UNIVERSITY OF CALIFORNIA

Los Angeles

Mechanisms Regulating Selective Gene Activation During the Innate Immune Response

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

by

Ann-Jay Tong

2015

ABSTRACT OF THE DISSERTATION

Mechanisms Regulating Selective Gene Activation

During the Innate Immune Response

by

Ann-Jay Tong

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2015

Professor Stephen Smale, Chair

The cells of the innate immune system are responsible for the first line of defense against foreign dangers. Recognition of pathogens results in the transcriptional upregulation of a stimulus-specific inflammatory gene program to counteract infection and initiate adaptive immune responses. An appropriate response is necessary to resolve infection, but excessive inflammation can damage host tissues and lead to inflammatory diseases. Therefore, it is critical to understand how inflammatory responses are selectively achieved in response to diverse stimuli. The dissertation describes two studies that attempt to better understand the regulatory mechanisms underlying selective gene activation in response to inflammatory stimuli. The first study explores how signaling pathways, transcription factors, and chromatin act in concert to shape the inflammatory gene program. Using genome-wide techniques to interrogate chromatin-associated RNA, insight was gained into the lipid A-induced transcriptional cascade in macrophages. A quantitative analysis of transcription factor binding combined with kinetic and expression data derived from loss-of-function mutant mouse strains have allowed us to identify co-regulated genes, particularly those regulated by NF-κB, interferon response factor-3, and

serum response factor. Furthermore, subsets of co-regulated secondary response genes were found to play distinct roles in immunity, underscoring the diverse mechanisms underlying selective gene activation. This has revealed insight into the unique regulatory logic for each inflammatory gene, and serves as a framework for understanding selective gene activation in various physiological settings. The second study utilizes the findings from the first study to investigate the mechanisms of LPS tolerance. A global interrogation of the effects of LPS tolerance in macrophages revealed a broad downregulation of gene expression in the tolerant state. In addition, a large subset of inducible genes exhibited prolonged transcription even after the tolerizing dose of LPS was removed but before the second LPS treatment, which could be partially explained by the presence of other cytokines mediating their activation in the tolerant state. Furthermore, previously described negative regulators of LPS signal transduction were expressed at higher levels in tolerant macrophages, including those inhibiting signals proximal to the TLR4 receptor. Together, the framework for understanding the regulatory logic of selective gene activation that can be utilized to unravel the mechanisms underlying diverse inflammatory settings.

The dissertation of Ann-Jay Tong is approved.

Linda Baum

Douglas Black

Genhong Cheng

Stephen Smale, Committee Chair

University of California, Los Angeles

2015

DEDICATION

I would like to dedicate this research to my parents, and to my best friend and partner Kent, for their support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

FIGURES

TABLES

ACKNOWLEDGEMENTS

Chapter 1 contains reprints of tables and figures with permission. Table 1-1 is a reprint granted by Elsevier (License #3606161198383). Figure 1-1 is a reprint granted by Nature Publishing Group (License #3606170157971).

Chapter 2 is a version of a manuscript currently in preparation. Xin Liu performed RNA-sequencing experiments of 5-minute lipid A stimulated macrophages, MAP kinase inhibitor-treated macrophages, and ChIP-sequencing experiments of serum response factor. Grant Barish conducted the ChIP-sequencing expriments of RelA.

Appendix A is a reprint of an article titled “Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions” published in Cell *150(2)*, 279-290, 2012. Dev Bhatt, Amy Pandya-Jones, Iros Barozzi, Michelle Lissner, Gioacchino Natoli, Douglas Black, and Stephen Smale were co-authors in this study.

Appendix B is a reprint of an article titled “Splicing kinetics and transcript release from the chromatin compartment limit the rate of lipid A-induced gene expression” published in RNA *19*, 811-827, 2013. Amy Pandya-Jones, Dev Bhatt, Chia-Ho Lin, Stephen Smale, and Douglas Black were co-authors in this study.

Appendix C is a version of a manuscript titled “Age Related Gene Expression Differences in Monocytes from Human Neonates, Young Adults, and Older Adults” submitted to *PLoS One.* Michelle Lissner, Brandon Thomas, Kathleen Wee, Tobias Kollman, and Stephen Smale were co-authors in this study.

