, the *RelA* dependence of a gene was determined by the percent expression in *RelA*^{-/-} samples at the earliest time point in which the WT samples were induced by at least 3-fold.

ChIP-Seq

ChIP-seq libraries were prepared using the Kapa LTP Library Preparation Kit (Kapa Biosystems). ChIP-seq was performed as described (Barish et al., 2010; Lee et al., 2006) with minor modifications, using anti-RelA (Santa Cruz Biotechnology, sc-372), anti-IRF3 (Santa Cruz, sc-9082), or anti-SRF (Santa Cruz, sc-335) antibodies.

Reads were aligned to the mouse genome (NCBI37/mm9) with Bowtie2. Uniquely mapped reads were used for peak calling and annotation using HOMER (Heinz et al., 2010). Peaks were called if they passed a false discovery rate of 0.01 and were enriched over input. Called peaks were considered for downstream analysis if peaks from at least 4 of 7 replicates were overlapping within 200 bp for RelA and five of five replicates were overlapping within 300 bp for SRF. Peaks were annotated to the nearest TSS.

ATAC-Seq

ATAC-seq libraries were prepared using the Nextera Tn5 Transposase kit (Illumina) as described (Buenrostro et al., 2015) with slight modifications. Libraries were single-end sequenced (50 bp) on an Illumina HiSeq 2000. Reads were mapped to the mouse genome (NCBI37/mm9) using Bowtie2. Reads

were removed from the subsequent analysis if they were duplicated, mapped to mitochondrial genome, or aligned to unmapped contiguous sequences. Promoter accessibility was calculated by totaling all reads within -500 to +150 bp relative to the TSS. The reads were converted to reads per million (RPM) by dividing by the total number of reads per sample. The average RPM from four replicates was used to quantify the fold increase in promoter accessibility.

Motif Analysis

The promoters of genes (-500 to +150 bp) were used for motif analysis unless otherwise indicated. The strongest p50:RelA binding site within each promoter was identified using a PBM dataset (Siggers et al., 2012). Transfac PWMs were used to identify the best matching SRF and IRF3 binding sites in promoters using Pscan (Zambelli et al., 2009).

ACCESSION NUMBERS

All data are accessible through GEO Series accession number GSE67357.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2016.01.020>.

AUTHOR CONTRIBUTIONS

A.-J.T., X.L., B.J.T., and M.M.L. designed and performed most experiments and wrote the manuscript. M.R.B., M.D.S., A.L.A., and G.D.B. performed the RelA ChIP-seq experiments, and S.T.S. designed experiments and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Christopher Glass, Alexander Hoffmann, Steven Ley, Ranjan Sen, and Trevor Siggers for helpful discussions, and the UCLA Broad Stem Cell Research Center Core for sequencing. This work was supported by NIH grants R01GM086372 (S.T.S.), P50AR063030 (S.T.S.), T32CA009120 (A.-J.T.), and T32GM008042 (B.J.T.), and by the China Scholarship Council and Whitcome pre-doctoral training program (X.L.).

Received: July 17, 2015

Revised: November 1, 2015

Accepted: January 13, 2016

Published: February 25, 2016

REFERENCES

- Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN- β promoter. *Cell* 103, 667–678.
- Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326, 257–263.
- Barish, G.D., Yu, R.T., Karunasiri, M., Ocampo, C.B., Dixon, J., Benner, C., Dent, A.L., Tangirala, R.K., and Evans, R.M. (2010). Bcl-6 and NF-kappaB cis-tromes mediate opposing regulation of the innate immune response. *Genes Dev.* 24, 2760–2765.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J.R. Stat. Soc.* 57, 289–300.
- Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* 150, 279–290.

- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* *109*, 21.29.1–21.29.9.
- 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.
- Freaney, J.E., Kim, R., Mandhana, R., and Horvath, C.M. (2013). Extensive cooperation of immune master regulators IRF3 and NF κ B in RNA Pol II recruitment and pause release in human innate antiviral transcription. *Cell Rep.* *4*, 959–973.
- Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* *47*, 810–822.
- Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.-L., et al. (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* *32*, 317–328.
- Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* *138*, 129–145.
- 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.
- Kagan, J.C., Su, T., Horng, T., Chow, A., Akira, S., and Medzhitov, R. (2008). TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. *Nat. Immunol.* *9*, 361–368.
- Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* *1*, 729–748.
- Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* *7*, 709–715.
- Lim, C.-A., Yao, F., Wong, J.J.-Y., George, J., Xu, H., Chiu, K.P., Sung, W.-K., Lipovich, L., Vega, V.B., Chen, J., et al. (2007). Genome-wide mapping of RE-LA(p65) binding identifies E2F1 as a transcriptional activator recruited by NF-kappaB upon TLR4 activation. *Mol. Cell* *27*, 622–635.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* *18*, 1509–1517.
- Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., et al. (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc. Natl. Acad. Sci. USA* *100*, 12247–12252.
- Metcalfe, S.M. (2011). LIF in the regulation of T-cell fate and as a potential therapeutic. *Genes Immun.* *12*, 157–168.
- Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. *Cell* *152*, 157–171.
- Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. *Cell* *129*, 1111–1123.
- Posern, G., and Treisman, R. (2006). Actin' together: serum response factor, its cofactors and the link to signal transduction. *Trends Cell Biol.* *16*, 588–596.
- Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* *29*, 436–442.
- Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* *138*, 114–128.
- Ramsey, S.A., Klemm, S.L., Zak, D.E., Kennedy, K.A., Thorsson, V., Li, B., Gilchrist, M., Gold, E.S., Johnson, C.D., Litvak, V., et al. (2008). Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS Comput. Biol.* *4*, e1000021.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* *26*, 139–140.
- Shih, H.Y., Sciumè, G., Poholek, A.C., Vahedi, G., Hirahara, K., Villarino, A.V., Bonelli, M., Bosselut, R., Kanno, Y., Muljo, S.A., and O'Shea, J.J. (2014). Transcriptional and epigenetic networks of helper T and innate lymphoid cells. *Immunity* *40*, 23–49.
- Siggers, T., Chang, A.B., Teixeira, A., Wong, D., Williams, K.J., Ahmed, B., Ragoussis, J., Udalova, I.A., Smale, S.T., and Bulyk, M.L. (2012). Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF- κ B family DNA binding. *Nat. Immunol.* *13*, 95–102.
- Toshchakov, V., Jones, B.W., Perera, P.-Y., Thomas, K., Cody, M.J., Zhang, S., Williams, B.R.G., Major, J., Hamilton, T.A., Fenton, M.J., and Vogel, S.N. (2002). TLR4, but not TLR2, mediates IFN-beta-induced STAT1alpha/beta-dependent gene expression in macrophages. *Nat. Immunol.* *3*, 392–398.
- Treisman, R. (1994). Ternary complex factors: growth factor regulated transcriptional activators. *Curr. Opin. Genet. Dev.* *4*, 96–101.
- Zambelli, F., Pesole, G., and Pavesi, G. (2009). Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.* *37*, W247–W252.
- Zhao, B., Barrera, L.A., Ersing, I., Willox, B., Schmidt, S.C.S., Greenfield, H., Zhou, H., Mollo, S.B., Shi, T.T., Takasaki, K., et al. (2014). The NF- κ B genomic landscape in lymphoblastoid B cells. *Cell Rep.* *8*, 1595–1606.

Supplemental Figures

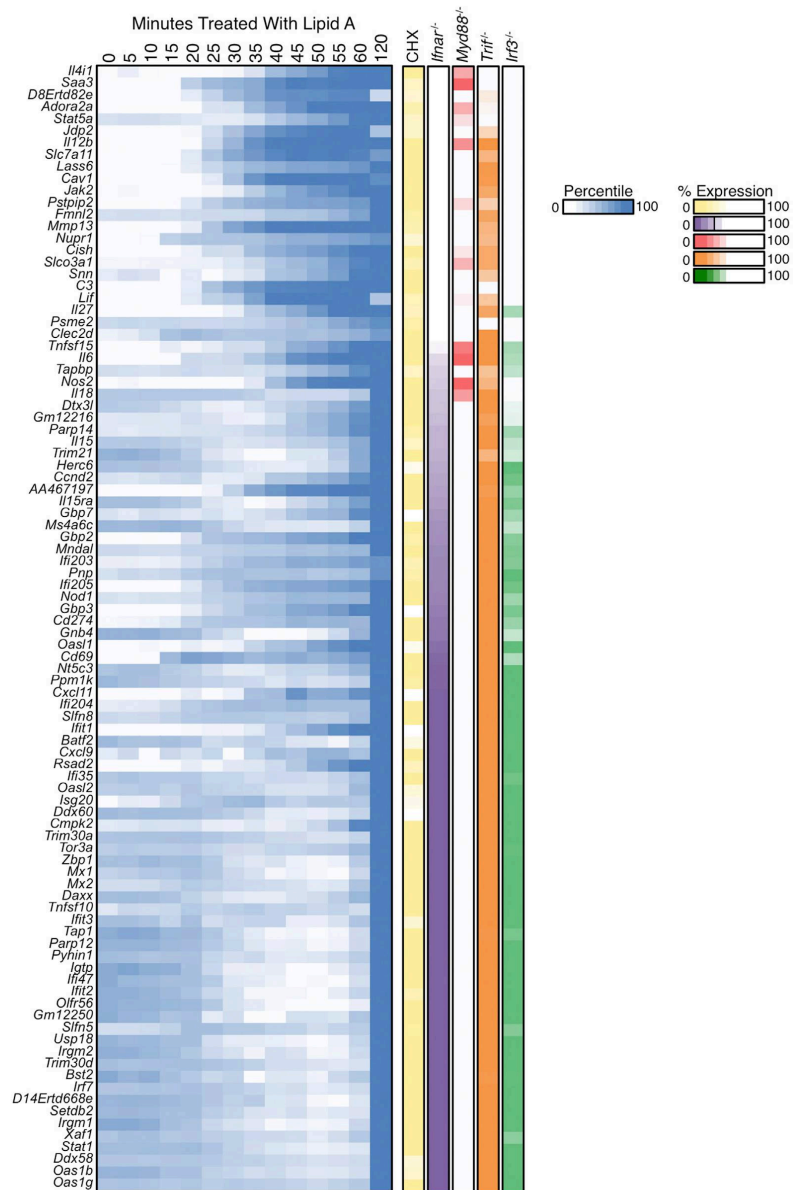


Figure S1. IFNAR-Independent and IFNAR-Dependent Secondary Response Genes, Related to Figure 2

An expanded version of Figure 2A is shown to include gene names for each 10-fold significantly induced secondary response gene. This expanded version also includes a column indicating the percent expression in cycloheximide-treated BMDMs.



Cluster Assignment

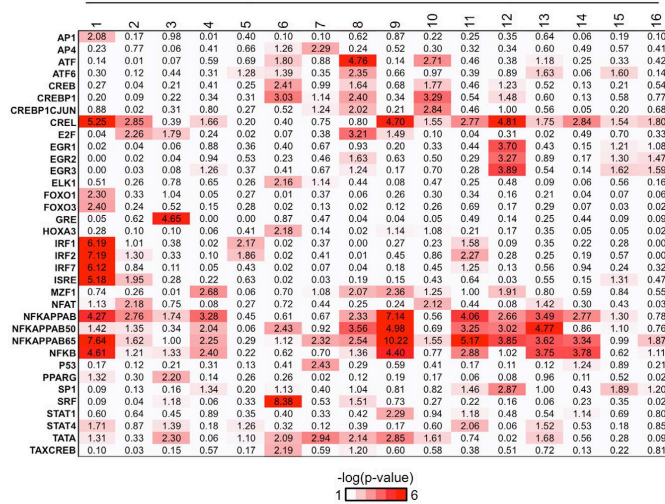


Figure S3. Promoter Motif Analysis of Primary Response Gene Clusters, Related to Figure 3
 Overrepresented transcription factor binding motifs are shown for each cluster, 1-16. The genes were clustered as described in Figure 3D. The transcription factor families are shown on the left, in alphabetical order. The color intensity is proportional to the negative log of the p value.

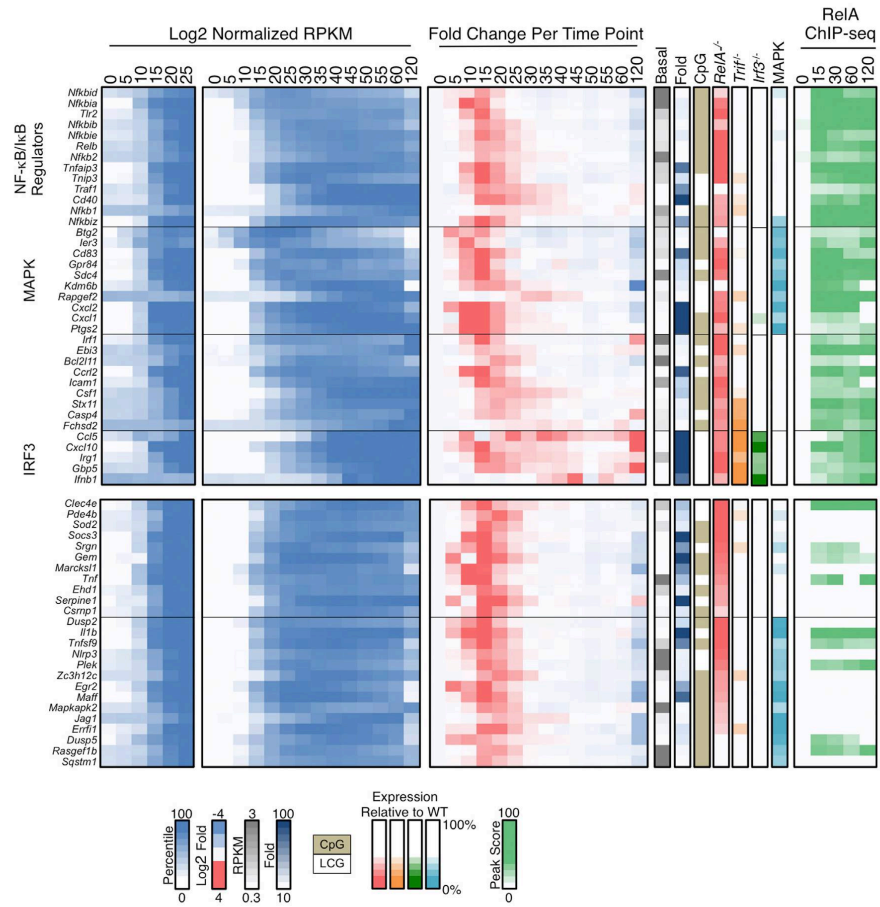


Figure S4. Putative NF-κB Target Genes and the Genes that Exhibit Similar Kinetics and/or RelA Dependence, Related to Figure 5
 An expanded version of Figures 5A and 5B is shown to include gene names for each putative NF-κB target and other genes that may be enhancer regulated NF-κB target genes.

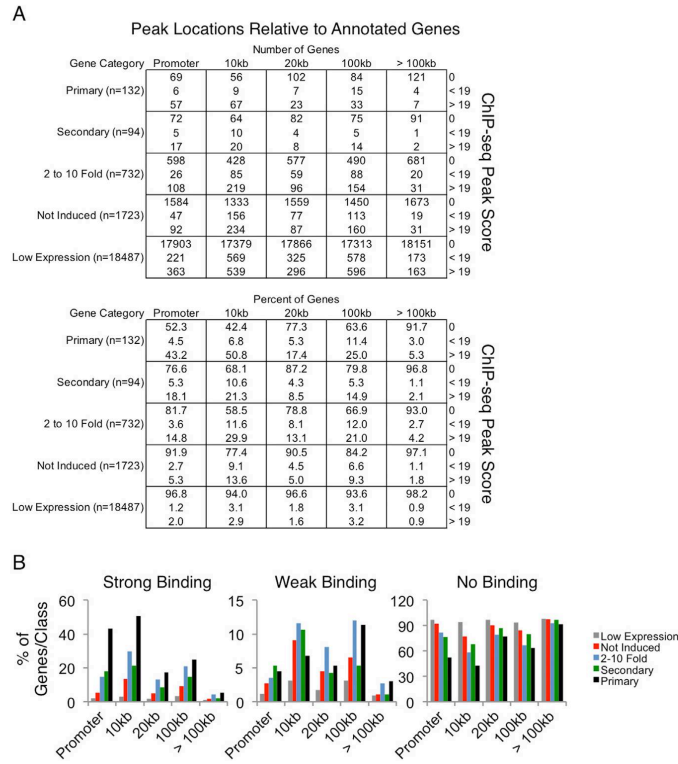


Figure S5. The Position of RelA Peaks Relative to the Transcription Start Sites of All Genes, Related to Figure 4
 (A) For each annotated gene in each gene category (primary, secondary, 2-10 fold induced, not induced but expressed, and unexpressed), RelA binding peaks were identified at the following distance ranges relative to the transcription start site (TSS): promoter, 10 kb, 20 kb, 100 kb, and >100 kb. The promoter was designated as the region spanning -500 to +150 relative to the TSS. Peaks included those identified either upstream or downstream from the TSS. The annotated RelA peaks were then grouped based on their ChIP-seq peak score (>19 or <19). If a gene did not have a peak in the indicated region, a score of 0 was given to that gene. The top table represents the number of genes in each group, and the bottom table indicates the percent of genes in each group relative to the gene class.
 (B) The distribution of RelA peaks as shown in the bottom table of (A) is shown as a bar graph. Strong binding indicates a RelA peak score >19, and weak binding indicates a RelA peak score <19.

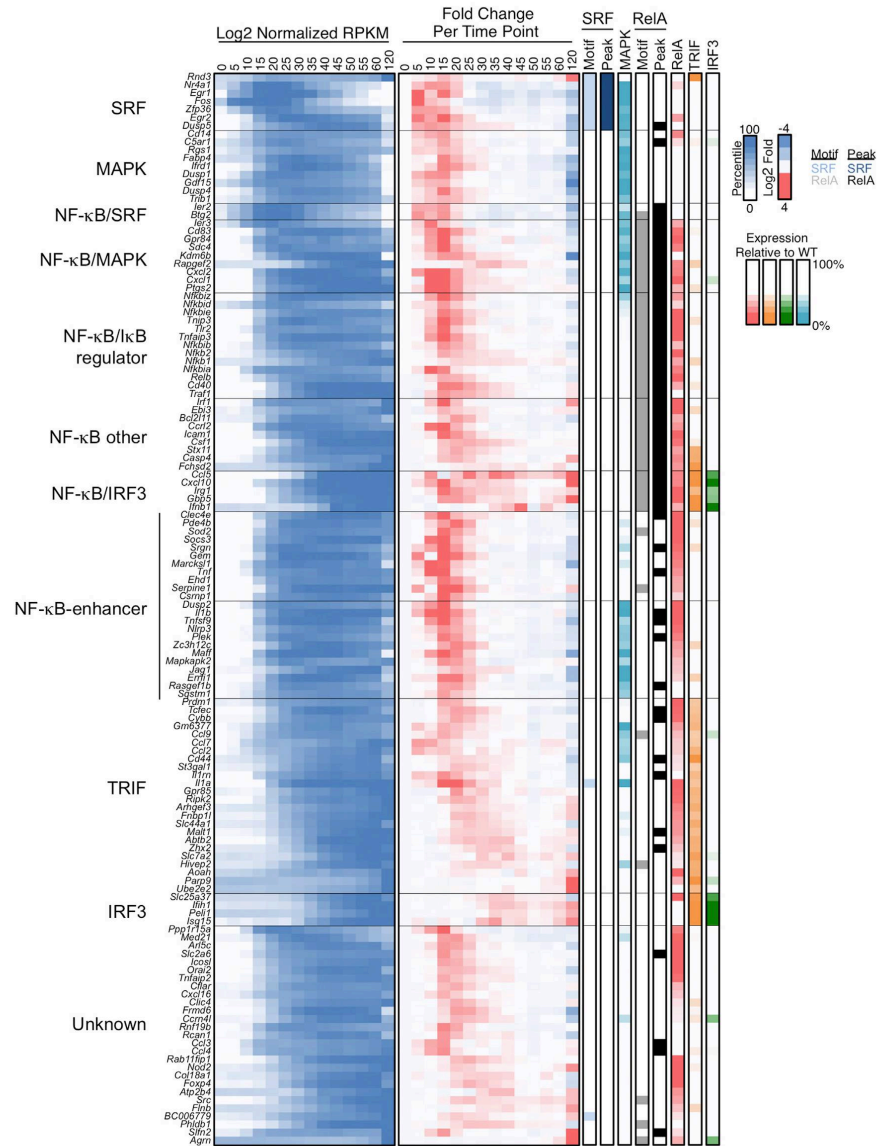


Figure S6. Final Classification of the Primary Response Genes, Related to Figure 7
 The 132 primary response genes were grouped based on their regulation by SRF or RelA, dependence on MAPK, TRIF, or IRF3. The left heatmap represents log₂ normalized expression values, and the right heatmap represents the log₂ fold change relative to the previous time point. To the right of the heatmaps are columns indicating the following, from left to right: the presence of a strong SRF motif, a strong SRF binding peak, expression in MAPK-inhibited BMDMs, a strong RelA motif, a strong RelA binding peak, expression in *Rela*^{-/-} FLMs and in *Trif*^{-/-} and *Irf3*^{-/-} BMDMs. See also Figure S7.