All of the work described in the dissertation was directed by Stephen Smale.

The work was supported by the UCLA Tumor Immunology Training Program (USHHS Ruth L. Kirschstein Institutional National Research Service Award # T32 CA009120).

VITA

2006	Bachelor of Arts, Molecular and Cell Biology/Cell and Developmental Biology
	University of California
	Berkeley, California
2006-2008	Laboratory Technician
	California Institute for Medical Research
	San Jose, California
2009-2010	Teaching Assistant
	Department of Life Sciences
	University of California, Los Angeles
	Los Angeles, California
2012-2015	Tumor Immunology Training Grant Recipient
	University of California, Los Angeles
	Los Angeles, California

PUBLICATIONS

Smale S.T., Plevy S.E., Weinmann A.S., Zhou L., Ramirez-Carrozzi V.R., Pope S.D., Bhatt D.M., **Tong A.J.**: Toward and understanding of the gene-specific and global logic of inducible gene transcription. *Cold Spring Harbor Symposium of Quantitative Biology*, Volume LXXVIII, April 2014.

Pandya-Jones A., Bhatt D.M., Lin C.H., **Tong A.J.**, Smale S.T., Black D.L.: Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* 19*,* 811-826, June 2013.

Bhatt D.M., Pandya-Jones A., **Tong A.J.**, Barozzi I., Lissner M., Natoli G., Black D.L., Smale S.T.: Transcript dynamics of pro-inflammatory genes uncovered by RNA-Seq analysis of subcellular RNA fractions. *Cell* 150*,* 279-290, July 2012.

Clemons K.V., Martinez M., **Tong A.J.**, Stevens D.A.: Resistance of MBL gene-knockout mice to experimental systemic aspergillosis. *Immunology Letters* 120, 105-107*,* February 2010.

Clemons K.V., Capilla J., Sobel R.A., Martinez M., **Tong A.J.**, Stevens D.A.: Comparative efficacies of lipid-complexed amphotericin B and liposomal amphotericin B against coccidioidal meningitis in rabbits. *Antimicrobial Agents and Chemotherapy* 53, 1858-1862*,* May 2009.

Van Asbeck E., Clemons K.V., Martinez M., **Tong A.J.**, Stevens D.A.: Significant differences in drug susceptibility among species in the *Candida parapsilosis* group. *Diagnostic Microbiology and Infectious Disease* 62, 106-109*,* September 2008.

PRESENTATIONS

Tong A.J., Liu X., Thomas B.J., Lissner M.L., Smale S.T.: Understanding the transcriptional program during macrophage activation. Cold Spring Harbor Laboratory Meetings: Gene Expression and Signaling in the Immune System, Cold Spring Harbor, NY, April 2014.

Tong A.J., Bhatt D.M., Liu X., Smale S.T.: Understanding the transcriptional program of macrophage activation. 78th Cold Spring Harbor Symposium on Quantitative Biology: Immunity & Tolerance, Cold Spring Harbor, NY, May 2013.

Tong A.J., Bhatt D.M., Smale S.T.: Understanding the selective activation of pro-inflammatory genes by RNA-Seq. ImmunologyLA, Los Angeles, CA, May 2012.

Tong A.J., Bhatt D.M., Smale S.T.: Understanding the selective activation of pro-inflammatory genes by RNA-Seq. Cold Spring Harbor Laboratory Meetings: Gene Expression and Signaling in the Immune System, Cold Spring Harbor, NY, April 2012.

Tong A.J., Bhatt D.M., Smale S.T.: Selective gene activation during an inflammatory response. ImmunologyLA, Malibu, CA, April 2011.

CHAPTER 1

Introduction

Transcriptional Regulation of the Innate Immune System

A. Pathogen Recognition and Response in the Innate Immune System

The immune response is central to the well-being and survival of all living organisms. A major task of the immune system is to distinguish self from non-self and generate a suitable response. Appropriate detection of foreign dangers such as pathogens is critical to protect against diseases and maintain host survival. In jawed vertebrates, the immune system is composed of two major strategies, the innate and adaptive immune systems. Because it is utilized by both vertebrates and invertebrates, the innate immune response is considered the more primitive form of immunity and is the first line of defense against foreign and environmental insults such as invading microorganisms and tissue injury[1–5]. Recognition of external threats occurs in a relatively non-specific and generic manner, and immunity is short-lived. Upon recognition of foreign insults, cells of the innate immune system such as macrophages and dendritic cells recruit and direct other cells to the site of infection or injury by cytokine and chemokine secretion, release of antimicrobial factors that can directly combat infection, and initiate tissue repair and recovery[1,3]. In vertebrates, the innate immune response also serves as the key trigger of the adaptive immune response. In addition to antigen presentation to cells of the adaptive immune system, innate immune cells must provide the context of the infection by inducing proper signals such as specific cytokine release and co-stimulatory molecule expression that instruct adaptive immune cells to mount the appropriate response[5–7]. The adaptive immune system, conserved in all jawed vertebrates, is composed of B and T lymphocytes that can be activated in a pathogen-specific manner by antigen-presenting cells (APCs) such as macrophages and dendritic cells, resulting in the elimination of the specific pathogen. Although activation of the adaptive immune system is delayed because it relies on signals from the innate immune response, adaptive immunity is long lasting and protective against subsequent infection by the same pathogen. Antigens from pathogens are processed into unique peptides and coupled to major histocompatibility complex (MHC) receptors on the

cell surface of APCs. In order to be able to recognize the diverse array of peptides, each lymphocyte bears a unique antigen-specific receptor established through V(D)J recombination. High affinity interactions between the antigen-MHC complex on APCs and antigen receptor on lymphocytes result in clonal expansion of the antigen-specific lymphocyte, removal of the specific pathogen and pathogen-infected cells, and development of long term memory to prevent subsequent infection by the pathogen.

The immune response to foreign pathogens critically relies on the innate immune system. The initial innate immune response, although relatively non-specific, occurs rapidly upon detection to establish a pro-inflammatory and antimicrobial state in the host. Innate immune cells then recruit and activate adaptive immune cells through antigen presentation, resulting in a long term pathogen-specific response and protection. Thus, the innate immune response plays a vital role in mediating both early and long-term immunity against infection.

A1. Pathogen Receptors

The innate immune system is critical to immediate detection of pathogen or tissue damage and initiating a specific response towards the detected threat. In order to induce the appropriate immune response to clear infection and maintain immunity, the innate immune system must have the ability to sense a diverse array of microorganisms that infect hosts both extracellularly and intracellularly. Although the innate immune response is antigen non-specific, sensing mechanisms exist to differentiate between a wide range of pathogens. Mammalian innate immune cells as well as other nonprofessional cells of the immune system express a family of pattern recognition receptors (PRRs) found in various cellular locations. Unlike antigen receptors generated by somatic DNA recombination to establish antigen specificity, PRRs are germline-encoded receptors that have evolved to sense evolutionarily conserved elements found on microorganisms. These pathogen-associated molecular patterns (PAMPs) are

molecular structures uniquely found on the microorganism, thus providing the innate immune system with a mechanism to differentiate between self and non-self. Several classes of PRRs have been identified and characterized, including the Toll-like receptors, NOD-like receptors, RIG-I-like receptors, and cytosolic DNA sensors[6,8]. Members within each family of PRRs detect a unique PAMP, and effective sensing rapidly activates the innate immune response, leading to induction of proinflammatory cytokines, release of antimicrobial factors, and eventual clearance of pathogen.