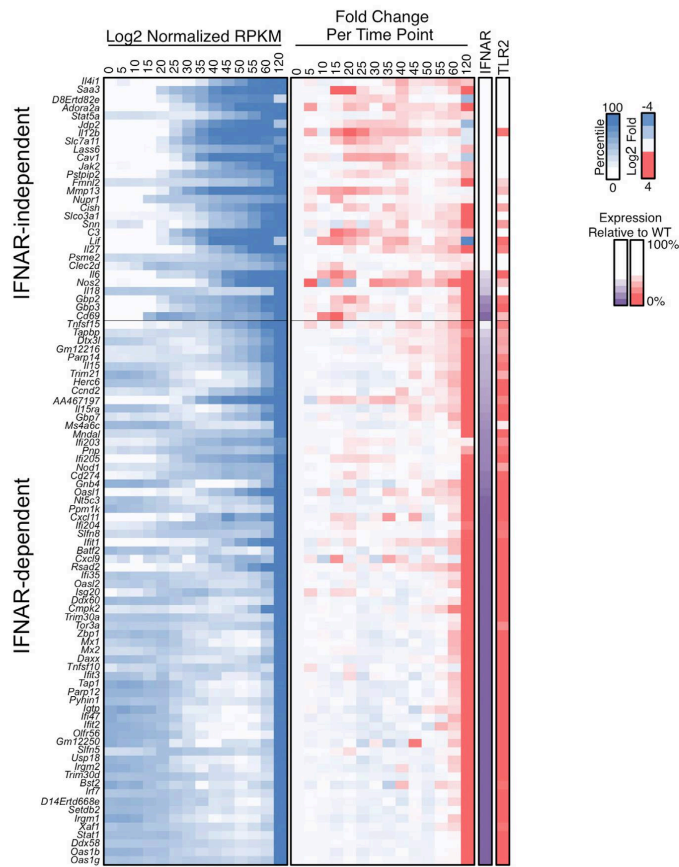


Figure S7. Final Classification of the Secondary Response Genes, Related to Figure 7
 The 94 secondary response genes were grouped based on their dependence on IFNAR. To the right of the heatmaps are columns indicating the following, from left to right: expression in *Ifnar*^{-/-} BMDMs, and expression in Pam-stimulated WT BMDMs. See also Figure S6.

CHAPTER 3

A Gene-Centric Approach Unravels the Selective and Combinatorial Regulation of SRF Targets

Abstract

Although ChIP-seq data reveal that a transcription factor can bind to thousands of genomic sites, expression data suggest that only a subset of these sites contribute to transcription. This discrepancy indicates that transcription factor binding is often insufficient for transcriptional activation and the functional sites require selective regulation. In this study, we described a gene-centric, stringent system approach to investigate the selective mechanisms regulating serum response factor (SRF) in a lipid A response. SRF can activate transcription through the combinatorial interaction with cofactors such as ternary complex factor (TCF). By combining well-defined gene clusters, motif data, and ChIP-seq data, we found that the Toll-like receptor 4 (TLR4) agonist lipid A selectively and strongly activated a small number of SRF targets. All promoter-activated SRF targets share a unique, strict, and conserved SRF-TCF cassette. This SRF-TCF cassette is essential for the strong promoter SRF binding and the cooperative binding of TCF. Furthermore, formation of ternary complex at the SRF-TCF cassette also requires an open chromatin at the promoter. Enhancer-activated SRF targets possess SRF and TCF motifs similar to promoter SRF-TCF cassettes. The function of SRF-TCF cassettes is further confirmed in different cell types and by different stimuli. These findings demonstrate that the selective and combinatorial regulation of SRF targets require strict motif rules. They also exemplify a gene-centric, stringent system approach to identify functional binding events that directly contribute to transcriptional induction.

Introduction

Discovered thirty years ago, serum response factor (SRF) is a well-studied transcription factor that led to many important discoveries in gene regulation. SRF was originally discovered as a factor that can rapidly activate *c-fos* gene within 15 minutes of serum or growth factor stimulation (Greenberg and Ziff, 1984; Treisman, 1986; Norman et al., 1988). SRF binds to a consensus sequence CC(A/T)₆GG, named the CArG box. This CArG box exists at the promoters of many SRF targets. SRF targets often respond rapidly to environmental stimuli through existing signaling molecules. Thus, they are called immediate early genes, or primary response genes. Conversely, genes that exhibit late kinetics and require new protein synthesis are called secondary response genes (K R Yamamoto et al., 1976; Herschman, 1991). The separation of primary and secondary response genes is often the first step to dissect stimulus-induced transcriptional cascades (Ramirez-Carrozzi et al., 2009; Tong et al., 2016).

Serum response factor (SRF) is an MADS (MCM1, Agamous, Deficiens, and SRF) box family transcription factor, which contains a conserved MADS box domain for DNA-binding. Mammalian SRF is ubiquitously expressed in many cell types and is critical for cell survival and cellular responses to many environmental stimuli. SRF can react to mitogens including serum, growth factors, cytokines, lipopolysaccharide, and tetradecanoylphorbol-13-acetate. Mitogens can activate SRF targets regulating cell cycle, cell proliferation, apoptosis, cell differentiation, circadian rhythm, and cytoskeleton functions (Posern and Treisman, 2006; Medjkane et al., 2009; Gerber et al., 2013). Cell-specific activation of SRF targets is often involved in maintaining cell identity and

regulating cell-specific functions (Miano, 2010). For example, SRF can collaborate with muscle-specific factors to regulate the expression of many muscle-specific genes regulating the development of cardiac muscle cells, smooth muscle cells, and skeletal muscle cells. Heart-specific mutation of SRF impairs cardiac development and is embryonic lethal (Parlakian et al., 2004). In hematopoietic stem cell progenitors, SRF is responsible for cell migration and cell seeding in response to chemokine signaling (Costello et al., 2015). In macrophages, SRF is essential for phagocytosis by regulating cytoskeletal functions (Sullivan et al., 2011).

Although SRF can undergo several post-translational modifications, active SRF is insufficient for gene activation. SRF activity also depends on the combinatorial interaction with co-factors. SRF binds to DNA constitutively and interacts with at least two competing co-factors: ternary complex factor (TCF) transcription factors (Elk-1, SAP1, and Net), and myocardin-related transcription factors (MRTF) (myocardin, MRTF-A, MRTF-B). Stimulus-induced MAPK pathway can phosphorylate and activate TCF (Dalton and Treisman, 1992; Posern and Treisman, 2006). TCF binds to an Ets site and interacts with SRF through a flexible domain, which allows for stable ternary complex formation with variable spacing between the Ets site and the CA₂G box (Treisman et al., 1992). Because the formation of ternary complex stabilizes TCF binding on DNA, TCF can tolerate a suboptimal Ets site. Among MRTF transcription factors, myocardin is selectively expressed and constitutively active in smooth muscle and cardiac cells (Pipes et al., 2006; Posern and Treisman, 2006). The other two MRTF members, MRTF-A and MRTF-B, are widely expressed in many cell types and are

induced by Rho-actin pathway. In resting cells, actin monomer binds to MRTF and inhibit MRTF nuclear translocation. Activated Rho pathway induces actin polymerization and releases actin from MRTF, resulting in MRTF activation and nuclear translocation (Vartiainen et al., 2007; Pawłowski et al., 2010). MRTF does not bind directly to DNA in a sequence-specific manner; it is recruited to the SRF motif by SRF to activate transcription. Because MRTFs and TCFs bind to the same interface on SRF, they can compete for SRF binding, resulting in either gene activation or repression (Murai and Treisman, 2002; Zaromytidou et al., 2006; Wang et al., 2004; Gualdrini et al., 2016). Therefore, by using different co-factors, SRF targets can integrate different extracellular signals into transcriptional outcomes and balance different cellular responses.

Previous biochemical and computational studies have given us many insights on the mechanisms regulating SRF targets. Site selection experiments discovered that the consensus motif of SRF, the CArG box, is a degenerate sequence (Pollock and Treisman, 1990). Computational analyses have identified hundreds of potential promoter CArG boxes in human and mouse genomes (Sun et al., 2006). The emerging next-generation sequencing technologies, including chromatin-immunoprecipitation sequencing (ChIP-seq) and Hi-C, uncovered thousands of *in vivo* SRF binding sites in different cell types. SRF ChIP-seq data suggest that SRF can bind to canonical CArG boxes as well as non-canonical CArG boxes with mismatches (Esnault et al., 2014; He et al., 2011; Sullivan et al., 2011). SRF binding at different genomic locations contribute to the differential gene expression patterns. For example, SRF binds to promoters to regulate ubiquitously expressed genes; SRF binds to enhancers established by pioneer

factors to activate cell-specific genes (Sullivan et al., 2011; Kim et al., 2010; He et al., 2011). Because SRF activity is dependent on co-factor interaction, most SRF peaks are constitutive and coincide with either MRTF or TCF binding (Gualdrini et al., 2016; Esnault et al., 2014). SRF ChIP-seq data also suggest that SRF binding is insufficient for gene activation. Selective recruitment of MRTF or TCF factors is critical for transcriptional activation (Vasudevan and Soriano, 2014). Although previous SRF ChIP-seq analyses have revealed many characteristics of SRF, their peak-centric statistical approaches are restricted by large sample sizes and are therefore biased towards co-regulatory mechanisms. To unveil the selective mechanisms regulating SRF function, we need to use a different strategy.

In this study, we describe a gene-centric approach to identify the motif rules that selectively activate SRF targets in the innate immune response. By identifying SRF targets and studying the motif characteristics, binding strength, and DNA context, we prevented the interference of non-functional sites and focused on the mechanisms controlling functional SRF sites. This enabled us to characterize a small subset of SRF targets sharing a unique promoter SRF-TCF cassette responsible for SRF activity. Owing to the relative few ChIP-seq peaks and abundant evidence from previous studies, SRF serves as a good example to optimize the strategies to interrogate the selective mechanisms regulating transcription.

Results

General Features of SRF ChIP-Seq in Macrophages

To understand the mechanisms regulating SRF activity in the lipid A response, we performed ChIP-seq experiments in bone marrow-derived macrophages (BMDMs) stimulated with lipid A for 0, 15, 30, 60, and 120 minutes. We eliminated the non-specific peaks that also showed up in SRF knockout samples (Sullivan et al., 2011). Consistent with the previous studies, SRF binds to DNA before stimulation (Li et al., 2003; Nissen et al., 2001). Most peaks bind constitutively and remain unchanged within two hours of lipid A stimulation (Figure 3-1A). Because weaker peaks are less reproducible and are more likely to represent technical artifacts, we treated samples of five time points as replicates and uncovered 718 reproducible peaks reaching a peak score threshold of 10. Strikingly, these SRF peaks are highly enriched in the promoter region (Figure 3-1B and C). Almost one fifth of the SRF peaks fall into the promoter region (-500 to +150 relative to TSS).

Although more than half of the promoters contain CpG islands, 91.6% of the SRF promoter peaks contain CpG islands with an obs/expt CpG ratio greater than 0.6 (Figure 3-1D). Because CpG-island promoters are too rigid to form stable nucleosomes, they are generally constitutively accessible to transcription factors without the need of chromatin remodeling (Ramirez-Carrozzi et al., 2009). Thus, they are often associated with housekeeping genes and other ubiquitously expressed genes. The prevalence of CpG-island promoters might suggest that SRF binding at promoters tend to regulate ubiquitously expressed genes. In contrast to promoters, fewer than 15% of the SRF

enhancer peaks have CpG islands (Figure 3-1E), and their CpG content is similar to active enhancer peaks identified by H3K27ac ChIP-seq peaks. The differential CpG content in SRF promoter and enhancer peaks might simply reflect the depletion of CpG in non-promoter regions of mammalian genomes. It might also reflect the different strategies that SRF deploys to regulate genes involved in different biological functions. One study found that SRF regulates ubiquitously expressed genes by promoter binding and regulates cell-specific genes by binding to enhancers established by pioneer factors (Sullivan et al., 2011; Heinz et al., 2010). Because promoters are better characterized than enhancers, we decide to start our analysis from SRF promoter peaks.

Identify SRF Targets Induced by Lipid A

In macrophages, lipid A-induced TLR4 signaling can trigger the MAPK pathway, which in turn phosphorylates TCF and induces transcriptional activation of SRF targets. Previously, we stimulated bone marrow-derived macrophages (BMDMs) with lipid A and monitored nascent transcript expression by collecting samples every five minutes during the first hour of stimulation (Tong et al., 2016). To prevent the interference of mRNA stability, we isolated nascent transcripts associated with chromatin and measured transcript levels by RNA-sequencing. Because SRF targets are often transiently induced between 10 to 30 min, a large number of time points can capture the peak transcription of SRF targets. Using stringent criteria that include a fold change threshold of 10 and a maximal expression threshold of 3 RPKM, we identified 226 lipid A-induced genes.

To uncover the common features and potential rules for SRF to activate transcription, we focused on the well-defined SRF targets. We found that lipid A activated only eight SRF targets (Figure 3-2A). These eight SRF targets are distinct from other lipid A-induced genes by their early transient transcriptional kinetics. They are activated by 5 min and peaked by 20 min. They are transiently induced because the upstream of SRF-TCF, MAPK, is transiently activated during the lipid A response. Consistently, all SRF targets are strongly repressed by a cocktail of MAPK inhibitors (ERK inhibitor, PD0325901; p38 inhibitor, BIRB796). All eight SRF targets are categorized as primary response genes, which depend on existing signaling molecules and are insensitive to the inhibitor of translation, cycloheximide. Six of the SRF targets (*Egr1*, *Egr2*, *Dusp5*, *Fos*, *Zfp36*, *Nr4a1*) have SRF binding and CpG islands at their promoters (Figure 3-2B). Another two SRF targets (*Ier2*, *Btg2*) have SRF binding at enhancer regions within 10kb upstream of their transcription start sites (TSSs), which coincide with the active enhancer mark H3K27ac (Figure 3-2B).

Interrogate Functional SRF Peaks Combining Motif Data and ChIP-seq Data.

To understand the requirements for functional SRF binding that lead to gene activation, we evaluated the correlation of promoter motif strength and peak strength of different gene classes. By their activation mechanisms and transcriptional kinetics, we separated the mouse genome into five categories: primary response genes, secondary response genes, 2- to 10-fold induced genes, not induced genes, and low expression genes. We compared the promoter SRF ChIP-seq peak score and the motif score using the Transfac database (Wingender et al., 1996; Kaplun et al., 2016). Only primary response

genes exhibit a clear correlation between promoter motif strength and peak strength. While no weak promoter SRF motif has binding, seven of nine primary response genes with strong promoter motifs have SRF binding (Figure 3-3A). Notably, six of these seven strong promoter motifs belong to previously defined SRF targets by RNA-seq. Although another two primary response genes possess strong promoter SRF motifs, their SRF sites reside farther (-306 and -331) upstream of transcription start sites than SRF targets, and therefore they may be occupied by nucleosomes and are inaccessible to SRF binding. Consistently, these two SRF sites are not conserved in mammalian species. In contrast to primary response genes, none of the other gene classes show such strong correlation between strong motif and strong binding. While 5.3% of the primary response genes contain strong promoter SRF motif with strong binding, only less than 0.7% of genes in the other classes show similar strong binding and strong motif (Figure 3-3A and B). Many classes have promoter SRF binding without strong SRF motifs, which might represent indirect or non-functional binding. The enrichment of strong motifs and strong binding at primary response genes suggests that strong motifs combining strong binding is necessary for SRF transcriptional activity in a lipid A response.

Because functionally important *cis*-regulatory elements are often conserved, we quantified the conservation score of the SRF motifs by PhastCons (Siepel et al., 2005). In primary response genes, six of the seven strong promoter motifs with strong binding are highly conserved with a score greater than 0.997 (Figure 3-3C, in black). In contrast, only 6 of 123 weak promoter SRF motifs of primary response genes reach such high

conservation score. Interestingly, none of the other two strong SRF motifs without binding is conserved, suggesting they might not contribute to transcription in the lipid A response. Promoters with strong SRF motifs and strong binding are more conserved regardless of their gene classes. This indicates that strong motifs with strong binding might be functionally important in different contexts. Although many promoters do not respond to lipid A stimulation, they might react to other stimuli through different mechanisms. The enhancer sites, however, show a different trend. Although they are more conserved than the predicted strong enhancer SRF motifs without binding, enhancer peaks with strong motifs are less conserved than promoter peaks with strong motifs. It is possible that many enhancer SRF peaks react to cell-specific and stimulus-specific signals, and therefore they are less conserved than the promoter sites regulating ubiquitously expressed genes, which are essential for developmental and cellular functions. Unlike promoter peaks, strong enhancer SRF motifs do not exhibit stronger conservation than weak enhancer SRF motifs. This indicates that the regulation of SRF at promoter and enhancer sites require different mechanisms.

Identify a Promoter SRF-TCF Cassette Required for Gene Induction

Although strong SRF motifs with strong binding are highly correlated with transcriptional activation by lipid A, another 39 promoters with strong SRF motifs and strong binding failed to strongly induce transcription. This implies that either SRF activity depends on other factors such as TCF, or there are subtle differences in SRF motifs that motif score cannot reveal. To address these possibilities, we compared the promoter SRF motifs in different gene classes. Because ETS family transcription factor TCF can bind to SRF

and form a ternary complex, which can stabilize TCF binding and tolerate a suboptimal Ets site near SRF site, we also scanned for the nearest Ets site using the motif-searching tool Pscan (Zambelli et al., 2009).

Intriguingly, all six SRF targets with promoter SRF binding share a strict SRF-TCF cassette (Figure 3-4A). Although the reported SRF consensus sequence can tolerate either A or T in the center six base pairs, we found all six SRF targets shared a more stringent SRF motif, which includes an ATA sequence on one side of the motif. These sites also contain a consensus Ets motif that is within 19bp of SRF motif. This is consistent with the previous discovery that TCF can physically interact with SRF through a flexible domain and tolerate short and variable spacing between the two motifs (Treisman et al., 1992). Thus, we have defined a unique SRF-TCF cassette shared by all six SRF targets: an SRF motif CCATA(A/T/C)(A/T)(A/T)GG, a TCF motif (CC/CA/GA)GGA, with 3 to 19-bp motif spacing. We also noticed that an additional primary response gene, *Rnd3*, also contains an SRF motif and an Ets motif, but the motif spacing between the two sites is only 1bp. *Rnd3* is not activated by SRF in the lipid A response, probably because the motif distance is too small for ternary complex formation. Structural evidence also suggests that ternary complex formation requires more than 1-bp spacing (Hassler, 2001). Furthermore, EMSA experiments have shown that SRF and TCF failed to form a ternary complex on a DNA probe with only 1-bp spacing between the SRF and TCF motifs (Treisman et al., 1992).

When we used the above definition of the SRF-TCF cassette to examine the other 39 promoters combining strong SRF binding with strong motifs, we found only 2 of the 39 promoters contain similar SRF-TCF cassettes: *Fosb* and *Glipr1*. Interestingly, both genes showed rapid transcriptional kinetics like SRF targets. Both genes peaked by 20 min of lipid A stimulation and were repressed by MAPK inhibitors (Figure 3-4B and C). Although *Fosb* is induced more than ten fold, its maximal RPKM is below our threshold of 3. Although *Glipr1* exceeds our maximal RPKM threshold, its fold change is 5.9, which is lower than our fold change threshold of 10. Because *Fosb* is a close related family member of another SRF target, *Fos*, it is likely that these two genes are regulated by similar mechanisms. Although *Fosb* and *Glipr1* missed our stringent criteria for strongly induced genes, they have substantial transcriptional induction and expression levels, so they appear to be functional SRF targets that were mis-categorized in our previous analysis. Thus, using this defined SRF-TCF cassette, we accidentally found another two SRF targets induced by lipid A. This further indicates that the SRF-TCF cassette may be sufficient to identify SRF targets regulated by promoter binding.