The first identified and most widely studied PRRs are the Toll-like receptor (TLR) family. Although initially identified as a regulator of development in fruit fly embryos[9], Toll was later shown to play a critical role in fruit fly resistance to infection[9,10]. This led to a series of key advances including the cloning of the first human TLR gene presently known as *TLR4*[11], demonstration that activation of this human Toll receptor preceded NF-κB transcription factor activation and cytokine production[11], and subsequent identification of lipopolysaccharide (LPS), an outer membrane component distinctly found on Gram-negative bacteria, as the ligand for TLR4[12–15]. LPS is composed of three parts: the O-antigen, the core oligosaccharide, and lipid A, the major component contributing to the immunostimulatory effects of LPS[16]. To date, 10 TLRs in humans and 12 TLRs in mice have been identified, each specializing in detection of a unique PAMP derived from microorganisms ranging from bacteria, virus and fungi. TLR1, 2, 4, 5, and 6 are expressed on the cell surface to sense extracellular threats such as bacteria and fungi, while TLR3, 7, 8, and 9 are restricted to the endocytic compartments to detect viral and bacterial nucleic acid[1]. Although TLRs were the first discovered and most extensively studied family of pathogen receptors, other PRR families nevertheless play non-redundant roles in pathogen recognition and innate immune activation. RIG-I-like receptors (RLRs) are specialized to detect cytosolic RNA from cells infected with RNA viruses, while NOD-like receptors (NLRs) are equipped to detect a wide range of cytoplasmic PAMPs[3]. The PAMPs recognized by various

PRRs are summarized in Table 1-1[1]. Together, the diverse collection of PRRs utilized by innate immune cells to counteract the wide range of microorganisms infecting host cells and tissue demonstrate the mechanisms by which the innate system distinguishes self from foreign danger. Crucial to establishing immunity against pathogen is the generation of a suitable immune response, as defense against an extracellular microorganism requires a response different from events elicited in response to viral infection. Therefore, insight into events downstream of PRR activation is critical to advance our understanding of how the innate immune system coordinates a specific response program that leads to clearance of pathogen and long lasting immunity.

A2. TLR4 Signal Transduction

An essential role of the innate immune system is not only to detect foreign insults but to effectively transmit instructive signals from the pathogen receptor involved to the nucleus of the responding cell in order to upregulate transcription of particular inflammatory genes encoding proteins directing an immune response specific for the PRR activated. This requires coordination of receptors with molecules capable of triggering signaling cascades, resulting in activation of transcription factors that induce expression of distinct inflammatory genes.

Specificity of the innate immune response is partially achieved by the adaptor molecules recruited to the intracellular tails of PRRs such as TLRs. TLRs are membrane-spanning receptors characterized by their N-terminal leucine-rich repeats (LRRs) and cytoplasmic Toll/IL-1R (TIR) homology domain[17]. The TIR domain is homologous to the cytoplasmic domain of the IL-1R and is capable of eliciting an immune response through homotypic interactions with other TIR domain-containing signaling adaptors. There are currently six signaling adaptors that contain TIR domains and theoretically able to interact with TLRs: MyD88, TIRAP, TIR domain-containing adaptor inducing IFN-β (TRIF or TICAM-1), TRAM, SARM, and BCAP[2,18]. MyD88

associates with all TLRs except TLR3, and TIRAP is required to bridge both TLR4 and TLR2 to MyD88. On the other hand, TRIF is recruited to the TIR domain of TLR3 and TLR4, with TLR4 additionally requiring TRAM for TRIF-mediated signaling[18]. Thus, TLR4 is the only TLR member known to activate both MyD88-dependent and TRIF-dependent signaling cascades. SARM and BCAP adaptor molecules are more recently discovered and less well understood. Although human SARM molecules have an inhibitory role in TRIF-signaling, SARM expression is limited to the brain and not essential for macrophage function in mice[1,18,19]. Less is known about BCAP, but recent studies have hypothesized an immunosuppressive role of BCAP by linking TLRs to PI3K activation[20]. Still, it remains unclear which TLRs associate with SARM and BCAP. Nevertheless, the classical TIR domain-containing adaptors are central to the TLR response, and specific signals emanating from the activated TLR are dictated in part by the adaptors recruited.

TLR4, the prototypic TLR, is unique in its ability to sequentially activate the distinct MyD88-dependent and TRIF-dependent signaling cascades to trigger a specific immune response (Figure 1-1). Importantly, while initiation of the MyD88-dependent signaling pathway occurs at the plasma membrane and culminates in the transcription of pro-inflammatory type genes, activation of the TRIF-dependent signaling pathway takes place from endosomes and leads to expression and synthesis of type I interferons (IFNs), critical regulators of the antiviral response[21–23]. Because TLR4 is unable to bind to LPS alone, LPS recognition occurs with the aid of accessory proteins and receptors such as LPS binding protein (LBP), CD14, and MD-2[18]. As a result of LPS detection, TLR4 oligomerizes and rapidly undergoes conformational changes to recruit bridging adaptor TIRAP to the receptor at the plasma membrane. This in turn recruits MyD88 to the signaling apparatus. Death domains (DD) on MyD88 allow for homotypic interactions with other DD-containing signaling molecules such as IL-1R-associated-kinases-4 (IRAK-4), one of four members of the IRAK family of kinases that play roles in early signal

transduction events in response to pathogen recognition[2]. IRAK4 subsequently interacts with TNFR-associated factor 6 (TRAF6), an E3 protein ubiquitin ligase that auto-ubiquitinates as well as ubiquitinates TGF-β-activated kinase 1 (TAK-1). TAK-1 then phosphorylates IkB kinase-β (IKK-β) and mitogen-activated kinase kinase 6 (MAPKK6), ultimately activating the NF-κB and MAPK signaling cascades, respectively. Soon after LPS triggering of TLR4 and MyD88-dependent signaling events take place, TLR4 is endocytosed in a dynamin GTPase-mediated event that requires CD14 as well as the TRAM adaptor[21,22]. TRAM is responsible for shuttling of TLR4 from the plasma membrane to the early endosomes, allowing TRIF-dependent signaling events to occur. Activated TRIF associates with both TRAF6 and receptor-interacting protein 1 (RIP1), resulting in late-phase activation of the NF-κB signaling pathway[18]. Importantly, activated TRIF also interacts with TRAF3, which in turn associates with TANK binding kinase (TBK1) and IKK*i,* two critical kinases mediating activation of the interferon regulatory factor 3 (IRF3) transcription factor to regulate IFN-β gene expression[1,19]. Therefore, while TLR4 utilizes both MyD88-dependent and TRIF-dependent signaling cascades, the requirement of receptor endocytosis for TRIF-dependent signal transduction explains the delayed activity of TRIF-mediated signals. MyD88 signals from the plasma membrane and results in activation of NF-κB and MAPK signaling cascades, while TRIF signals from the endosome to initiate the late-phase of the NF-κB and MAPK signaling cascades and uniquely activates IRF3, resulting in IFN-β gene expression and antiviral activity. The signaling events downstream of TLRs as well as the transcription factors that they activate are depicted in Figure 1-1[9]. Thus, differential utilization of adaptor molecules downstream of pathogen recognition contributes to the specificity of the innate immune response.

A3. Transcriptional Regulation of the TLR4 Response

Precise control of gene expression by transcription factors activated in response to a microbe or other foreign danger is essential to maintain homeostasis and immunity. As discussed in the previous section, pathogen receptor detection of foreign dangers such as PAMPs initiates numerous signal transduction pathways, leading to the activation of transcription factors and other molecules that orchestrate gene expression to coordinate a response that is specific to the stimulus detected. Induction of the inflammatory and antiviral programs during the innate immune response is due to the transcriptional induction of genes encoding for cytokines such as TNF-α, IL-1β, and IFN-β. Controlled expression of these genes are essential for effective defense against pathogens or host insults, as well as for prevention of unnecessary immune responses that can damage host tissues. Therefore, efforts have been made to understand the role of these transcription factors in effective immune function. Because TLR4 is considered the prototypical Toll-like receptor and critically relies on both MyD88 and TRIF signaling events for an effective immune response, signaling cascades emanating from this receptor will be discussed with greater emphasis.

A3.1 NF-κB

Discovered over 25 years ago in B cells[24], perhaps the most crucial and well-studied transcription factor involved in the regulation of the immune response is NF-κB, and its capacity to be activated by both MyD88-dependent and TRIF-dependent signaling pathways bolsters its importance for the immune response[25]. Moreover, NF-κB is central in the response to numerous other stimuli such as TNF-α and IL-1[26,27], in response to T and B cell activation[24,28,29], and additionally influences expression of genes outside of the immune system in other cell types[30–32]. The breadth of this transcription factor family suggests a complexity underlying its mechanisms of activation to result in a specific transcriptional output.