Only Nine Additional Promoters Contain SRF-TCF Cassettes

Because we defined the promoter SRF-TCF cassette solely from six SRF targets, it is possible that we might experience a survivor bias; SRF-TCF cassettes might also exist in other promoters but is unnecessary or insufficient for gene induction. To address this problem, we scanned 21013 promoters in the mouse genome for additional SRF-TCF cassettes. In addition to the above eight SRF targets, the defined SRF-TCF cassette only exists in another nine promoters. All of these nine promoters lack SRF binding and

were not activated by lipid A (Figure 3-5A). It is likely that their SRF-TCF cassettes lack transcription factor binding because of nucleosome occupation.

Because promoter CpG content often dictates nucleosome stability, we first assessed their chromatin states by comparing the promoter CpG content between SRF targets and these nine promoters lacking SRF binding. While seven of eight SRF targets have CpG ratio greater than 0.9, none of the nine promoters without SRF binding reach such high CpG ratio (Figure 3-5B). The lower CpG content at the nine promoters lacking SRF binding supports that their promoters tend to possess well-positioned nucleosomes and are inaccessible for SRF binding.

Consistent with the above CpG content data, ATAC-seq data also showed that all SRF targets have accessible promoters. In contrast, only three of the nine promoters without SRF binding are accessible with strong ATAC-seq signals (Figure 3-5C). H3K4me3 mark for active promoters also aligned with the CpG content and ATAC-seq data. While all SRF targets have strong H3K4me3 signal, only three of the nine promoters lacking SRF binding (*Rhoj*, *Tnfsf14*, and *Gpr183*) are active promoters with strong H3K4me3 signals. However, SRF motifs of these three promoters are located beyond the functional range of SRF motifs in SRF targets (Figure 3-5D). While promoter SRF motifs of SRF targets are between 42bp to 410bp upstream of the TSS, one promoter SRF motif without SRF binding (*Tnfsf14*) is located farther upstream of the TSS, which is far from the accessible promoter region. Another two promoter SRF sites without SRF binding (*Rhoj* and *Gpr183*) are at the proximal downstream of the TSS. These two SRF

sites may be occupied by the +1 nucleosome, a well-positioned nucleosome downstream of the TSS (Jiang and Pugh, 2009; Bai and Morozov, 2010). Therefore, the additional nine SRF-TCF cassettes probably lack SRF binding because they are located within inaccessible promoter regions.

Because eight of the nine genes with promoter SRF-TCF cassettes lacking SRF binding have undetectable expression in resting macrophages (data not shown) and appear to require chromatin regulation, their transcriptional activation may require other stimulus-specific or cell-specific transcription factors that can trigger chromatin remodeling. One of these nine genes, *Vil1*, is such an example. *Vil1* is selectively expressed in epithelial cells. Bacterial infection can activate *Vil1* by inducing SRF-TCF pathway in epithelial cells (Rieder et al., 2005).

Therefore, by evaluating the prevalence of the defined SRF-TCF cassette at all promoters, we confirmed that the defined SRF-TCF cassette is a unique feature of SRF targets. The defined SRF-TCF cassette is probably necessary for inducing SRF targets through the MAPK pathway. It also suggests that an accessible chromatin context is essential for ternary complex formation and function at the defined SRF-TCF cassette.

SRF Targets Possess the Strongest SRF Promoter Peaks

Further examination of SRF targets revealed that their promoter SRF-TCF cassettes also coincide with exceptionally strong SRF peaks. Among the strongest fifteen promoter peaks, eight of them contain strong SRF motifs. And six of these eight SRF

peaks with strong SRF motifs belong to SRF targets, often dwarfing most of the peaks on the same chromosome (Figure 3-6A and B). For example, the *Egr1* promoter peak is the strongest peak on chromosome 18 (Figure 3-6B). One likely explanation for this observation is the cooperative binding of TCF, which can stabilize SRF binding to DNA by forming a ternary complex (Shore and Sharrocks, 1994; Shore et al., 1996). In mouse embryonic fibroblasts, mutating TCF transcription factors strongly impaired SRF binding (Gualdrini et al., 2016).

However, another two SRF targets (*Zfp36* and *Fos*) containing promoter SRF-TCF cassettes do not exhibit exceptionally strong SRF binding. This suggests that the exceptionally strong binding at six SRF targets might be supported by other mechanisms as well. One likely mechanism is the collaboration of SRF with other transcription factors. To uncover potential collaborating transcription factors, we compared motif enrichment of promoters with stronger (peak score >30) and weaker (peak score <30) SRF binding. While all promoter peaks are highly enriched with the SRF motif, the stronger SRF peaks are also enriched with motifs of CREB/ATF family transcription factors (Figure 3-6C). To determine whether CREB can promote SRF binding, we evaluated the CREB motif strength and CREB-SRF motif spacing in all promoters with strong SRF binding and strong SRF motifs (Figure 3-6D and E). Interestingly, all six SRF targets with exceptionally strong binding contain TCF motifs and strong CREB motifs (motif score >0.87) within 70bp of SRF motifs. Moreover, all these six SRF target promoters have strong CREB ChIP-seq peaks (Figure 3-6E and F). Another two promoters (*Egr4* and *Filip1l*) also have strong CREB motifs near SRF

motifs. However, both promoters lack strong CREB binding. This is probably because their CREB sites are inaccessible by nucleosome occlusion. Although we cannot exclude the possibility that CREB binding at these six targets may be a coincidence, it is very likely that the proximal CREB binding may enhance SRF binding to six SRF targets through direct or indirect mechanisms. For example, CREB might physically interact and stabilize the SRF-TCF complex. Or it might facilitate SRF binding by bending DNA or excluding nucleosomes to increase chromatin accessibility.

Besides CREB, there may be other explanations for the exceptionally strong SRF binding at six SRF targets. Another two genes, *Srf* and *Bcl2l11*, also have exceptionally strong SRF binding and strong SRF motifs, but lack strong CREB motifs near SRF motifs (Figure 3-7A). However, both of their promoters contain more than one SRF site, which can potentially enhance the binding affinity of SRF (Figure 3-7A). To evaluate the importance of motif copy number, we scanned for additional SRF motifs at the promoters of SRF targets. Four of the SRF targets with the strongest peaks (*Egr1*, *Egr2*, *Dusp5*, and *Glipr1*) have more than one SRF site at their promoters. In *Egr1* promoter, there are six SRF sites and multiple Ets sites spanning two SRF peaks (Figure 3-2B, Figure 3-7A and B). Interestingly, *Egr2* shares a similar pattern with *Egr1*. *Egr2* also has two promoter peaks covering multiple SRF and Ets sites. It is very likely that *Egr1* and *Egr2* use similar mechanisms to regulate transcription by SRF and TCF. And it is also interesting to understand whether the two adjacent promoter peaks function cooperatively or redundantly. Taken together, these findings suggest that the exceptionally strong binding of SRF at SRF targets might depend on more than one

mechanism. Besides the role of TCF in stabilizing SRF binding, the binding of CREB near SRF and multiple copies of SRF motifs might also contribute to the exceptionally strong binding of SRF.

To determine whether the above mechanisms for promoter peaks also apply for enhancer peaks, we examined the SRF motif strength and peak strength of the 552 SRF enhancer peaks (Figure 3-8A). Because it is difficult to correlate enhancer peaks with their targets, and it is hard to separate functional enhancers by gene transcription, we did not observe any clear correlation between the motif strength and peak strength. To assess whether the strong SRF enhancer peaks correlates with Ets motifs or CREB motifs, we searched for the nearest Ets motifs and the best CREB motifs near SRF motifs. Because the motif requirements for functional binding at enhancers may differ from those at promoters, we used less stringent criteria for enhancer SRF-TCF motifs than the defined promoter SRF-TCF cassette. We first found the best SRF motif; then we searched for the nearest Ets motif and the best CREB motif near an SRF motif.

The strongest enhancer peaks exhibit a different motif pattern from the strongest promoter peaks. While six of eight strongest promoter peaks contain strong SRF, Ets, and CREB motifs, only two of sixteen strongest enhancer peaks contain all three motifs. Notably, these two enhancer peaks are near *Btg2* and *ler2*—two SRF targets identified by expression data (Figure 3-2A and Figure 3-8B). Another potent enhancer peak, containing only SRF and Ets motifs, is around 3kb upstream of *Egr3*, a gene that is closely related to another two SRF targets, *Egr1* and *Egr2*. Although *Egr3* has relatively

low expression level (RPKM<1), which is below our threshold for expressed genes (RPKM>3), it is rapidly and transiently induced like other SRF targets (Figure 3-8C). And its induction was sensitive to MAPK inhibitors. Another two enhancer peaks contain both SRF motifs and CREB motifs but lack Ets motifs near SRF motifs. Consequently, their closest genes, *Junb* and *Ubald1*, are not strongly induced by lipid A. Although the motif sequences at enhancer peaks are slightly different from the defined promoter SRF-TCF cassettes of SRF targets, we found that SRF-TCF motifs combining strong SRF binding are sufficient to identify functional enhancer peaks. Thus, by studying the mechanisms of SRF regulation at promoters, we were able to extend similar analyses to enhancers and found that ternary complex formation can regulate transcription at both promoters and enhancers.

MRTF May Regulate Other Promoter SRF Peaks with Strong Motifs

Besides the six SRF targets, another two genes also contain exceptionally strong promoter SRF peaks and strong SRF motifs: *Srf* and *Bcl2l12* (Figure 3-6A). Because these two genes lack the defined promoter SRF-TCF cassette and are not activated by the MAPK pathway, they might be regulated by another SRF cofactor, MRTF. Upon activation of Rho-actin pathway, MRTF can translocate into nucleus and bind to SRF to activate SRF targets. Published data in fibroblasts showed that Cytochalasin D (CytoD), a strong activator of Rho-actin pathway, induced *Srf* and *Bcl2l12* by more than five fold (He et al., 2011, Figure 3-9A and B). Moreover, serum stimulation, a mild activator of MRTF, led to MRTF binding at the promoter of *Srf*, coinciding with SRF binding (Figure 3-9C). These findings suggest that MRTF might regulate other promoter SRF peaks that

lack the defined SRF-TCF cassettes.

To evaluate whether the other 37 promoters combining strong SRF peaks and strong SRF motifs are regulated by the SRF-MRTF pathway, we compared gene induction by CytoD in fibroblasts and by lipid A in macrophages (Figure 3-9D and E). While most SRF targets were preferentially induced by lipid A, 12 of 37 genes with strong SRF motif and binding were preferentially induced by CytoD (fold change>5, max RPKM>3). Moreover, seven of these CytoD-inducible genes also have MRTF peaks at their promoters following serum stimulation (Figure 3-9D). The other five genes lack MRTF binding, probably because serum stimulation is a weaker activator of MRTF than CytoD. Another 25 genes were not strongly induced by CytoD. Their activation might require stronger stimulation or other factors. Taken together, these data support that MRTF might regulate a subset of promoter SRF peaks with strong motifs that lack SRF-TCF cassettes. Because MRTF binds to SRF only after stimulation and it might not stabilize SRF binding like TCF, the SRF peaks regulated by MRTF are weaker than those with SRF-TCF cassettes.

Ternary Complex Formation Can Enhance SRF Binding and H3K4me3 Mark

Because TCF binding coincides with the strongest SRF peaks at promoters and enhancers, we first evaluated its contribution to SRF binding and transcription by mutating Ets sites in immortalized macrophages using CRISPR. To avoid the effects of redundant sites, we mutated the SRF or Ets site at the *Nr4a1* promoter, which has only one SRF site and one Ets site. To get maximal gene induction, we stimulated the

immortalized macrophages with lipid A and serum simultaneously. For each CRISPR mutation, we selected two mutant clones as biological replicates. Mutating the Ets site strongly diminished SRF binding at the *Nr4a1* promoter (Figure 3-10A). This supports the hypothesis that TCF binds cooperatively with SRF and enhances SRF binding to DNA. However, while mutating the SRF site decreased SRF binding to less than 5% of maximal SRF binding, mutating the TCF site still retained about 25% of SRF binding. The residual SRF binding in the absence of TCF indicates that SRF can bind autonomously. In contrast, TCF binding at the *Nr4a1* promoter seems to be completely dependent on SRF. The binding of a TCF family member SAP1 is decreased to similar level by mutating either the SRF or Ets site (Figure 3-10B).

Most SRF-regulated promoters possess CpG islands and strong H3K4me3 marking (Deaton and Bird, 2011; Vavouri and Lehner, 2012). It's still obscure how H3K4me3 mark is deposited. To determine whether SRF binding contributes to H3K4me3 deposition, we measured H3K4me3 levels in CRISPR mutants (Figure 3-10C). Interestingly, mutating either the SRF or Ets site can reduce half of *Nr4a1* promoter H3K4me3 signal. This indicates that SRF binding alone is insufficient for optimal H3K4me3 marking; the formation of ternary complex at basal level is required for optimal H3K4me3 marking. Studies have shown that histone modification is associated with the active ternary complex, possibly by directly or indirectly recruiting histone modifiers (Esnault et al., 2017). And the other half of H3K4me3 signal might depend on other mechanisms. For example, CpG-binding proteins can bind to CpG islands and recruit methyltransferase to promote H3K4me3 marking (Thomson et al., 2010).

Consistent with the ChIP data, mutating either the SRF or Ets site significantly impaired *Nr4a1* gene induction (Figure 3-10D). This indicates that the potent gene induction requires forming an active ternary complex. Although loss of ternary complex reduced more than 70% of maximal transcription, there was still more than 20-fold gene induction. This suggests that other transcription factors may also contribute to *Nr4a1* transcription. For example, MAPK-activated CREB may independently induce the transcription of *Nr4a1*. Thus, by investigating the SRF target, *Nr4a1*, we demonstrated that SRF could bind and recruit TCF to optimize H3K4me3 level and gene induction.

Ternary Complex at the *Egr3* Enhancer Stabilizes SRF Binding and Promotes Transcription

To assess whether TCF-SRF interaction also regulates enhancer function, we mutated the Ets site at the *Egr3* enhancer, which has the strongest SRF peak on chromosome 14 (Figure 3-11A, Figure 3-6B). Interestingly, the *Egr3* enhancer has the same SRF site and TCF site as the *Nr4a1* promoter. But they have different motif spacing. Because *Egr3* has a second SRF site at the promoter peak, which might be redundant or dependent on the enhancer SRF site, we individually mutated the promoter and enhancer sites in immortalized macrophages using CRISPR.

Consistent with our hypothesis that TCF can stabilize SRF binding, mutating the enhancer Ets site compromised the binding of SRF to the *Egr3* enhancer. Mutating the enhancer Ets site has similar effects as mutating the enhancer SRF site (Figure 3-11B). Thus, like *Nr4a1* promoter, the *Egr3* enhancer might also possess a ternary complex formed by the cooperative binding of SRF and TCF. However, there is still about one

third of residual SRF binding at the *Egr3* enhancer in the absence of the enhancer SRF site. Two likely explanations may account for the residual SRF binding. First, there may be another cryptic SRF site at the *Egr3* enhancer. Although we found another weaker SRF motif, its mutation did not affect SRF binding or *Egr3* transcription (data not shown). It is possible that these two enhancer SRF sites are redundant to each other, but the stronger one is more critical. Second, promoter SRF binding might contribute to the residual enhancer SRF binding through promoter-enhancer interaction.

Likewise, mutating the promoter SRF site did not affect SRF binding at the enhancer (Figure 3-11C). This indicates that the enhancer SRF binding is independent of the promoter SRF site. Conversely, mutating the enhancer SRF site or Ets site did not affect promoter SRF binding; only mutating the promoter SRF site impaired promoter SRF binding. Thus, SRF binding at the promoter and enhancer are independent of each other. Probably because there is only one SRF site at the promoter, loss of SRF binding was greater at the promoter than that at the enhancer. It is also possible that the promoter SRF site, rather than the enhancer SRF site, is essential for promoter-enhancer looping. While mutating the enhancer SRF site still allows for promoter-enhancer interaction and the detection of SRF binding at the enhancer, mutating the promoter SRF might abolish promoter-enhancer interaction and therefore prevent detecting SRF binding at the promoter.

SAP1 binding was similarly curtailed by mutating either the SRF or Ets site at the *Egr3* enhancer (Figure 3-11D). Therefore, like the *Nr4a1* promoter, TCF binding at the *Egr3*

enhancer is also dependent on SRF. Furthermore, loss of the promoter SRF site did not affect TCF binding at the enhancer. This again confirms that ternary complex formation at the enhancer is independent of the promoter SRF site.

Mutating either the enhancer SRF site, TCF site, or promoter SRF site reduced more than 70% of the maximal *Egr3* transcription (Figure 3-11E). The reduced gene induction suggests that ternary complex formation at the enhancer is critical for *Egr3* gene induction. Because the enhancer and promoter mutation exhibit similar loss of gene induction, both the enhancer and promoter may be required for *Egr3* induction, possibly by promoter-enhancer looping. However, none of these mutants reduced the gene expression to basal level. The residual 30% of transcription might be explained by the activity of the promoter or enhancer alone. It is also likely that other lipid A-activated transcription factors can contribute to *Egr3* transcription. For example, we found that CREB binds to a strong CREB motif on the *Egr3* promoter in CREB ChIP-seq data (data not shown). This suggests that SRF and CREB may act independently to promote the transcription of *Egr3*.

We did not detect any change of histone marks for active enhancer (H3K27ac) or active promoter (H3K4me3) in any of these mutants (data not shown). This might result from the redundant roles of promoter and enhancer. This might also imply that histone mark deposition is dependent on gene-specific mechanisms at *Egr3*.

Promoter SRF-TCF Cassette Is Correlated with Robust and Ubiquitous Transcriptional Induction

To understand the cell-specific binding patterns of SRF, we compared the SRF ChIP-seq data from five different cell types: BMDMs, 3T3 fibroblasts, cortical neurons, HL-1 cardiac muscle cells, and C2C12 myocytes (Esnault et al., 2014; Kim et al., 2010; He et al., 2011). Because we have more confidence in identifying targets regulated by promoter peaks than enhancer peaks, we first compared SRF promoter peaks in different cell types. When comparing all promoter peaks, we found up to 78 unique promoter peaks in each cell type, but not all of them may be regulated by direct or functional binding (Figure 3-12A, left). To enrich the functional sites, we further compared the strong peaks combining strong motifs. This revealed only fewer than 5 unique promoter peaks in each cell type. These promoter peaks are generally weaker peaks and lack CpG islands at promoters (Figure 3-12A, right). And many of these promoters regulate cell-specific genes. For example, one neuron-specific SRF promoter peak is associated with *Pou3f4*, a gene encoding a neuron-specific transcription factor. It is likely that this small subset of cell-specific promoter SRF peaks tends to form stable nucleosomes and requires chromatin remodeling initiated by cell-specific transcription factors.

All these five cell types share 41 SRF promoter peaks containing strong SRF motifs. These 41 promoters include seven SRF targets bearing promoter SRF-TCF cassettes. Only one SRF target, *Glipr1*, is not among these 41 promoters. *Glipr1* lacks a CpG-island promoter. It's likely that the SRF binding to *Glipr1* promoter requires chromatin

remodeling maintained by another macrophage-specific transcription factor. Because SRF predominantly binds to CpG-island promoters, which are constitutively accessible without the requirement of chromatin remodeling, SRF can ubiquitously bind to these promoters in many cell types.

Next, we compared published transcriptional data using different activators of SRF in three cell types: lipid A stimulation in BMDMs; serum stimulation in 3T3 fibroblasts; and KCl depolarization in neurons (Tong et al., 2016; Esnault et al., 2014; Kim et al., 2010). All these stimuli strongly induced the MAPK-TCF pathway. Serum stimulation also mildly activates Rho-actin pathway and MRTF. When we compared the 41 genes with strong promoter SRF binding and strong SRF motifs, we found only six genes induced more than five fold in all cell types (Figure 3-12C). All of them are the SRF targets identified in the lipid A response (Figure 3-4A). This suggests that the defined promoter SRF-TCF cassette is functional in different cell types and may be essential for the ubiquitous gene induction. Therefore, by embedding the conserved TCF-SRF cassettes in accessible CpG-island promoters, the SRF-TCF ternary complex can bind ubiquitously in different cell contexts and promote robust and transient transcriptional activation by different stimuli.

Discussion

By carefully documenting the motif characteristics of SRF peaks associated with SRF targets, we discovered that the MAPK-TCF pathway selectively regulates a small subset of SRF targets through a unique promoter SRF-TCF cassette. This strict motif

requirement refined the model of the combinatorial regulation by multiple transcription factors. This study also demonstrated that our gene-centric method has the potential to uncover the mechanistic details of transcriptional regulation.

By comparing the transcriptional induction, motif strength, and binding strength in well-defined gene sets, we were able to separate the direct and functional binding events that contribute to transcription. The correlation between strong binding, strong SRF motifs, and strong conservation at primary response genes suggest that SRF activity requires direct binding to a strong SRF motif. The differential CpG content and motif conservation at promoters and enhancers indicate that SRF binds to promoters and enhancers with different mechanisms to differentially regulate gene expression. Previous studies have found that SRF can regulate ubiquitously expressed genes through promoters, and SRF can regulate expression of cell-specific and stimulus-specific transcription through enhancers (Sullivan et al., 2011). Because many cell-specific and stimulus-specific genes evolved later than genes regulating development and essential cellular functions, these cell-specific and stimulus-specific enhancers might not be as conserved as promoters.

Because MAPK-induced SRF targets are relatively few, and there are many studies on individual SRF targets, we were able to scrutinize each SRF target. By comparing their features of motif, peak strength, and expression, we found unique patterns emerged. We discovered that promoter-regulated SRF targets share a strict SRF-TCF cassette. Although previous studies have examined individual SRF targets and identified

consensus SRF motifs using the site selection method and *de novo* motif discovery, we found that the SRF-TCF targets require even stricter motif sequences (Pollock and Treisman, 1990; Sullivan et al., 2011; Esnault et al., 2014;). The strict rules explain why the defined SRF-TCF targets are highly selective and rare in the mouse genome. Besides the eight SRF targets, the characterized promoter SRF-TCF cassette only exists in 9 of 21013 promoters that lack SRF binding. Furthermore, the SRF-TCF cassettes of SRF targets are more conserved than those without SRF binding. While seven of the eight promoter SRF motifs of SRF targets are conserved, only five of the nine promoter SRF motifs lacking SRF binding are conserved. We also found these nine SRF motifs lack CpG islands at promoters and probably reside in inaccessible chromatin. This suggests that these SRF-TCF cassettes lacking SRF binding might evolve later and require recruiting chromatin-remodeling complexes by cell-specific or stimulus-induced transcription factors.

Besides the eight SRF targets, we found another 37 genes containing strong promoter SRF binding and strong SRF motifs. At least one third of them are regulated by another cofactor of SRF, MRTF. Interestingly, some genes can be activated by either TCF or MRTF. Because TCF and MRTF can competitively bind to a common interface at SRF, these genes might be controlled by the integrated signals from both pathways (Murai and Treisman, 2002; Zaromytidou et al., 2006; Gualdrini et al., 2016). In contrast, some genes can be activated by only one pathway. These genes might employ a selective mechanism to prevent the interference of another pathway. While SRF targets are sensitive to any MAPK-TCF activator, many MRTF targets do not respond to all types of

MRTF stimuli. For example, a weak activator of MRTF, serum stimulation, failed to activate twelve genes that were inducible by cytochalasin D in the same cell type. The differential responses to MRTF activators indicate that different stimuli can selectively regulate MRTF through yet unknown mechanisms. Besides MRTF-regulated SRF targets, strong SRF promoter binding and strong motifs also exist at the promoters of another 25 genes. There is no evidence supporting their regulation by the SRF-TCF or SRF-MRTF pathways. Therefore, their activation might require selective post-translational modifications, other cofactors of SRF, or collaboration with other transcription factors (Posem and Treisman, 2006; Sealy et al., 1997; Watson et al., 1997).

We also found that SRF targets are unique in their exceptionally strong SRF binding. Several mechanisms can possibly explain this phenomenon. The most likely explanation is the supportive role of TCF, which can bind cooperatively with SRF in a ternary complex and stabilize SRF binding. By CRISPR mutation experiments, we found that TCF can enhance SRF binding not only at promoters but also at enhancers. Three strongest enhancer peaks with SRF and TCF motifs are all in the vicinity of previously identified SRF targets. Therefore, these SRF-TCF motifs at enhancers might also possess a ternary complex that regulates transcriptional induction by lipid A.

Besides the supportive role of TCF, the proximate CREB binding might also contribute to the strong binding of SRF. It is interesting to clarify whether CREB can promote SRF binding through direct or indirect mechanisms. It is possible that CREB binding can

bend the DNA or exclude nucleosomes to increase accessibility for ternary complex formation. Alternatively, CREB can facilitate SRF binding and transcription by participating in a large complex. It has been shown that the active SRF-TCF complex can recruit CREB binding protein (CBP), which also interacts with CREB at *c-fos* promoter (Nissen et al., 2001; Ramirez et al., 1997). Because CREB can independently recruit CBP, CREB can probably promote the interaction between CBP and ternary complex and assist with basal transcriptional machinery assembly. CREB is also known to regulate several SRF targets (Ahn et al., 1998; Herndon et al., 2013; Vialou et al., 2012; Ramanan et al., 2005). Because most SRF targets are involved in key cellular functions including cell proliferation, cell cycle regulation, apoptosis, cell growth, and metabolism, they might require stringent regulation by more than one mechanism. In addition to SRF-TCF regulation, CREB can add another layer of regulation to SRF targets, refining the expression pattern for stimulus-specific responses. And by integrating different intracellular and extracellular signals, the combination of SRF and CREB regulation may sensitize the cellular response to environmental changes.

By mutating the SRF and Ets sites at the promoter of *Nr4a1*, we confirmed that TCF supports the strong binding of SRF. While SRF can bind by itself, TCF alone cannot bind to DNA. This suggests that SRF might be able to recruit TCF on the *Nr4a1* promoter. Although many studies showed that SRF could recruit TCF, there is also evidence indicating that TCF can recruit SRF to form a ternary complex (Treisman et al., 1992; Latinkic et al., 1996). It is likely that the subtle sequence differences in TCF members can affect the interaction with SRF and DNA. DNA context can also affect

SRF and TCF interaction. For example, both SAP1 and Elk-1 contain C-terminal inhibitory sequences that prevent the autonomous binding to DNA (Treisman et al., 1992). By mutating SRF and Ets sites, we found that ternary complex is also required for the optimal H3K4me3 marking at the *Nr4a1* promoter. The underlying mechanism is still unclear. It is likely that the SRF-TCF ternary complex can recruit histone modifiers (Esnault et al., 2017). It is also likely that the loss of ternary complex binding can affect the nucleosome structure and histone modifications.

By CRISPR mutation experiments, we confirmed that TCF can facilitate SRF binding at the *Egr3* enhancer. Moreover, *Egr3* induction depends on both the promoter and enhancer SRF sites. Similar to *Egr3*, another two Egr family members, *Egr1* and *Egr2*, also possess two SRF peaks upstream of the TSS. It will be interesting to investigate whether these two peaks are redundant or interdependent to each other. Because *Egr1*, *Egr2*, and *Egr3* belong to the same family and share similar patterns of SRF binding, their regulation by SRF might depend on similar mechanisms. However, comparing to *Egr1* and *Egr2*, *Egr3* is distinct in at least three aspects. First, while *Egr1* and *Egr2* have two SRF promoter peaks, *Egr3* has one promoter peak and one enhancer peak. Second, *Egr3* has fewer copies of SRF motifs than *Egr1* and *Egr2*. Third, although all three genes exhibit similar transcriptional kinetics after stimulation, *Egr3* has a much lower expression level than *Egr1* and *Egr2*. The first two differences can possibly explain the lower expression level of *Egr3*. Because both the promoter and enhancer are required for the optimal induction and the *Egr3* enhancer peak is far from the TSS, *Egr3* expression may be limited by enhancer-promoter interaction. This can result in

less potent transcription than *Egr1* and *Egr2*, which both depend on promoter regulation. It is also likely that fewer copies of SRF motifs can lead to lower SRF binding affinity to *Egr3*. But it is also possible that the additional SRF motifs of *Egr1* and *Egr2* can enhance transcription by other mechanisms. Because Egr members regulate genes of different biological functions, SRF might control the differential expression of different family members by altering their motif copies and locations (Poirier et al., 2008; O'Donovan et al., 1999).

Interestingly, lipid A-induced SRF targets are the only genes that are inducible by different MAPK activators and in different cell types. The robust and ubiquitous transcriptional induction of SRF targets probably derives from the combination of CpG-island promoters and the unique SRF-TCF cassettes. Because CpG-island promoters form unstable nucleosomes and are constitutively accessible in all cell types, they allow the formation of ternary complex in resting cells, which can rapidly respond to various environmental stimuli through MAPK pathway activation.

Thus, by revisiting a classical model of combinatorial regulation using a gene-centric method, we revealed strict motif rules that selectively regulate the SRF-TCF targets. Because SRF lacks have any functionally redundant family member and interacts with only one cofactor at a time, it represents a relatively simple model of combinatorial regulation. But the strategy in our study has demonstrated the potential to answer more complicated questions. In future studies, we can use this strategy to investigate how different transcription factors act in concert to induce gene transcription in more

complicated combinatorial regulation models involving more factors. We can also use this strategy to explore and compare the regulatory mechanisms of different transcription factor family members. And this will deepen our understanding of the mechanisms and underlying biological significance that account for the selective transcriptional activation by different environmental stimuli.

Materials and Methods

Cell Culture and Reagents

Bone marrow cells were isolated from 6- to 10-week-old C57BL/6 male mice. They were incubated in medium containing M-CSF for six days for macrophage differentiation. Bone marrow-derived macrophages were stimulated with 100ng/ml lipid A (Sigma L6895) for 0, 15, 30, 60, and 120min before cross-linking for ChIP experiments. J2 virus-immortalized macrophages were also from C57BL/6 mouse. Immortalized macrophages were incubated in medium containing 0.5% fetal bovine serum overnight. The next day, they were stimulated by 100ng/ml lipid A in media with 20% fetal bovine serum.

Chromatin Immunoprecipitation and ChIP-seq Library Preparation

After stimulation, 40 million bone marrow-derived macrophages were cross-linked with 1% formaldehyde (Fisher, PI-28908). Nuclei pellets were isolated and sonicated using Misonix 3000 sonicator to fragments between 200bp to 1000bp. Chromatin lysate was incubated with SRF antibody (Santa Cruz, sc-335), SAP1 antibody (Santa Cruz, sc-13030), or H3K4me3 antibody (Millipore, 07-473) overnight. The immune complex was

pulled down by Protein G Dynabeads (Invitrogen, 10004D) and washed for four times. The purified immune complex was incubated with proteinase K (ThermoScientific, EO0491) at 60°C overnight for reverse cross-linking and protein digestion. IP DNA was purified by phenol-chloroform (Sigma, P3803). DNA concentration was quantified by Qubit kit (Thermo Fisher, Q32854). ChIP-seq library was prepared using KAPA LTP library preparation kit (KK8500) following the manufacturer's instructions.

ChIP-seq Read Mapping and Processing

ChIP-seq library samples were sequenced on Illumina HiSeq 2000 platform. Single-end 50bp reads were aligned to mouse mm9 genome using Bowtie2. Unique mapped reads were kept for peak calling using HOMER (Heinz et al., 2010) with enrichment over input and FDR less than 0.01. Non-specific peaks were eliminated by comparing with SRF ChIP-seq data from SRF KO macrophages. Peak read density was calculated by HOMER peak annotation function and heat maps were generated using Java TreeView software. SRF peaks were annotated as promoter peaks if they are within -600bp to +250bp relative to TSS. CpG content is calculated by dividing the number of observed CpG by the number of expected CpG. Promoter region is defined as -500bp to +150bp relative to TSS.

Chromatin RNA Extraction, RNA-seq Library Preparation and Data Analysis

See Chapter 2 EXPERIMENTAL PROCEDURES.

Motif Analysis and Conservation Analysis

SRF, TCF, and CREB motifs were searched using TRANSFAC and JASPAR transcription factor data base by Pscan (Wingender et al., 1996; Mathelier et al., 2016; Zambelli et al., 2009). Promoter region is defined as -500bp to +150bp relative to TSS. The best SRF motif identified by motif M00186 was used for analysis of each SRF peak. Ets motifs (M00032, M00074, M00007, M00025, MA0080.3, MA0081.1, MA0098.2, MA0028.1, and MA0076.2) were used to search the nearest Ets site near SRF motif. CREB motif M00039 was used to find the best CREB motif in promoter or enhancer peaks.

Conservation score is quantified using UCSC PhastCons placental mammal data (Siepel et al., 2005). Conservation score is quantified as the average PhastCons score over a 10bp SRF motif. The best SRF motif of each region was used for conservation analysis. Enhancer SRF motifs in the mouse genome were identified using HOMER findMotifsGenome.pl function. Strong motifs were identified with SRF motif score (M00186) greater than 0.89.

CRISPR

J2 virus-immortalized macrophages were diluted and seeded in 96-well plates to obtain single cell colonies. Single cell-derived colony was expanded for CRISPR experiment. Single guide CRISPR sequences targeting transcription factor binding sites were designed using MIT CRISPR Designer (<http://crispr.mit.edu/>). They were cloned into lentiviral vector lentiCRISPRv2 (Addgene, 52961), which expresses both Cas9 and guide RNA. Lentiviral vectors and lentiviral packing plasmids psPAX2 (Addgene, 12260)

and pMD2.G (12259) were transfected into 293T cells. Lentiviral media were collected 36hr and 60hr after transfection. One million single cell-derived macrophages were plated in 6-well plates. Each well was incubated with 4ml of lentiviral medium and centrifuged at 2000rpm for 1.5 hours. Two spin infections were performed. Three days after the second spin infection, macrophages were cultured in medium containing 12.5ug/ml puromycin for seven days. Puromycin selection media were replaced every two days. Puromycin-resistant macrophages were scraped, diluted, and plated into single cells in 96-well plates. Single cell colonies were expanded for 2 weeks. Cells were lysed and protein is digested by proteinase K at 55°C overnight. Genomic DNA is precipitated by isopropanol. DNA pellets were washed with 70% ethanol and dissolved in water. Primers targeting the transcription factor sites were used to select mutant clones. The region flanking the mutation sites of candidate clones were PCR amplified and sent for Sanger sequencing. Two mutants for each CRISPR guide RNA were selected for qRT-PCR and ChIP experiments.

RNA extraction and qRT-PCR

Two million cells were lysed in TRI reagent (Molecular Research Center, TR118). RNA was extracted using Qiagen RNeasy kit following the manufacture's instructions. 1ug RNA was used to synthesize cDNA. Levels of cDNA were quantified by qPCR using BioRad CFX384 Real-Time PCR machine.

Accession Number

RNA-seq, ChIP-seq, and ATAC-seq data were deposited in Gene Expression Omnibus (GEO) database. BMDM chromatin RNA-seq, SRF ChIP-seq and ATAC-seq are under session number GSE67357. H3K27ac ChIP-seq data are under session number GSE38379. 3T3 fibroblast RNA-seq and ChIP-seq data are under session number GSE45888. Neuron SRF ChIP-seq data are under session number GSE21161. C2C12 SRF ChIP-seq data is under session number GSE36024. HL-1 SRF ChIP-seq data is under session number GSE21529.

Figure Legends

Figure 3-1. General Features of SRF ChIP-Seq in Lipid A-Activated Macrophages

Bone marrow-derived macrophages were stimulated with lipid A for 0, 15, 30, 60, and 120 min. Samples were collected for SRF ChIP-seq experiments. (A) The heat map shows the read density of SRF ChIP-seq peaks in a 6-kb window, centered in peaks called in at least one sample (by HOMER, false discovery rate <0.01). Peaks are ranked by the average peak score. The color indicates the read value. (B) The pie chart displays the genomic distribution of 718 reproducible peaks. Promoter region is defined as -500 to +150 bp relative to the TSS. (C) The genomic distribution of SRF ChIP-seq peaks is compared with the distribution of different genomic regions. (D) The CpG content of 166 promoters with SRF binding (in red) is compared with 21168 promoters (in black) in the mouse genome. The y-axis shows CpG content of promoters, which equals to the number of observed CpG divided by the number of expected CpG. The x-axis shows the percent of promoters in each category. The table compares the CpG content between SRF-regulated promoters and all promoters. (E) The CpG content of

552 SRF enhancer peaks (in red) is compared with 26434 enhancers with H3K27ac peaks (in black). The y-axis shows the CpG content, which equals to the number of CpG divided by the number of expected CpG in a 395-bp window. The x-axis shows the percent of enhancers in each category. The table compares the CpG content between SRF-regulated promoters and all promoters.

Figure 3-2. Identify SRF Targets by RNA-seq and ChIP-seq

(A) The heat map shows log₂-transformed expression levels of eight SRF targets. Macrophages were stimulated by lipid A with 5 min intervals in the first hour including a 2 hr stimulation. The blue shade indicates the expression level. The blue and red heat map shows the fold change of each time point relative to the previous time point. The CpG column shaded in brown indicates genes with CpG-island promoters (observed/expected CpG ratio is greater than 0.6). The MAPK column in blue indicates genes inhibited by MAPK inhibitors (ERK inhibitor PD0325901; p38 inhibitor BIRB0796) by at least three fold. (B) The left panel shows the genome browser snapshots of SRF binding at the promoters of six SRF targets: *Egr1*, *Egr2*, *Dusp5*, *Fos*, *Zfp36*, and *Nr4a1*. The right panel shows SRF binding to distal regions of two SRF targets (*Btg2* and *Ier2*), which overlap with H3K27ac enhancers. The red arrow indicates the TSS and the direction of transcription.

Figure 3-3. Characterize SRF Peaks by Peak Strength and Motif Strength

(A) The left scatter plot displays the peak score of promoter peaks (x-axis) and the motif score by TRANSFAC transcription factor database (y-axis) of 132 primary response

genes. The right scatter plot shows the peak score and motif score of other gene classes: secondary response genes, 2- to 10-fold induced genes, not induced genes, and low expression genes. The horizontal dashed line indicates the motif score threshold of 90. The vertical dashed line indicates the peak score threshold of 10. (B) Tables display the number and percent of promoters divided by motif score and peak score in five gene classes. (C) The line graph shows the conservation scores of SRF sites for SRF peaks at primary response genes (left), all promoters (middle), and enhancer sites (right). Conservation score is quantified as the average score spanning the 10-bp SRF motif using UCSC PhastCons data. Promoter SRF sites and enhancer SRF sites with binding are identified by Pscan. Strong SRF motifs at enhancers without binding are predicted by HOMER.

Figure 3-4. SRF Targets Share a Unique Promoter SRF-TCF Cassette

(A) The table lists SRF motif, TCF motif, and motif spacing at the promoters of six SRF targets in primary response genes and another two SRF targets in other gene classes. (B) The scatterplot shows the fold change of nascent transcripts induced by lipid A (x-axis) and the percent of maximal expression in the presence of MAPK inhibitors (ERK inhibitor PD0325901 and p38 inhibitor BIRB0796). The horizontal dashed line indicates the expression threshold of 33%. The vertical dashed line indicates the fold change threshold of 5. SRF targets in primary response genes are highlighted in red. Another two SRF targets, *Glipr1* and *Fosb*, are labeled in the right bottom square. (C) The heat map shows nascent transcript expression of six SRF targets in primary response genes and genes in other classes that also have strong SRF motifs and strong SRF binding at

promoters. Normalized expression by RPKM is shown in blue scale. The fold change relative to the previous time point was shown in blue to red scale. MAPKi column in blue indicates genes sensitive to MAPK inhibition.

Figure 3-5. Identify Promoters Containing Defined SRF-TCF Cassettes and Lacking SRF Binding

(A) The table lists SRF motif, TCF motif, locations of motifs, motif spacing, and CpG content of nine genes in the genome that contain the defined promoter SRF-TCF cassettes. CpG content equals the number of CpG divided by the number of expected CpG in promoter. (B) The dot plot displays the promoter CpG content of eight SRF targets (in red) and nine genes with promoter SRF-TCF cassettes lacking SRF binding (in black). The dotted horizontal line indicates the CpG obs/exp threshold of 0.6. (C) The dot plot compares the levels of H3K4me3 (in green) and ATAC-seq signal (in black) between eight SRF targets (left) and nine genes with promoter SRF-TCF cassettes lacking SRF binding (right). The levels of H3K4me3 ChIP-Seq and ATAC-seq are evaluated by RPKM of the called peaks. The dotted horizontal line indicates the RPKM threshold of 2. (D) The dot plot compares the position of promoter SRF motif relative to the TSS between eight SRF targets (in red) and nine genes with promoter SRF-TCF cassettes lacking SRF binding (in black). The dotted horizontal lines indicate position -410 and +1 relative to the TSS.