Classically, the initiation of the NF-κB signaling cascade begins with phosphorylation of the IκB kinase (IKK) complex, composed of IKKα, IKKβ, and NEMO, by upstream kinases such as TAK-1[33,34]. Phosphorylation and activation of the IKK complex result in IκB protein phosphorylation, ubiquitination, and ultimately degradation. IκBs such as IκBα, IκBβ, and IκBε are inhibitory proteins that sequester NF-κB dimers in the cytoplasm, rendering them inactive[35]. Degradation of the IκBs, predominantly IκBα during the TLR response, allows translocation of NF-κB dimers into the nucleus and subsequent binding to κB sites at regulatory elements of genes. NF-κB comprises five family members: p65 (RelA), RelB, c-Rel, p50 (p105), and p52 (p100)[36]. All members of this family are characterized by their N-terminal Rel homology domain (RHD) that is responsible for interaction with IκBs, DNA binding, and dimerization. Because of their ability to form both homo- and heterodimers, NF-κB can dimerize in 15 possible combinations. The transcriptional activation domain (TAD) that confers the ability to induce gene expression is present only on the C-terminal regions of p65, RelB, and c-Rel. Therefore, unless p50 or p52 heterodimerize with a TAD-containing NF-κB member, they are transcriptionally inactive but retain the ability to bind DNA and may function as transcriptional repressors.

Additional modes of NF-κB regulation occur post-translationally and through DNA binding preferences. Phosphorylation of p65 at Ser276 is mediated by protein kinase A (PKA) to promote interaction with coactivators CREB binding protein (CBP) and p300[37]. The two cofactors subsequently mediate the acetylation of p65 at Lys310 to enhance transcription of NF-κB target genes[38]. Notably, loss of phosphorylation of p65 at Ser276 only impaired subsets of NF-κB-responsive genes, suggesting differences in the requirements of these genes for activation[33]. Other kinases such as mitogen- and stress- activated protein kinases (MSK1 and MSK2) also phosphorylate p65 at Ser276[39]. Notably, MSK1 and MSK2 are activated by MAPK signaling pathways p38 and ERK, suggesting crosstalk between NF-κB and MAPK[33].

Furthermore, recent studies have elucidated differences in dimer-DNA specific binding preferences and uncovered nonconsensus κB motifs[40]. Thus, selective gene activation by NF-κB is not only limited to cell type specificity or dimer composition, but is contributed to by post-translational modifications, coactivators involved, as well as dimer binding preferences to the regulatory sequences of target genes. The layer of mechanisms by which NF-κB can regulate gene expression emphasizes the complexity of mechanisms underlying specificity of the immune response.

A3.2 IRF3

Another major signaling pathway initiated due to LPS sensing by TLR4 is one that results in the activation and translocation of the IRF3 transcription factor. In addition to their role in hematopoietic cell development and promoting adaptive immune responses, members of the IRF family of transcription factors such as IRF3 play an important role in response to pathogen sensing and are involved in diverse immune functions. IRFs are indispensable for the induction of genes encoding the type I interferons (IFNs) and are central to establish antiviral immunity[41]. TLR4 is only one of two TLRs capable of activating the IRF3 signaling pathway, thus underscoring the unique and selective regulatory role of IRF3 in response to TLR4 signaling.

The transcription factor family of interferon regulatory factors (IRFs) comprises nine members, IRF1-9[41,42]. They are characterized by a conserved N-terminal 120 amino acid DNA binding domain and, with the exception of IRF1 and IRF2, a conserved C-terminal domain that is homologous to the C-terminal domain of the SMAD family of transcription factors. Although less is known about the precise functions of the C-terminal domain, it is thought to mediate protein interaction. Nevertheless, the DNA binding domain of IRF family members form a helix-turn-helix domain that is capable of recognizing a common DNA motif known as the interferon-stimulated response element (ISRE). ISREs are found in the regulatory regions of genes

encoding for type I IFNs as well as in the regulatory regions of numerous other genes involved in maintaining immunity[43].

Traditionally, IRFs have been studied in the context of viral infection. Even before the discovery of PRRs, it was observed that viral infections such as Newcastle disease virus (NDV; an ssRNA virus) and encephalomyocarditis virus (EMCV; an ssRNA virus) resulted in type I IFN expression[44,45]. This observation could be seen across all cell types, occurring through what we now know as the cytosolic PRRs discussed in the previous section. The first IRF discovered to function in type I IFN production was IRF1 through detection of viral RNAs by the cytosolic PRRs[46]. However, subsequent gene deletion studies determined that IRF1 is not essential for type I IFN production. This ultimately led to the discovery that IRF3 and IRF7 are largely responsible for type I IFN expression[45,47]. Although IRF3 and IRF7 are highly homologous, IRF3 remains constitutively expressed in macrophages. Inactive IRF3 resides in the cytoplasm, and is activated upon phosphorylation to allow either homo- or hetero- dimerization with IRF3 or IRF7, respectively, and translocation into the nucleus[48–50]. On the other hand, IRF7 is minimally expressed in macrophages during steady state and its transcription is induced by type I IFN signaling[51,52]. Furthermore, biochemical studies have identified dimer-specific binding preferences to DNA[43]. Thus, although IRF3 and IRF7 both play critical roles during the innate immune response and offer a degree of redundancy, differences in their regulation suggest the selective roles they play in immunity.

In the context of TLR4, activation of IRF3 is mediated through the TRIF-dependent module of the TLR4 signaling network[53], as discussed above. Upon activation and endocytosis of TLR4, the TRIF signaling adapter is activated through the bridging adapter TRAM[21,23]. This ultimately results in TANK binding kinase (TBK1) and IKK*i* -mediated phosphorylation of IRF3 at carboxy-terminal serine residues 385 and 386, and between residues 396 and 405 at a serine and threonine cluster[48,50]. Phosphorylated IRF3 homodimerizes or heterodimerizes with IRF7

and translocates into the nucleus where IRF3 dimers can associate with co-activators such as CBP or p300 to bind target genes and initiate transcription[48,50,54,55]. TLR4 is one of only two TLR family members, the other being TLR3, capable of activating IRF3 to regulate transcription of IFN-β[53].Thus, the IRF3 signaling pathway represents a unique and specialized mechanism utilized by innate immune cells to mount a specific immune response.

A3.3 MAPK

The mitogen-activated protein kinases (MAPK) are a family of signal transducing enzymes that are involved in the cellular response to a diverse range of stimuli including oxidative stress, heat shock, inflammatory stimuli, and pathogens. The numerous members within this family of kinases, as well as the diversity of roles it plays in cellular regulation such as cell proliferation, survival, and programmed death indicate that MAPK signaling cascades are essential to cell integrity and function. The following section focuses on the role of MAPKs in transcriptional regulation and its contribution to the innate immune response.

The first MAP kinase discovered was extracellular signal regulated protein kinase 1 (ERK1) followed by a closely related family member ERK2[56,57]. The two kinases were found to play critical roles in growth factor signaling cascades, hence the term mitogen-activated. Of the 14 MAPKs that have been described, ERK1, ERK2, p38α, Jun N-terminal kinase 1 (JNK1), and JNK2 have been implicated to be involved in mediating innate immune responses[58]. Each of these three MAPK groups, ERK, p38, and JNK, play a critical role in transducing signals in their respective signaling cascades. Generally, MAPK signaling cascades occur with a series of at least three kinases: a MAPK kinase kinase (MAP3K) that phosphorylates and activates a MAPK kinase (MAP2K), which in turn phosphorylates and activates a MAPK. The activated MAPKs then have the capability to either directly activate transcription factors or activate downstream kinases that regulate transcription factor activity. MAPK substrates contain a consensus motif

for phosphorylation, [Ser/Thr]-Pro[56]. Each MAPK has specific MAP2Ks that regulate their activities: MEK1/2 for ERK1/2, MKK3/6 for p38, and MKK 4/7 for JNK1/2[59]. On the other hand, the MAP3Ks involved in activation of the MAP2Ks are more complex and diverse. For example, although the MAP3K RAF is critical for activation of MEK1/2 following growth factor and antigen receptor stimulation[60], the tumor progression locus 2 (TPL2) MAP3K has been shown to be critical in MEK1/2 activation for the ERK1/2 signaling pathway during innate immune responses[61,62]. Additionally, at least a dozen MAP3Ks, including MEKK1-4, TAO1/2, and MLK2/3, have been shown to have the capacity to activate both the p38 and JNK signaling pathways[59], and specificity is partially dependent on the stimulus and cell type. Specificity of the MAP3Ks and MAP2Ks involved has also been attributed to scaffolding proteins that link different members of the cascade together[63,64]. Thus, although MAPKs can be activated by an array of upstream regulators, the diverse range of processes regulated by MAPK signaling pathways suggests mechanisms are in place to achieve specificity of the response.