Figure 3-6. Interrogate the Potent SRF Binding at SRF Targets

(A) The scatterplot shows SRF peak score (x-axis) and SRF motif score quantified by

TRANSFAC database (y-axis) at all promoters. The grey horizontal dashed line indicates the motif score threshold of 89.5. The vertical dashed line indicates the peak score threshold of 10 (left, used for strong SRF binding) and 67 (right, used for top 15 promoter peaks). Eight SRF targets are highlighted in red. Top 15 peaks are highlighted in blue. (B) Genome browser snapshots display six chromosomes containing SRF peaks of SRF targets and *Egr3*. (C) Motif analysis comparing the strong motifs with peak score greater than 30 and strong motifs with peak score lower than 30. The p value (in red) of the enriched motifs is quantified by Pscan. (D) The scatterplot compares CREB and SRF motif spacing identified by Pscan (x-axis) and motif score (y-axis) at promoters with strong SRF binding and motifs. The grey dashed horizontal line indicates the CREB motif score threshold of 0.87. The grey dashed vertical line indicates the SRF-CREB motif spacing threshold of 70bp. Eight SRF targets are highlighted in red. Two SRF targets (*Fos* and *Zfp36*) outside the top left square are labeled. Another two top promoter peaks (*Srf* and *Bcl2l12*) with strong motifs are also labeled. (E) The table compares the eight promoters at the top left square in (D) by their CREB motif score, SRF-CREB motif spacing, CREB ChIP-seq peak score, and SRF ChIP-seq peak score. (F) Promoter sequences of six SRF targets, which rank among top 15 strongest promoter peaks. Position relative to the TSS is shown on left. SRF sites are in red; TCF sites are in blue; and CREB sites are in green.

Figure 3-7. Strong Promoter SRF Peaks Possess Proximate CREB Binding and Multiple SRF Motifs

(A) The table compares SRF targets and another two genes (*Srf* and *Bcl2l12*) by the

number of SRF site, minimal motif spacing between SRF sites, CREB site near SRF site, TCF site near SRF site, and SRF peak score. (B) Promoter sequences of *Egr1* and *Egr2*, which have multiple SRF sites and TCF sites. Position relative to the TSS is shown on left. SRF sites are in red; TCF sites are in blue; and CREB sites are in green.

Figure 3-8. SRF-TCF Motifs at Enhancer SRF Peaks Correlate with SRF Targets

(A) The scatterplot shows SRF peaks score (x-axis) and SRF motif score quantified by TRANSFAC database (y-axis) at enhancer peaks. Enhancer SRF peaks containing Ets motifs within 20bp of SRF motifs are in blue. Enhancer SRF peaks containing CREB motifs within 70bp of SRF motifs are in yellow. Enhancer SRF peaks containing both Ets motifs within 20bp of SRF motifs and CREB motifs within 70bp of SRF motifs are in red. The horizontal dashed line indicates the motif score threshold of 89.2. The vertical dashed lines indicate the SRF ChIP-seq peak score threshold of 10 (left, for strong peaks) and 88 (right, for top 25 strongest enhancer peaks). The closest genes of 25 strongest enhancer peaks in top right square are labeled: two peaks in red represent *Btg2* and *Ier2*; one peak in blue represents *Egr3*; two peaks in yellow represent *Junb* and *Ubald1*. (B) Enhancer SRF peak sequences of *Btg2*, *Ier2*, and *Egr3* are shown. Position relative to the TSS is shown on left. SRF sites are in red; TCF sites are in blue; and CREB sites are in green. (C) The line graph indicates RNA-seq RPKM values measuring the nascent transcript levels of *Egr3* stimulated by lipid A in the presence of DMSO (in black) control or MAPK inhibitors (in blue).

Figure 3-9. The Role of MRTF in Regulating Promoters with Strong SRF Binding

and Strong Motifs

(A) The bar graphs compare the mRNA expression of *Srf* induced by 100ng/ml lipid A in BMDMs. (B) The bar graph compares mRNA expression of *Bcl2/12* stimulated by 2uM cytochalasin D (CytoD) in 3T3 fibroblasts. (C) Genome browser snapshots show that SRF binding coincides with MRTF binding at the promoter of *Srf* following serum stimulation in 3T3 fibroblasts. (D) The table summarizes genes with strong promoter binding and SRF motifs by their responses to cytochalasin D (CytoD) and MRTF binding in 3T3 fibroblasts. (E) The scatterplot shows the fold change of nascent transcript induced by lipid A in macrophages (x-axis) and the fold change of mRNA induced by cytochalasin D (CytoD) in 3T3 fibroblasts (y-axis) for genes with strong promoter binding and strong SRF motifs. Eight SRF targets are in red. Genes with inducible MRTF binding by serum stimulation are in green. Other genes with strong promoter SRF binding and strong motifs are in grey.

Figure 3-10. Evaluate the Roles of the Ternary Complex at the *Nr4a1* Promoter by CRISPR

Bar graphs show SRF binding (A), SAP1 binding (B) or H3K4me3 (C) at the *Nr4a1* promoter in control, mutant lacking the SRF site (sgSRF), and mutant lacking the TCF site (sgTCF). J2 virus-immortalized bone marrow-derived macrophages are stimulated with 100ng/ml lipid A and 20% serum. The data shown represent an average of three biological replicates. (D) Bar graph shows the fold change of the *Nr4a1* mRNA induced by 100ng/ml lipid A and 20% serum in immortalized macrophages. The data shown represent an average of three biological replicates. Error bars indicate the standard

error. **p < 0.01; *p < 0.05; n.s.=not significant.

Figure 3-11. Evaluate the Roles of the Ternary Complex at the *Egr3* Enhancer by CRISPR

(A) Genome browser snapshots show the enhancer SRF peak and promoter SRF peak of *Egr3*. The enhancer peak contains one TCF site (in blue) and one SRF site (red). The promoter contains one SRF site (in red). Bar graphs show SRF binding at the *Egr3* enhancer (B) or promoter (C) in control, mutant lacking the enhancer SRF site (sgEnh_SRF), mutant lacking the enhancer TCF site (sgEnh_TCF), and mutant lacking the promoter SRF site (sgPro_SRF). J2 virus-immortalized bone marrow-derived macrophages are stimulated with 100ng/ml lipid A and 20% serum. (D) Bar graph shows SAP1 binding at the *Egr3* enhancer in control, mutant lacking the enhancer SRF site (sgEnh_SRF), mutant lacking the enhancer TCF site (sgEnh_TCF), and mutant lacking the promoter SRF site (sgPro_SRF). (E) Bar graph shows the fold change of *Egr3* mRNA induced by 100ng/ml lipid A and 20% serum in immortalized macrophages. The data shown represent an average of three biological replicates. Error bars indicate the standard error. **p < 0.01; *p < 0.05; n.s.=not significant.

Figure 3-12. Robust Induction of SRF Targets by Different Stimuli in Different Cell Types

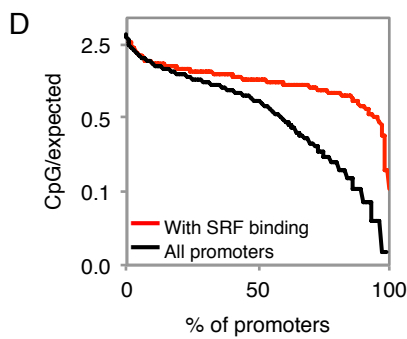
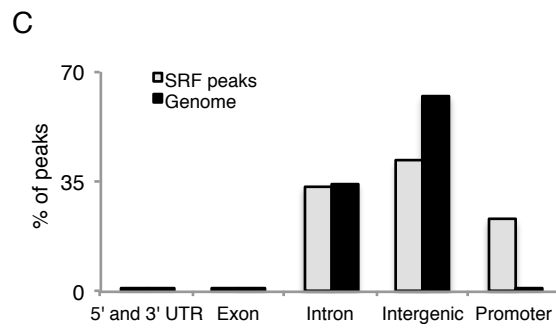
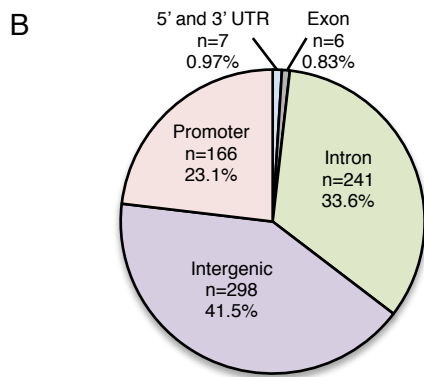
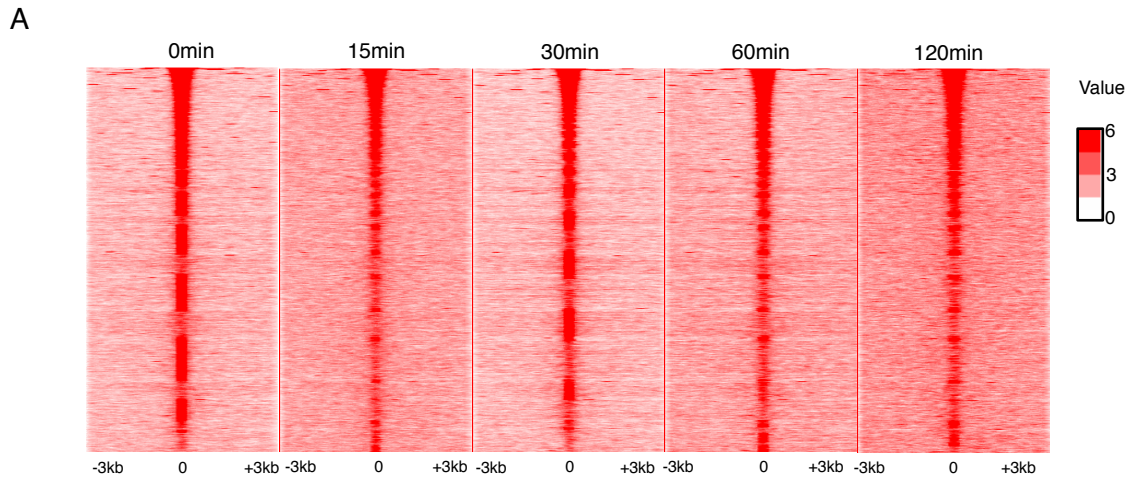
(A) The Venn diagram compares all SRF promoter peaks (left) or SRF promoter peaks with strong motifs (right) overlapping between 3T3 fibroblasts, BMDMs, cortical neurons, HL-1 cardiac muscle cells, and C2C12 myocytes. (B) The table lists cell-

specific promoter peaks with strong motifs by peak score and promoter CpG content. (C) The dot plot compares the induction of 41 genes with strong promoter binding and strong motifs in five cell types. Gene induction is quantified by RNA-seq data in different cell types. Nascent transcript induction by lipid A in BMDMs is in red. Messenger RNA induction by serum stimulation in 3T3 fibroblasts is in green. Messenger RNA induction by KCl depolarization in cortical neurons is in blue. The grey vertical dashed line indicates the fold change threshold of 5.

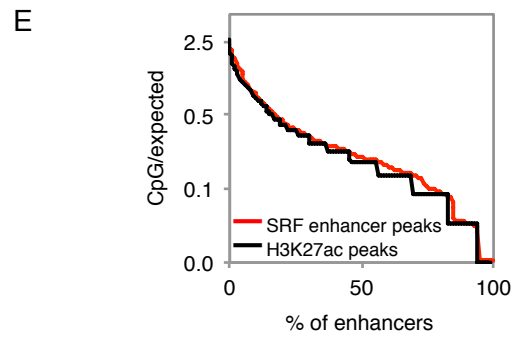
Table 3-1. List of Genes with SRF Binding or SRF Motifs

The list shows the Refseq IDs, gene symbols, aliases, and functions of lipid A-induced SRF targets, genes with promoter SRF-TCF cassettes lacking binding, and other genes mentioned in this article.

Figure 3-1. General Features of SRF ChIP-Seq in Lipid A-Activated Macrophages



	Number of promoters			Percent of promoters		
CpG/expected ratio	>0.6	0.2-0.6	<0.2	>0.6	0.2-0.6	<0.2
With SRF binding	142	10	3	91.6	6.5	1.9
All promoters	11708	4645	4811	55.3	21.9	22.7



	Number of enhancers			Percent of enhancers		
CpG/expected ratio	>0.6	0.2-0.6	<0.2	>0.6	0.2-0.6	<0.2
SRF enhancer peaks	78	199	275	14.1	36.1	49.8
All enhancers	3309	8709	14416	12.5	32.9	54.5

Figure 3-2. Identify SRF Targets by RNA-seq and ChIP-seq

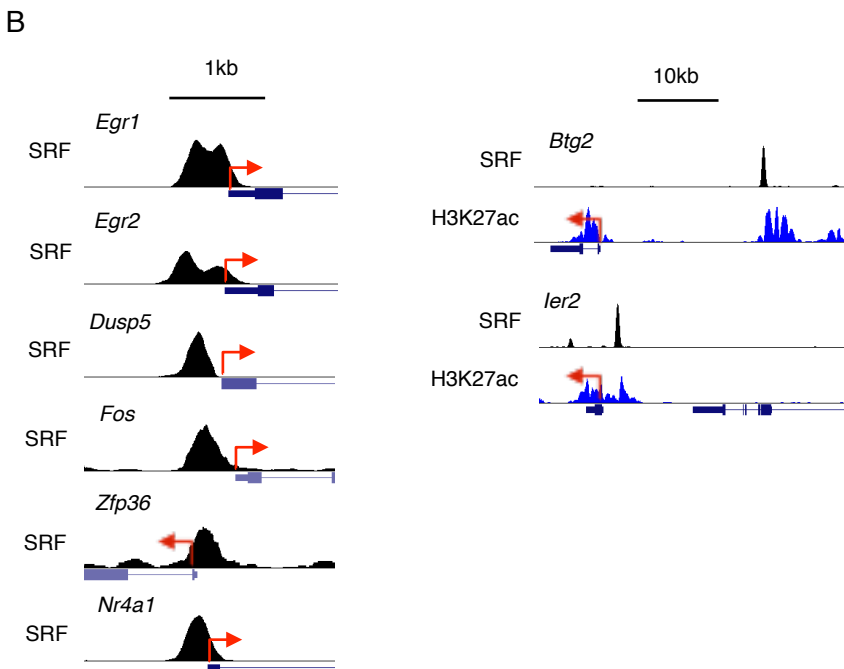
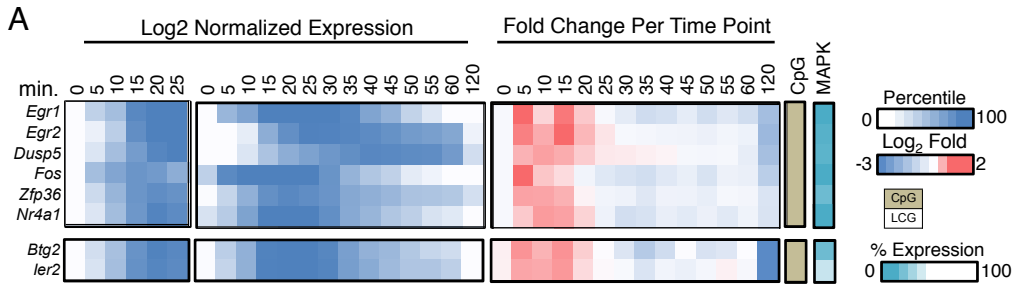


Figure 3-3. Characterize SRF Peaks by Peak Strength and Motif Strength

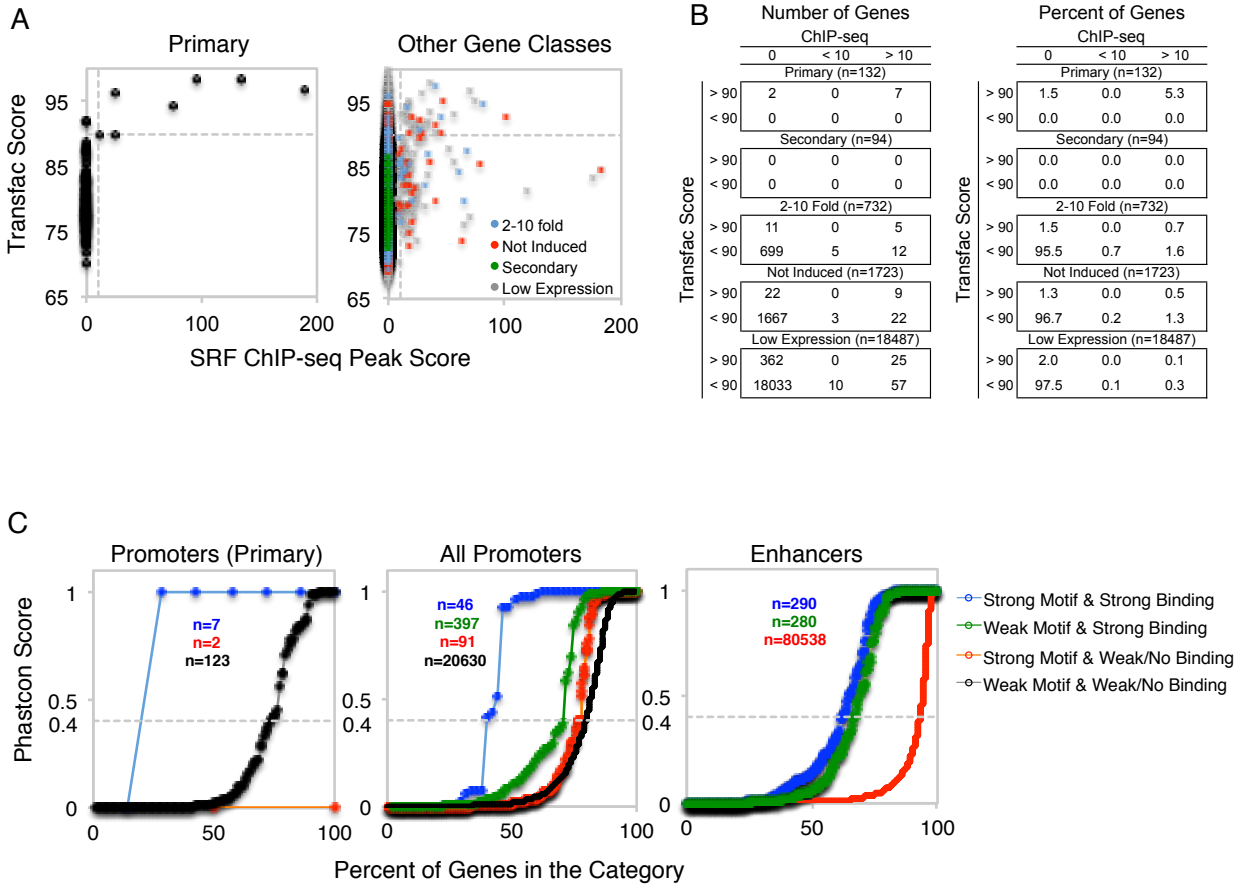
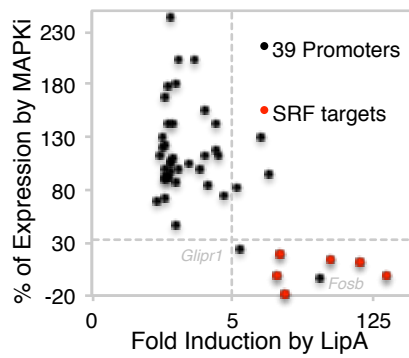


Figure 3-4. SRF Targets Share a Unique Promoter SRF-TCF Cassette

A

	Gene	Spacing	SRF Motif	TCF motif
Primary Response Genes	<i>Egr1</i>	4	C C A T A T A A G G	C C G G A
	<i>Dusp5</i>	12	C C A T A T A A G G	C A G G A
	<i>Egr2</i>	3	C C A T A T A T G G	G A G G A
	<i>Fos</i>	3	C C A T A T A G G	C A G G A
	<i>Nr4a1</i>	19	C C A T A C A A G G	C A G G A
Other Gene Classes	<i>Zfp36</i>	7	C C A T A A A A G G	C A G G A
	<i>Fosb</i>	13	C C A T A T A A G G	C A G G A
	<i>Glpr1</i>	5	C C A T A T A T G G	C C G G A

B



C

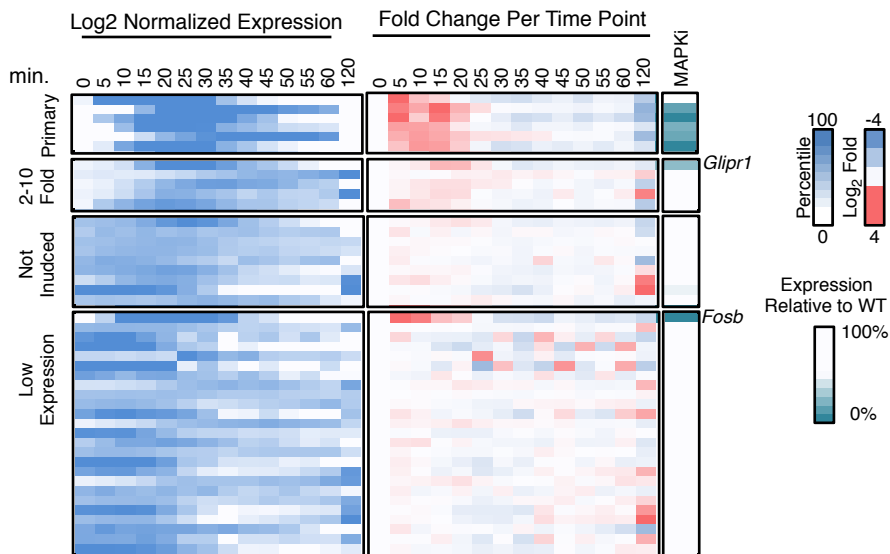
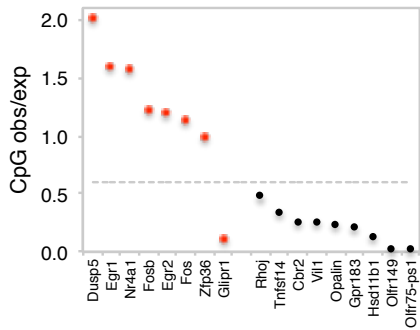


Figure 3-5. Identify Promoters Containing Defined SRF-TCF Cassettes and Lacking SRF Binding

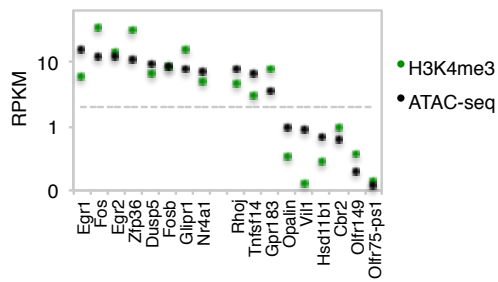
A

Gene	SRF							TCF			CpG obs/exp									
	TSS Distance	Motif						TSS Distance	Motif	Spacing (bp)										
<i>Olfir149</i>	-450	C	C	A	T	T	T	T	G	G	-442	C	A	G	G	A	G	3	0.03	
<i>Rhoj</i>	1	C	C	A	T	T	T	A	G	G	16	C	A	G	G	A	A	6	0.48	
<i>Vil1</i>	-75	C	C	A	T	A	T	A	A	G	G	-58	C	A	G	G	A	A	10	0.27
<i>Gpr183</i>	27	C	C	A	T	A	A	A	A	G	G	12	C	A	G	G	A	A	11	0.21
<i>Tnfrsf14</i>	-446	C	C	A	T	A	T	A	T	G	G	-424	C	A	G	G	A	A	14	0.35
<i>Opalin</i>	-65	C	C	A	T	A	C	A	T	G	G	-44	C	A	G	G	A	A	14	0.24
<i>Olfir75-ps1</i>	-145	C	C	A	T	A	A	T	T	G	G	-164	C	A	G	G	A	A	15	0.03
<i>Hsd11b1</i>	-442	C	C	A	T	A	T	T	T	G	G	-420	G	A	G	G	A	A	17	0.13
<i>Cbr2</i>	-110	C	C	A	T	A	T	T	T	G	G	-80	G	A	G	G	A	A	19	0.27

B



C



D

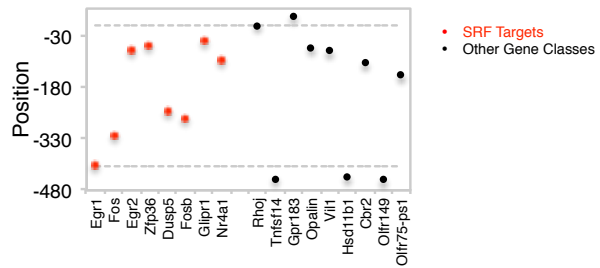


Figure 3-6. Interrogate the Potent SRF Binding at SRF Targets

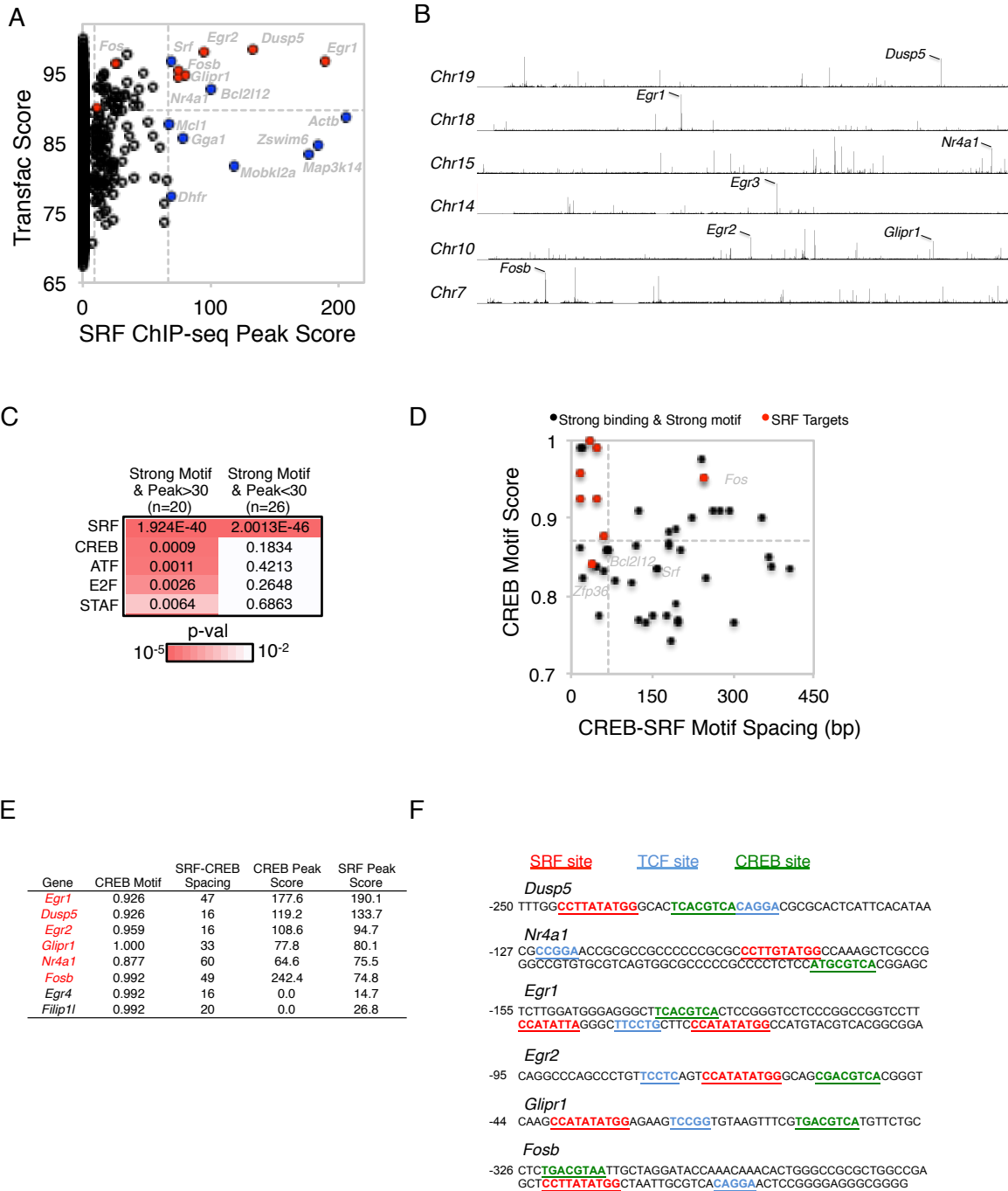


Figure 3-7. Strong Promoter SRF Peaks Possess Proximate CREB Binding and Multiple SRF Motifs

A

Gene	Number of SRF motifs	Min SRF Spacing	CREB site <70bp	TCF site <20bp	Peak score
<i>Egr1</i>	6	5	Yes	Yes	190.1
<i>Egr2</i>	3	35	Yes	Yes	94.7
<i>Dusp5</i>	2	0	Yes	Yes	133.7
<i>Glpr1</i>	2	341	Yes	Yes	80.1
<i>Nr4a1</i>	1	NA	Yes	Yes	75.5
<i>Fosb</i>	1	NA	Yes	Yes	74.8
<i>Fos</i>	1	NA	No	Yes	25.0
<i>Zfp36</i>	1	NA	No	Yes	10.7
<i>Bcl2l12</i>	2	416	No	No	100.2
<i>Srf</i>	2	10	No	No	68.8

B

SRF site TCF site CREB site

Egr1

-420 CGAC**CCGGA**AACG**CCATATAAGG**AG**CAGGA**AGGATCCCCG**CCGGA**ACA
 GA**CCCTATTGG**GCAGCG**CCCTATATGG**AGTGG**CCCAATATGG**CCCTGCC

-145 GAGGGCT**TCACGTCA**CTCCGGGTCTCCCGGCGGTCTT**CCATATTAGG**
 GCT**TCCTG**CTT**CCATATATGG**CCATGTACGTACGGCGGAGGCGGGCCC

Egr2

-450 GGAATGGCT**CCAAACAAGG**GCCGGGGAGGCGGAGCCGCCACT**CCGGA**
 TCTTCT**CCCTTTTGG**AAAGTCTCGGAGAACCAGAAATCTCCCGCCCAA

-100 GCTGCCAGGCCAGCCCTGT**TCCTC**AGT**CCATATATGG**GCAG**CGACGTCA**

Figure 3-8. SRF-TCF Motifs at Enhancer SRF Peaks Correlate with SRF Targets

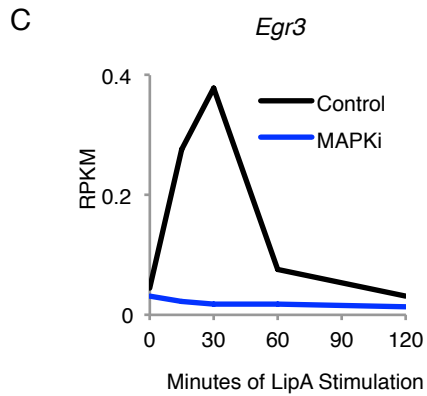
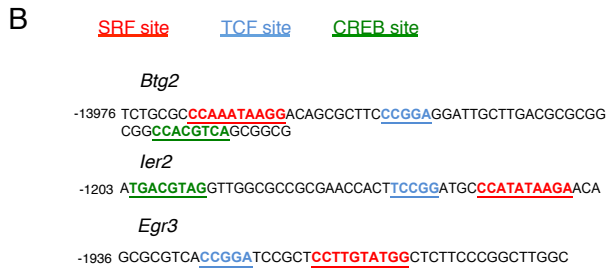
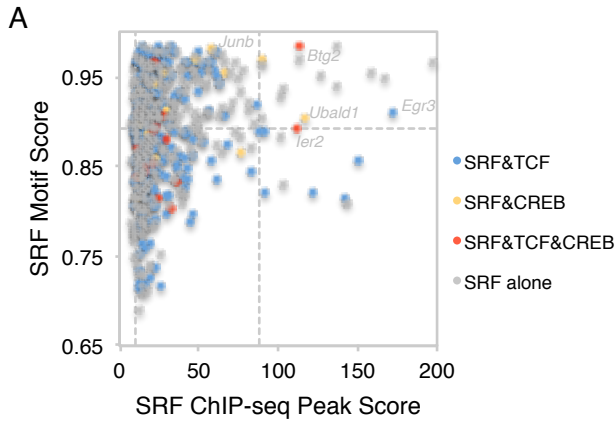


Figure 3-9. The Role of MRTF in Regulating Promoters with Strong SRF Binding and Strong Motifs

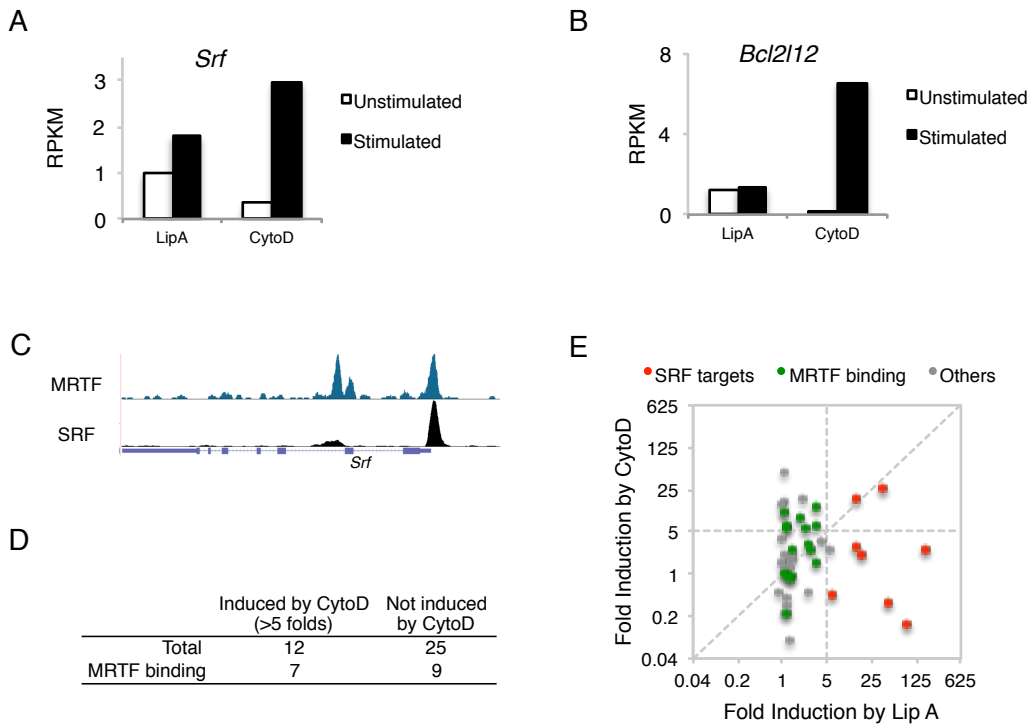


Figure 3-10. Evaluate the Roles of the Ternary Complex at the *Nr4a1* Promoter by CRISPR

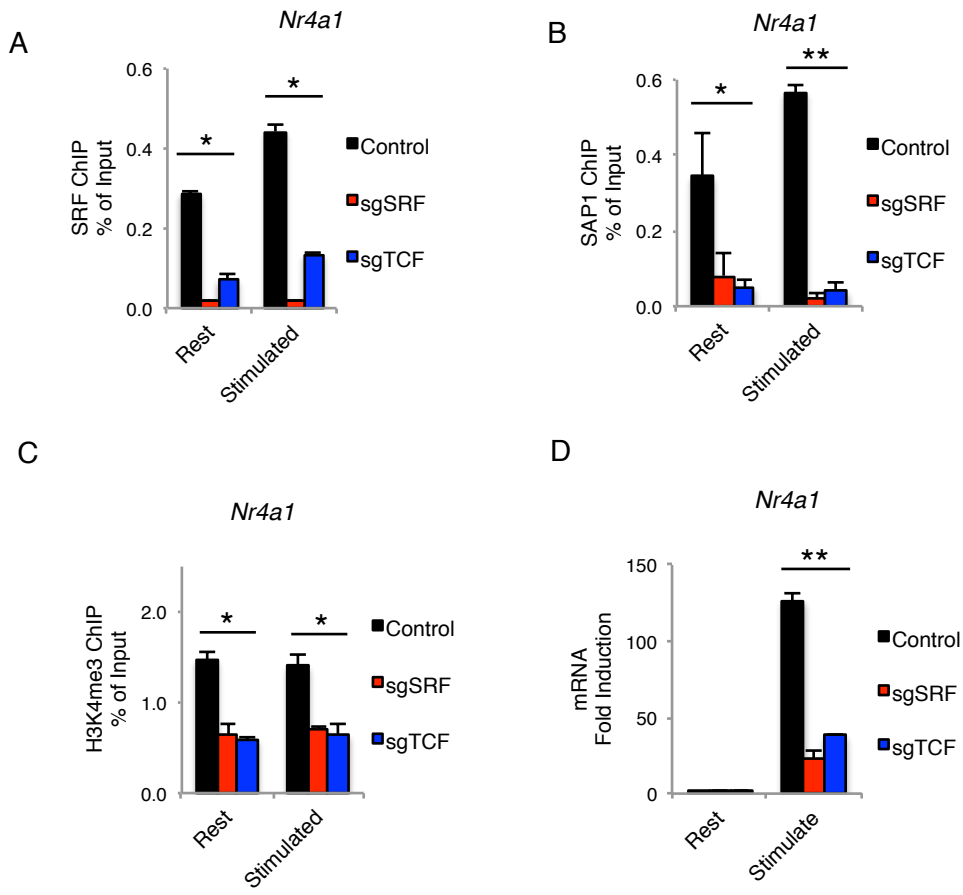
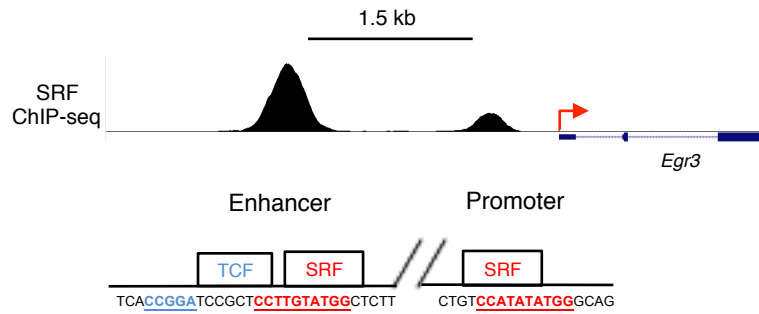
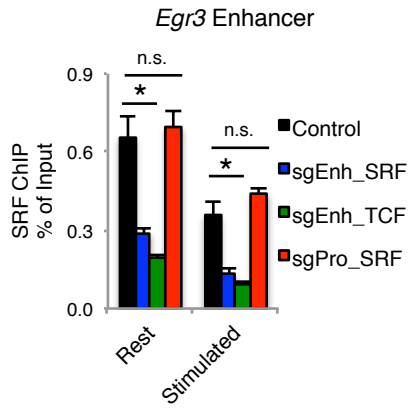


Figure 3-11. Evaluate the Roles of the Ternary Complex at the *Egr3* Enhancer by CRISPR

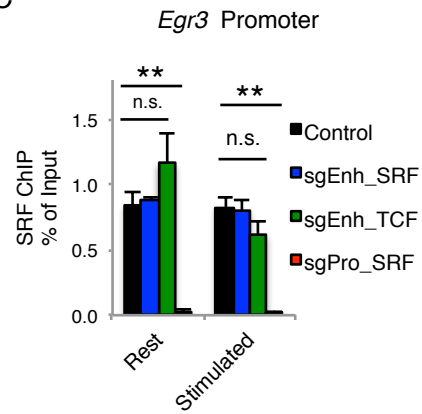
A



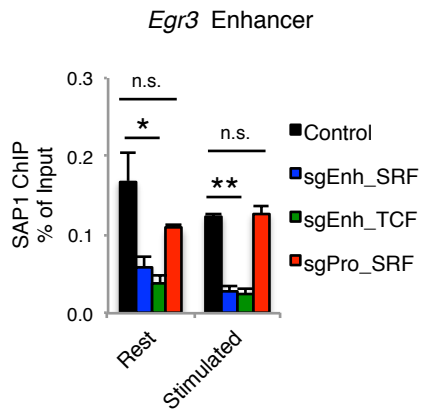
B



C



D



E

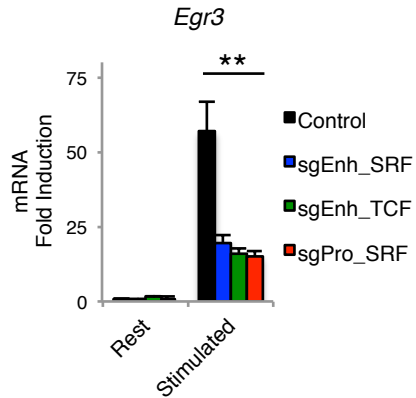


Figure 3-12. Robust Induction of SRF Targets by Different Stimuli in Different Cell Types

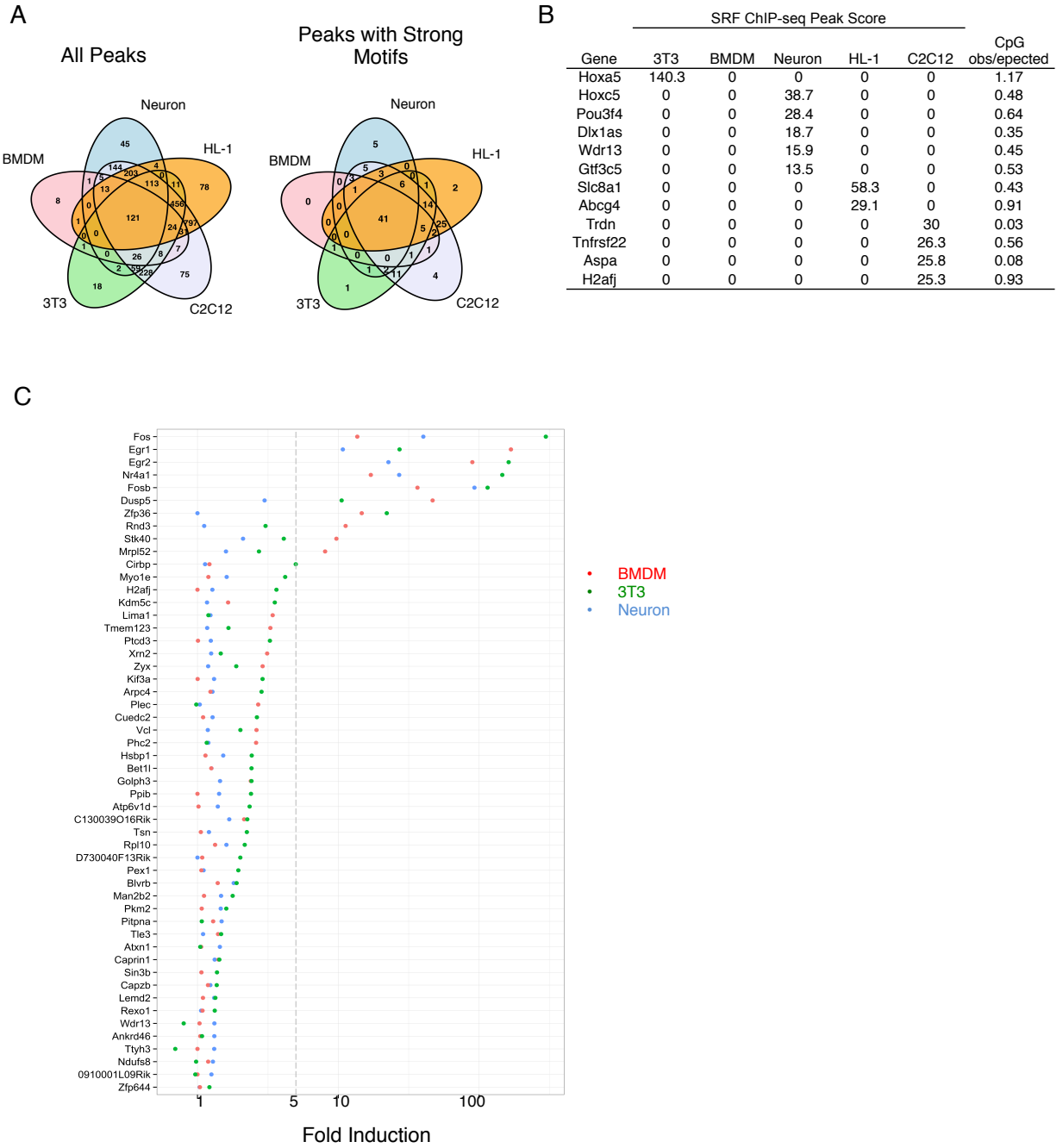


Table 3-1. List of Genes with SRF Binding or SRF Motifs

Refseq ID	Gene Symbol	Aliases	Function
SRF Targets			
NM_010234	<i>Fos</i>	D12Rfj1; c-fos; cFos	Transcriptional regulator
NM_011756	<i>Zfp36</i>	Gos24; Ttp; TIS11D; Zfp-36; Nup475; TISII; Tis11	RNA binding protein
NM_010444	<i>Nr4a1</i>	NGFIB; Hbr-1; TR3; NP10; Gfrp; Hbr1; NGFI-B; NUR77-2; NUR77-1; N10; nur77; TIS1; GFRP1; Hmr	Transcriptional regulator
NM_007913	<i>Egr1</i>	Krox24; TIS8; NGFIA; ETR103; NGF1-A; NGFI-A; Zfp-6; Zif268; A530045N19Rik; Krox-1; Egr-1; Krox-24; Zenk; egr	Transcriptional regulator
NM_010118	<i>Egr2</i>	Egr-2; NGF1-B; Zfp-25; Zfp-6; Krox-20; Krox20	Transcriptional regulator
NM_001085390	<i>Dusp5</i>	Gm337	Signaling molecule
NM_008036	<i>Fosb</i>	-	Transcriptional regulator
NM_028608	<i>Glipr1</i>	mRTVP-1; RTVP-1; 2410114O14Rik; RTVP1	Tumor suppressor
NM_007570	<i>Btg2</i>	AA959598; APRO1; Pc3; TIS21	Transcriptional regulator
NM_010499	<i>ler2</i>	Ch1; Pip92; AI317238	Transcriptional regulator
NM_018781	<i>Egr3</i>	Pilot	Transcriptional regulator
SRF-TCF Cassettes without Binding			
NM_207138	<i>Olf149</i>	MOR224-8; M31	G-protein-coupled receptor
NM_007621	<i>Cbr2</i>	MLCR	Enzyme
NM_153520	<i>Opalin</i>	Tmp10; Tmem10	Transmembrane protein
NM_183031	<i>Gpr183</i>	Ebi2	Signaling molecule
NM_019418	<i>Tnfrsf14</i>	LIGHT; HVEM-L; Tnlg1d; LTg; Ly113; HVEM-L	Cytokine
NM_023275	<i>Rhoj</i>	TC10L; TCL; 1110005O19Rik; AW210585; Arhj	Signaling molecule
NM_009509	<i>Vil1</i>	Vil	Signaling molecule
NM_008288	<i>Hsd11b1</i>	-	Enzyme
NR_033528	<i>Olf175-ps1</i>	V1; Olf175; MOR135-20	G-protein-coupled receptor
Other Genes			
NM_029410	<i>Bcl2l12</i>	Bcl2-L12; 2810475P17Rik; 5430429M05Rik; Bcl-L12	Anti-apoptotic factor
NM_020493	<i>Srf</i>	AW049942; AW240594	Transcriptional regulator
NM_028810	<i>Rnd3</i>	2610017M01Rik; AI661404; Rhoe; Arhe	Signaling molecule
NM_008562	<i>Mcl1</i>	AW556805; Mcl-1	Anti-apoptotic factor
NM_145929	<i>Gga1</i>	4930406E12Rik; AW209092; AU016030	Protein trafficking regulator
NM_010049	<i>Dhfr</i>	AI662710; AA607882; 8430436I03Rik; AW555094	Enzyme
NM_007393	<i>Actb</i>	E430023M04Rik; Actx; beta-actin	Cytoskeleton
NM_145456	<i>Zswim6</i>	2900036G02Rik; mKIAA1577	Metal ion binding
NM_172457	<i>Mobkl2a</i>	Mob3a; A630029F06; AV218468; 5330417K06Rik	Metal ion binding
NM_016896	<i>Map3k14</i>	aly; Nik	Signaling molecule
NM_020596	<i>Egr4</i>	pAT133; NGFIC; NGF1-C; NGFI-C	Transcriptional regulator
NM_001177871	<i>Filip1l</i>	Doc1; 4631422O05Rik	Signaling molecule
NM_008416	<i>Junb</i>	-	Transcriptional regulator
NM_145359	<i>Ubal1</i>	1500031H01Rik; Fam100a; BC013706	Unkonwn

Works Cited

- Ahn, S., Olive, M., Aggarwal, S., Krylov, D., Ginty, D.D., and Vinson, C. (1998). A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. *Mol. Cell. Biol.* *18*, 967–977.
- Bai, L., and Morozov, A.V. (2010). Gene regulation by nucleosome positioning. *Trends in Genetics* *26*, 476–483.
- Costello, P., Sargent, M., Maurice, D., Esnault, C., Foster, K., Anjos-Afonso, F., and Treisman, R. (2015). MRTF-SRF signaling is required for seeding of HSC/Ps in bone marrow during development. *Blood* *125*, 1244–1255.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* *25*, 1010–1022.
- Esnault, C., Gualdrini, F., Horswell, S., Kelly, G., Stewart, A., East, P., Matthews, N., and Treisman, R. (2017). ERK-induced activation of TCF family of SRF cofactors initiates a chromatin modification cascade associated with transcription. *Molecular Cell* *65*, 1081–1095.e5.
- Esnault, C., Stewart, A., Gualdrini, F., East, P., Horswell, S., Matthews, N., and Treisman, R. (2014). Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. *Genes Dev.* *28*, 943–958.
- Gerber, A., Esnault, C., Aubert, G., Treisman, R., Pralong, F., and Schibler, U. (2013). Blood-borne circadian signal stimulates daily oscillations in actin dynamics and SRF activity. *Cell* *152*, 492–503.
- Greenberg, M.E., and Ziff, E.B. (1984). Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* *311*, 433–438.
- Gualdrini, F., Esnault, C., Horswell, S., Stewart, A., Matthews, N., and Treisman, R. (2016). SRF co-factors control the balance between cell proliferation and contractility. *Molecular Cell* *64*, 1048–1061.
- Hassler, M. (2001). The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. *The EMBO Journal* *20*, 3018–3028.
- He, A., Kong, S.W., Ma, Q., and Pu, W.T. (2011). Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *PNAS* *108*, 5632–5637.

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. *Molecular Cell* 38, 576–589.

Herndon, C.A., Ankenbruck, N., Lester, B., Bailey, J., and Fromm, L. (2013). Neuregulin1 signaling targets SRF and CREB and activates the muscle spindle-specific gene *Egr3* through a composite SRF–CREB-binding site. *Experimental Cell Research* 319, 718–730.

Herrera, R.E., Shaw, P.E., and Nordheim, A. (1989). Occupation of the c-fos serum response element in vivo by a multi-protein complex is unaltered by growth factor induction. *Nature* 340, 68–70.

Herschman, H.R. (1991). Primary response genes induced by growth factors and tumor promoters. *Annual Review of Biochemistry* 60, 281–319.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10, 161–172.

K R Yamamoto, and Alberts, and B.M. (1976). Steroid receptors: elements for modulation of eukaryotic transcription. *Annual Review of Biochemistry* 45, 721–746.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.

Latinkic, B.V., Zeremski, M., and Lau, L.F. (1996). Elk-1 can recruit SRF to form a ternary complex upon the serum response element. *Nucl. Acids Res.* 24, 1345–1351.

Li, Q.-J., Yang, S.-H., Maeda, Y., Sladek, F.M., Sharrocks, A.D., and Martins-Green, M. (2003). MAP kinase phosphorylation-dependent activation of Elk-1 leads to activation of the co-activator p300. *The EMBO Journal* 22, 281–291.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 44, D110–D115.

Medjkane, S., Perez-Sanchez, C., Gaggioli, C., Sahai, E., and Treisman, R. (2009). Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nat Cell Biol* 11, 257–268.

Miano, J.M. (2010). Role of serum response factor in the pathogenesis of disease. *Lab Invest* 90, 1274–1284.

Murai, K., and Treisman, R. (2002). Interaction of serum response factor (SRF) with the Elk-1 B box inhibits rhoa-actin signaling to SRF and potentiates transcriptional activation by Elk-1. *Mol. Cell. Biol.* 22, 7083–7092.

Nissen, L.J., Gelly, J.-C., and Hipkind, R.A. (2001). Induction-independent recruitment of CREB-binding protein to the c-fos serum response element through interactions between the bromodomain and Elk-1. *J. Biol. Chem.* 276, 5213–5221.

Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55, 989–1003.

O'Donovan, K.J., Tourtellotte, W.G., Millbrandt, J., Baraban, J.M., O'Donovan, K.J., Tourtellotte, W.G., Millbrandt, J., and Baraban, J.M. (1999). The EGR family of transcription-regulatory factors: progress at the interface of molecular and systems neuroscience. *Trends in Neurosciences* 22, 167–173.

Parlakian, A., Tuil, D., Hamard, G., Tavernier, G., Hentzen, D., Concordet, J.-P., Paulin, D., Li, Z., and Daegelen, D. (2004). Targeted inactivation of serum response factor in the developing heart results in myocardial defects and embryonic lethality. *Mol. Cell. Biol.* 24, 5281–5289.

Pawłowski, R., Rajakylä, E.K., Vartiainen, M.K., and Treisman, R. (2010). An actin-regulated importin α -dependent extended bipartite NLS directs nuclear import of MRTF-A. *The EMBO Journal* 29, 3448–3458.

Pipes, G.C.T., Creemers, E.E., and Olson, E.N. (2006). The myocardin family of transcriptional coactivators: versatile regulators of cell growth, migration, and myogenesis. *Genes Dev.* 20, 1545–1556.

Poirier, R., Cheval, H., Mailhes, C., Garel, S., Charnay, P., Davis, S., and Laroche, S. (2008). Distinct functions of egr gene family members in cognitive processes. *Front Neurosci* 2, 47–55.

Pollock, R., and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res* 18, 6197–6204.

Posern, G., and Treisman, R. (2006). Actin' together: serum response factor, its cofactors and the link to signal transduction. *Trends in Cell Biology* 16, 588–596.

Ramanan, N., Shen, Y., Sarsfield, S., Lemberger, T., Schütz, G., Linden, D.J., and Ginty, D.D. (2005). SRF mediates activity-induced gene expression and synaptic plasticity but not neuronal viability. *Nat Neurosci* 8, 759–767.

Ramirez, S., Ali, S.A.S., Robin, P., Trouche, D., and Harel-Bellan, A. (1997). The CREB-binding Protein (CBP) Cooperates with the serum response factor for

transactivation of the c-fos serum response element. *J. Biol. Chem.* 272, 31016–31021.

Rieder, G., Tessier, A.J., Qiao, X.T., Madison, B., Gumucio, D.L., and Merchant, J.L. (2005). Helicobacter-induced intestinal metaplasia in the stomach correlates with elk-1 and serum response factor induction of villin. *J. Biol. Chem.* 280, 4906–4912.

Shore, P., and Sharrocks, A.D. (1994). The transcription factors Elk-1 and serum response factor interact by direct protein-protein contacts mediated by a short region of Elk-1. *Mol. Cell. Biol.* 14, 3283–3291.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050.

Sun, Q., Chen, G., Streb, J.W., Long, X., Yang, Y., Stoeckert, C.J., and Miano, J.M. (2006). Defining the mammalian CArGome. *Genome Res.* 16, 197–207.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086.

Tong, A.-J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. *Cell* 165, 165–179.

Treisman, R. (1986). Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *Cell* 46, 567–574.

Treisman, R., Marais, R., and Wynne, J. (1992). Spatial flexibility in ternary complexes between SRF and its accessory proteins. *EMBO J* 11, 4631–4640.

Vartiainen, M.K., Guettler, S., Larijani, B., and Treisman, R. (2007). Nuclear actin regulates dynamic subcellular localization and activity of the SRF cofactor MAL. *Science* 316, 1749–1752.

Vasudevan, H.N., and Soriano, P. (2014). SRF regulates craniofacial development through selective recruitment of MRTF cofactors by PDGF signaling. *Developmental Cell* 31, 332–344.

Vavouri, T., and Lehner, B. (2012). Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biology* 13, R110.

Vialou, V., Feng, J., Robison, A.J., Ku, S.M., Ferguson, D., Scobie, K.N., Mazei-Robison, M.S., Mouzon, E., and Nestler, E.J. (2012). Serum response factor and

cAMP response element binding protein are both required for cocaine induction of Δ FosB. *J. Neurosci.* 32, 7577–7584.

Wang, Z., Wang, D.-Z., Hockemeyer, D., McAnally, J., Nordheim, A., and Olson, E.N. (2004). Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression. *Nature* 428, 185–189.

Watson, D.K., Robinson, L., Hodge, D.R., Kola, I., Papas, T.S., and Seth, A. (1997). FLI1 and EWS-FLI1 function as ternary complex factors and ELK1 and SAP1a function as ternary and quaternary complex factors on the Egr1 promoter serum response elements. *Oncogene* 14, 213–221.

Zambelli, F., Pesole, G., and Pavesi, G. (2009). Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res* 37, W247–W252.

Zaromytidou, A.-I., Miralles, F., and Treisman, R. (2006). MAL and ternary complex factor use different mechanisms to contact a common surface on the serum response factor DNA-binding domain. *Mol. Cell. Biol.* 26, 4134–4148.

CHAPTER 4

Concluding Remarks: Conclusions and Future Directions

A balanced mammalian immune system is essential for host defense and homeostasis (Kotas and Medzhitov, 2015). Inefficient immune activation can cause susceptibility to cancer and pathogen infection. Over-activation of the immune system can cause allergy, autoimmunity, and chronic inflammatory diseases such as Crohn's diseases and rheumatoid arthritis (Eberl, 2016; Takeuchi and Akira, 2010). Novel therapies based on immune discoveries have proved to show efficacy in treating cancer and other diseases (Daniyan and Brentjens, 2017). To develop novel therapeutics for immune-related diseases, it is essential to understand the mechanisms regulating immune responses.

Upon pathogen infection, the innate immune cells often act as the first line of host defense. They can recognize pathogen components and damaged cell components by various pattern recognition receptors including Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), and nucleotide-binding oligomerization domain-like receptors (NLRs) (Brubaker et al., 2015; Takeuchi and Akira, 2010). Pattern recognition receptor activation will trigger many downstream pathways including MAPK, NF κ B, IRF, and PI3K. This signal cascade eventually leads to the transcriptional activation of thousands of pro-inflammatory genes and the production of antibacterial peptides, cytokines, and chemokine, which in turn attract adaptive immune cells and initiate the adaptive immune response. Activation of innate immune cells is a highly coordinated process that requires the precise regulation at transcriptional, post-transcriptional, translational, and post-translational steps. Because transcription acts as the first step for the production of inflammatory molecules in the innate immune response, we decided to dissect the

transcriptional cascades of a classical pattern recognition receptor: Toll-like receptor 4 (Li and Biggin, 2015; Vogel and Marcotte, 2012).

Although different stimuli can induce discrete subsets of pro-inflammatory and anti-viral genes, they activate innate immune cells through many common signaling pathways and transcription factors. The stimulus-specific induction of inflammatory genes through common regulators demands using selective gene regulatory mechanisms. While studies on inflammatory diseases often identify only a few cytokines and chemokines responsible for pathogenesis, it is difficult to inhibit them selectively. Chemical inhibitors generally target kinases and can affect many downstream genes. Antibody injection can selectively inhibit cytokines or chemokines, but it can lack selectivity for the inflammation sites and elicit systematic side effects. To develop more effective and selective therapies to treat inflammatory diseases, it is essential to understand both the common and gene-specific regulatory mechanisms controlling key inflammatory genes during inflammation.

To pursue our long-term goal to understand the selective regulatory mechanisms in the innate immune response, we started by studying a key pro-inflammatory gene that requires tight and selective regulation: *Il12b*. In Toll-like receptor 4 (TLR4) transcriptional cascades, *Il12b* is potently and selectively induced by the NFκB family member c-Rel (Sanjabi et al., 2005). *Il12b* induction is also optimized by many other transcription factors binding to its promoter and enhancer (Zhou et al., 2007; Bradley et al., 2003). Besides transcription factors, *Il12b* also requires stimulus-induced chromatin

remodeling at the promoter. The chromatin regulation of *I12b* precedes the binding of c-Rel and adds an additional layer of transcriptional regulation (Weinmann et al., 2001; Weinmann et al., 1999). Owing to the emerging next-generation sequencing technologies, we can apply our knowledge of *I12b* regulation to whole transcriptome studies and investigate similar or different regulatory strategies. We found that chromatin remodeling also controls the induction of many late primary response genes and secondary response genes. The requirement of chromatin remodeling is rooted in the CpG content of promoters. CpG-island promoters form unstable nucleosomes and are constitutively accessible, allowing for the rapid induction of many early primary response genes. Low CpG-island promoters form well-positioned nucleosomes that require chromatin remodeling, leading to delayed induction of late primary response genes and secondary response genes (Ramirez-Carrozzi et al., 2009).

Although many systems genomic analyses have dissected and uncovered some common transcriptional regulation mechanisms, their statistical approach relies on a large sample size, which can potentially miss unique mechanisms that only regulate a few genes. Thus, we need to employ a different strategy to explore gene-specific regulatory mechanisms. In this dissertation, we used a gene-centric, stringent system approach that highlights the quantitative nature of genomic data. In chapter 2, we used this approach to dissect the TLR4 transcriptional cascades and classified lipid A-induced genes by upstream signal pathways and transcription factors. We also explored the selective regulatory mechanisms by key transcription factors including NF κ B and SRF. To understand the selective and combinatorial regulation of transcription factors,

we investigated a classical SRF-TCF combinatorial regulation model and identified motif rules governing the activity of SRF and TCF.

Explore Gene-Specific Regulatory Mechanisms Using a Stringent Systems Approach

Many systems approaches assume that one transcription factor or pathway can regulate hundreds of genes. In these cases, statistical methods can efficiently reveal common regulatory mechanisms, but they may also miss selective mechanisms that regulate only a few genes. To reveal gene-specific regulatory mechanisms, we improved the experimental design and applied a stringent systems approach. Our strategies showed at least four advantages in unveiling the gene-specific regulatory mechanisms.

First, we measured the nascent transcript levels by isolating chromatin-associated RNA for RNA-seq analysis. This directly quantifies transcription rate and prevents the interference of mRNA stability. Second, we focused our analysis on strongly induced genes. Although key inflammatory genes are among the most strongly induced genes, strongly induced genes (fold change >10) only constitute 20% of all induced genes (fold change >2). Given that the strongly and weakly induced genes possess different transcriptional characteristics and are regulated by different mechanisms, separating strongly and weakly induced genes will prevent a biased conclusion towards the majority weakly induced genes. We can then apply the knowledge learned from strongly induced genes to studying weakly induced genes. Third, by focusing on strongly

induced genes, perturbation studies exhibited robust changes compared to control samples and gave us confidence to identify the targets regulated by different pathways and transcription factors. Last, we identified direct and functional protein binding events by combining the ChIP-seq and motif data. These data reveal that strong binding is highly correlated with strong motifs at the functional sites for two important transcription factors, NFκB and SRF.

By quantitatively dissecting TLR4 transcriptional cascades using stringent criteria, we classified the strongly induced genes by their kinetics and regulatory mechanisms. This also discloses several gene-unique regulatory mechanisms. We found that NFκB and IRF3 can selectively and collaboratively regulate five lipid A-induced genes through promoter binding. These five genes include *Ifnb1* and *Ccl5*. Interestingly, *Ifnb1* and *Ccl5* promoters use different strategies for NFκB and IRF3 interaction. While *Ifnb1* promoter is accessible before the cooperative binding of NFκB and IRF3, *Ccl5* promoter requires chromatin remodeling by IRF3 before NFκB binding. Another three NFκB/IRF3 genes also require strong remodeling at the promoter, but they probably depend on another factor CEBPβ. The differential regulation by NFκB and IRF3 demonstrates that critical inflammatory genes can achieve selective regulation by combining transcription factor binding, DNA context, and chromatin regulation.

Identify Rules Regulating the Functional Binding of SRF in a Lipid A Response

Besides five selective NFκB/IRF3 genes, we also found another small subset of genes selectively regulated by SRF. Although SRF binds to hundreds of promoters and

enhancers in the genome, only a small number of SRF targets were activated by lipid A. To understand the selective regulatory mechanisms of SRF targets, many studies used a peak-centric method. It first proposed the putative functional peaks and then tested them by statistical correlation with expression data. This method often lacks precision to understand the mechanistic details of complicated transcription events. To find the precise mechanisms regulating transcription, we employed a gene-centric method. We examined the shared characteristics and common patterns of SRF targets and identified rules that can explain the functional binding events.

By comparing the promoter SRF motifs of SRF targets, we found that functional SRF binding sites share a strict SRF-TCF cassette. This SRF-TCF cassette is rare; only another nine promoters in the mouse genome contain it, but they are inaccessible and thus are not inducible by lipid A. The characterized SRF-TCF cassette could also possibly explain the potent SRF binding at promoters of many SRF targets. We found that the SRF cofactor, TCF, can enhance SRF binding by forming a ternary complex. Ternary complex formation is required for the optimal H3K4me3 marking and transcriptional activation. Interestingly, SRF-TCF motifs also exist at enhancers that have exceptionally strong SRF binding near SRF targets. Thus, by studying the common motif features of SRF targets, we found that the selective activation of SRF targets require the combinatorial regulation of SRF and TCF. Because their SRF-TCF cassettes reside in CpG-island promoters, which are constitutively accessible, these SRF targets are also inducible in other cell types and are sensitive to different TCF activators. Given that most SRF targets regulate essential cellular functions including

cell growth, proliferation, and apoptosis, the selective regulation by SRF and TCF ensures these key genes can faithfully respond to various environmental stimuli in different cell contexts.

Implications and Future Directions

The analyses in this dissertation demonstrate our initial efforts to surface the mechanisms underlying the complicated transcriptional cascades in the innate immune response. They also provide many interesting observations that warrant further investigation. One frequent observation common for many transcription factors is the skewed distribution of ChIP-seq peak score (Figure 4-1). It is interesting that the peak score distribution does not exhibit a normal distribution. Normal distribution is common for many other biological features including the cell size, body height, and blood pressure (Marshall et al., 2012; A'HEARN et al., 2009; Wright et al., 2011). These biological features are tightly regulated by many factors within a narrow range. The stringent regulation will exclude any individual that falls out of the optimal range, resulting in a relatively homogeneous population close to an optimal value. On the other hand, the skewed distribution is more frequently seen in phenomena in social sciences such as the distribution of wealth (Berman et al., 2016; Bassani et al., 2014). In these cases, the population is rather heterogeneous; a small subset of extremely strong events dramatically differs from the majority of weak events. The sum of this small percent of extreme events usually has greater effects than the sum of all weak events.

The skewed distribution of ChIP-seq peak score implies that the transcription factor

binding *in vivo* might follow a different regulatory mechanism from other biological features that results in homogenous populations. The homogenous populations with normal distribution may result from factors excluding abnormal individuals beyond the optimal range by affecting their essential functions. The heterogeneous population of ChIP-seq peaks implies that non-functional binding *in vivo* does not adversely affect key cellular functions, and thus are not excluded by evolution. The prevalent redundant “non-functional” binding sites may prepare the cells to evolve new functional *cis*-regulatory elements in an ever-changing environment. But it is also likely the non-functional binding is a consequence of technical artifacts. Because the functional sites are relatively few (for example, only 8 of 166 SRF promoter peaks contribute to a lipid A response) in many contexts, studying all sites together will lead to biased conclusions that overshadow the regulatory mechanisms of functional sites. Thus, it is essential to separate the functional and non-functional sites using stimulus-specific criteria.

Many studies have endeavored to find a universal strategy to identify functional ChIP-seq peaks. Some studies defined functional sites with the strongest peaks, which are called “super enhancers.” They are usually much stronger or wider peaks than normal enhancers (Pott and Lieb, 2015; Whyte et al., 2013; Hnisz et al., 2013). Consistently, most of our functional SRF peaks rank among the strongest peaks. However, not all strong peaks are functional in a certain context. Some peaks may be active in one context, but not in another. It is tempting to hypothesize that peaks frequently activated by various environmental stimuli during evolution are more important, so they are more likely to be conserved and become stronger than other less important peaks. Thus, the

strong peaks may be more likely to represent functionally important sites during the evolution, which are often selective to different contexts or stimuli. But the context-specific regulatory mechanisms are largely unknown. The statistical methods often require large sample sizes and can potentially ignore the relatively few number of context-specific peaks. By identifying the patterns of lipid A-specific targets, we can avoid the limitation of statistics and identify rules that selectively activate important transcription factors. Although in this dissertation, we have identified motif requirements as a critical factor that distinguishes the functional sites of SRF in a lipid A response, we cannot exclude the possibility that other factors can also contribute to the functional binding and transcriptional activation. Other mechanisms might include motif location, chromatin state, and histone marks. Future studies can address the possibility of other mechanisms by studying the selective transcriptional regulation under different stimuli and in different cell types.

We focused our analyses on promoters because they are better characterized than enhancers. In future studies, we will extend our analysis to enhancers, which might require different regulatory mechanisms from promoters. Our CRISPR mutation experiments on *Egr3* revealed that both the enhancer and the promoter are essential for *Egr3* gene induction. This implies that the optimal induction of *Egr3* requires the collaboration of the enhancer and promoter through promoter-enhancer interaction. The step-wise mechanism of this induction remains unclear. We can try to understand the promoter-enhancer interaction by HiC and mutation experiments. Although we still lack the knowledge to identify the targets of enhancers, we can get initial insights from the

regulatory mechanisms of individual enhancers like the *Egr3* enhancer and extend our insights to broader studies. Understanding transcription factor regulation at enhancers will also allow for a better characterization of enhancers.

Although we have demonstrated that some transcription factors can activate transcription by altering chromatin features, it is still unclear how transcription factor binding leads to chromatin changes. In our studies, we find that IRF3 binding to the *Ccl5* promoter precedes SWI/SNF binding, chromatin remodeling, NFκB binding, and transcriptional induction. It is still unclear how IRF3 binds to an inaccessible region and initiates remodeling. It is very likely that it binds at the edge of a well-positioned nucleosome or on a small accessible window of the nucleosome. We need further investigation into the stepwise mechanistic details of this process in future studies. Intriguingly, IRF3 can also bind to accessible promoters that do not require chromatin remodeling. It will be interesting to understand whether IRF3 can detect chromatin contexts to apply different regulatory mechanisms. To understand the roles of IRF3 in initiating chromatin remodeling, we can also compare it with other transcription factors that can also induce chromatin remodeling, particularly transcription factors in the IRF family.

Besides combinatorial regulation, many other mechanisms could potentially contribute to selective transcriptional induction. One mechanism involves the selective binding of different transcription factor members that have slightly different motif preference. For example, NFκB family transcription factors can recognize and bind ubiquitously to many

canonical NFκB sites. However, the *I12b* promoter contains a non-canonical NFκB site, which is highly selective for c-Rel. This preference leads to selective inhibition of *I12b* in c-Rel knockout cells without affecting other NFκB targets (Sanjabi et al., 2005). Similar preference for a certain transcription factor member might also exist in other transcription families. For example, AP1 family transcription factors consist of Fos proteins and Jun proteins. Different combinations of Fos and Jun dimers can result in different binding preference (Mechta-Grigoriou et al., 2001; Karin et al., 1997).

Another possible selective regulatory mechanism is the use of alternative cofactors. For example, SRF activity requires two competing cofactors: MRTF transcription factors and TCF transcription factors. While TCF family members can form ternary complexes with SRF and activate genes regulating cell proliferation, apoptosis, and circadian clock, MRTF family members can bind to SRF and activate cytoskeletal genes. One member of the MRTF family, myocardin, is selectively expressed in muscle cells and regulates muscle-specific genes (Olson and Nordheim, 2010; Posern and Treisman, 2006;). Another example is CREB, which can interact with two cofactors: CBP and CRTC2 (Altarejos and Montminy, 2011; Mayr and Montminy, 2001). While the CBP-CREB interaction controls genes regulating broad functions, CRTC2-CREB can selectively activate genes regulating metabolic functions.

To get a more comprehensive understanding of mechanisms regulating lipid A transcriptional cascades, we can use our stringent systems approach to examine the binding profiles of other transcription factors. Besides NFκB, IRF3, and SRF, many

other transcription factors also contribute to the lipid A response, which include CEBP, AP1, and CREB. We can explore the ChIP-seq data from these factors with respect to three aspects.

First, we can dissect their binding profiles using our gene-centric approach and investigate the mechanisms that activate lipid A-responsive sites. Second, lipid A can induce different transcription factor members in the same family. It will be interesting to compare the common and unique properties of different transcription factor members. By comparing their kinetics, peak strength, motifs, and other characteristics, we can uncover the selective mechanisms underlying the redundant and differential use of transcription factor members. Furthermore, we can integrate transcription factor binding properties with other features including chromatin accessibility and histone mark data. The potential coordination of chromatin regulation and transcription factor will illuminate on the complexity of transcription factor regulation in the native environment. Third, several transcription factors often act in concert to induce transcription. After understanding how each transcription factor functions individually, we can then explore their combinatorial regulation. For example, besides NF κ B and IRF3, CEBP β also binds to the promoters of three NF κ B/IRF3 genes. These three NF κ B/IRF3 genes undergo chromatin remodeling before NF κ B and IRF3 binding. It is likely that CEBP β binds at these three NF κ B/IRF3 genes to initiate chromatin remodeling, which permits NF κ B and IRF3 binding and gene induction. Taking advantage of the well-defined lipid A-induced gene classes, we can investigate more selective mechanisms by combination of transcription factors. Eventually, by clarifying the roles of different transcription factors,

we can refine the classification of lipid A-induced genes. This will enable us to evaluate the contribution of key transcription factors during inflammation.

There are also many challenges for analyzing ChIP-seq data from different transcription factor family members. Because transcription factor family members may share homologous domains, it may be technically challenging to find specific ChIP antibodies for each transcription factor member. Furthermore, there is often functional redundancy between different family members, so knockout of one gene might not uncover all targets. Mutating all transcription factor members is technically difficult and is often lethal. The difficulty to identify targets might compromise the identification of functional ChIP-seq peaks. If a transcription factor is selectively expressed in macrophages, there is an alternative approach. We can compare the gene expression in other cell types. If a gene is selectively activated only in macrophages, then it is more likely to be a real target.

Taken together, by carefully dissecting the lipid A transcriptional cascades and uncovering selective regulatory mechanisms of transcription factors, we have demonstrated the efficacy and potential of mining the gene-specific regulatory mechanisms using our stringent systems approach. This approach will benefit future genomic studies and provides potential applications in medicine. With the emergence of cheaper personal sequencing technologies, it is likely that clinicians will adopt genomic analyses for personalized medicine in the future. By comparing the gene expression profiles of the control and patient samples, it is possible to make more accurate

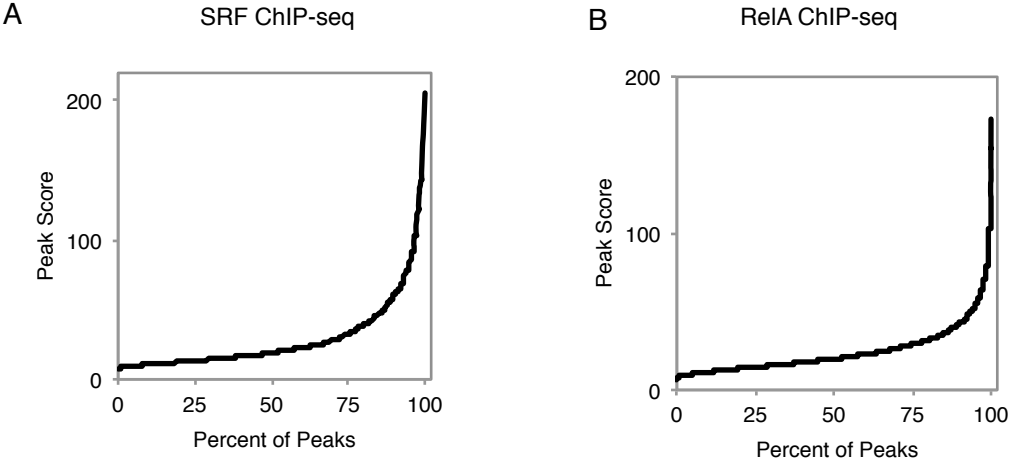
diagnoses and find better treatment options. However, inflammatory diseases often disrupt complicated gene networks involving both common and gene-specific regulatory mechanisms. Conventional systems approaches can detect large-scale expression changes regulated by common mechanisms, but our approach can also detect unique changes regulated by gene-specific mechanisms. The combination of both approaches will provide a precise diagnosis of complicated inflammatory diseases by identifying critical regulators in disease pathogenesis. Besides the applications in disease diagnosis, investigation of gene-specific mechanisms regulated by transcription factors, chromatin regulation, and other mechanisms will also facilitate the development of selective medications to treat human diseases with minimal side effects.

Figure Legend

Figure 4- 1 The Distribution of SRF and RelA ChIP-seq Peak Scores

(A) Line graph displays the peak score of 718 reproducible SRF ChIP-seq peaks identified by HOMER software. Peaks are ranked by peak strength. (B) Line graph displays the peak score of 8458 reproducible RelA ChIP-seq peaks identified by HOMER. Peaks are ranked by peak strength.

Figure 4-1 The Distribution of SRF and RelA CHIP-Seq Peak Score



Works Cited

A'HEARN, B., PERACCHI, F., and VECCHI, G. (2009). Height and the normal distribution: evidence from italian military data. *Demography* 46, 1–25.

Altarejos, J.Y., and Montminy, M. (2011). CREB and the CRTC co-activators: sensors for hormonal and metabolic signals. *Nat Rev Mol Cell Biol* 12, 141–151.

Bassani, D.G., Corsi, D.J., Gaffey, M.F., and Barros, A.J.D. (2014). Local distributions of wealth to describe health inequalities in india: a new approach for analyzing nationally representative household survey data, 1992–2008. *PLOS ONE* 9, e110694.

Berman, Y., Ben-Jacob, E., and Shapira, Y. (2016). The dynamics of wealth inequality and the effect of income distribution. *PLOS ONE* 11, e0154196.

Bradley, M.N., Zhou, L., and Smale, S.T. (2003). C/EBP β regulation in lipopolysaccharide-stimulated macrophages. *Mol. Cell. Biol.* 23, 4841–4858.

Brubaker, S.W., Bonham, K.S., Zanoni, I., and Kagan, J.C. (2015). Innate immune pattern recognition: a cell biological perspective. *Annual Review of Immunology* 33, 257–290.

Daniyan, A.F., and Brentjens, R.J. (2017). Immunotherapy: hiding in plain sight: immune escape in the era of targeted T-cell-based immunotherapies. *Nat Rev Clin Oncol advance online publication*.

Eberl, G. (2016). Immunity by equilibrium. *Nat Rev Immunol* 16, 524–532.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H., and Young, R.A. (2013). Transcriptional super-enhancers connected to cell identity and disease. *Cell* 155.

Karin, M., Liu, Z., and Zandi, E. (1997). AP-1 function and regulation. *Current Opinion in Cell Biology* 9, 240–246.

Kotas, M.E., and Medzhitov, R. (2015). Homeostasis, Inflammation, and Disease Susceptibility. *Cell* 160, 816–827.

Li, J.J., and Biggin, M.D. (2015). Statistics requantitates the central dogma. *Science* 347, 1066–1067.

Marshall, W.F., Young, K.D., Swaffer, M., Wood, E., Nurse, P., Kimura, A., Frankel, J., Wallingford, J., Walbot, V., Qu, X., et al. (2012). What determines cell size? *BMC Biology* 10, 101.

- Mayr, B., and Montminy, M. (2001). Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat Rev Mol Cell Biol* 2, 599–609.
- Mechta-Grigoriou, F., Gerald, D., and Yaniv, M. (2001). The mammalian Jun proteins: redundancy and specificity. *Oncogene* 20, 2378–2389.
- Olson, E.N., and Nordheim, A. (2010). Linking actin dynamics and gene transcription to drive cellular motile functions. *Nat Rev Mol Cell Biol* 11, 353–365.
- Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon- β enhanceosome. *Cell* 129, 1111–1123.
- Posern, G., and Treisman, R. (2006). Actin' together: serum response factor, its cofactors and the link to signal transduction. *Trends in Cell Biology* 16, 588–596.
- Pott, S., and Lieb, J.D. (2015). What are super-enhancers? *Nat Genet* 47, 8–12.
- Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by cpg islands and nucleosome remodeling. *Cell* 138, 114–128.
- Sanjabi, S., Williams, K.J., Sacconi, S., Zhou, L., Hoffmann, A., Ghosh, G., Gerondakis, S., Natoli, G., and Smale, S.T. (2005). A c-Rel subdomain responsible for enhanced DNA-binding affinity and selective gene activation. *Genes Dev.* 19, 2138–2151.
- Takeuchi, O., and Akira, S. (2010). Pattern recognition receptors and inflammation. *Cell* 140, 805–820.
- Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13, 227–232.
- Weinmann, A.S., Mitchell, D.M., Sanjabi, S., Bradley, M.N., Hoffmann, A., Liou, H.-C., and Smale, S.T. (2001). Nucleosome remodeling at the IL-12 p40 promoter is a TLR-dependent, Rel-independent event. *Nat Immunol* 2, 51–57.
- Weinmann, A.S., Plevy, S.E., and Smale, S.T. (1999). Rapid and selective remodeling of a positioned nucleosome during the induction of IL-12 p40 transcription. *Immunity* 11, 665–675.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.

Wright, J.D., Hughes, J.P., Ostchega, Y., Yoon, S.S., and Nwankwo, T. (2011). Mean systolic and diastolic blood pressure in adults aged 18 and over in the United States, 2001-2008. *Natl Health Stat Report* 1–22, 24.

Zhou, L., Nazarian, A.A., Xu, J., Tantin, D., Corcoran, L.M., and Smale, S.T. (2007). An inducible enhancer required for *Il12b* promoter activity in an insulated chromatin environment. *Mol. Cell. Biol.* 27, 2698–2712.