In the context of TLR signaling, TAK-1 is activated by TRAF6 during proximal TLR signaling events, as discussed above. TAK-1 acts as a MAP3K for both the p38 and JNK signaling pathways[65–67]. Additionally, TAK-1 indirectly activates MAP3K TPL2 to initiate the ERK signaling pathway[68,69]. In resting macrophages, TPL2 is held inactive in a complex with NF-κB member p105. Once TAK-1 is activated, it phosphorylates IκB kinase 2 (IKK2) to initiate proteolysis of p105, resulting in the release of TPL2 and subsequent activation of the ERK signaling pathway. Therefore, TAK-1 plays an indispensible role in MAPK activation and initiating innate immune responses.

Once activated by immune stimuli, MAPKs can both directly and indirectly affect the transcription of genes. Activated JNKs enter the nucleus to phosphorylate and increase the activity of transcription factors c-Jun and ATF2[70]. Additionally, the MAPK ERK1/2

phosphorylates and activates the Ets family of transcription factors such as ELK1 to form ternary complex factors (TCFs) with the constitutively bound and expressed serum response factor (SRF) in the regulatory region of genes such as *Fos*[71]. Fos, a transcriptional regulator, then dimerizes with active c-Jun members to form the AP-1 transcription factor complex. ERK1/2 and p38 can phosphorylate effector protein kinases such as MK2/3, MSK1/2, and RSKs that play diverse roles in regulating cellular processes[59]. In particular, MSK1 and MSK2 are nuclear kinases that phosphorylate and activate the transcription factor CREB1 to regulate its target genes[72]. Thus, the MAPK signal transduction pathways activate numerous transcription factors and effector kinases to play both direct and indirect roles in the innate immune response.

Taken together, the precise transcriptional output in response to an external signal is determined by the activation of specific combinations of transcription factors. These transcription factors must be able to associate with DNA at regulatory elements in a sequence-specific manner and recruit components of the transcription machinery, all in the context of chromatin. The following section will discuss chromatin biology and the constituents of the general transcription machinery, and their influence on the innate immune response.

B. Chromatin Influence and Transcriptional Regulation at the Promoter

Although it is well established that signaling pathways and the transcription factors they activate play an indispensable role during the innate immune response, chromatin architecture represents an additional barrier of selective regulation contributing to gene expression because the basal transcription machinery is unable to associate with nucleosomal DNA. Furthermore, the promoters of genes are dense with sequence-specific binding sites and other conserved elements that recruit transcription factors and the general transcription machinery. Together, these components provide an additional complex layer of regulation in the activation of genes during the innate immune response.

B1. Chromatin Biology

Chromatin is a molecular complex composed of DNA wrapped around histones to form nucleosomes. Each nucleosome core is composed of 147 base pairs of DNA wrapped around a histone octamer containing two copies of each histone protein: H2A, H2B, H3, and H4. Linker histones such as H1 and H5 and histone variants such as H2A.Z, H2A.X, and H3.3 contribute to the integrity of chromatin structures[73]. The tightly wrapped nature of DNA around a nucleosome therefore prevents access to these genomic regions by most *trans*-acting elements. For example, nucleosomes occluding regulatory regions such as promoters prevent binding of sequence-specific transcription factors that do not have the capability to bind inaccessible nucleosomal DNA. However, core histone tails protruding away from the nucleosome can be post-translationally modified to relax nucleosome structures, provide docking sites for other recruited proteins, or enforce higher order chromatin structures. Furthermore, chromatin remodeling enzymes can displace nucleosomes from its genomic position in an ATP-dependent manner to alter accessibility to DNA. Therefore, this macromolecular complex not only allows for the compaction of DNA, but modulation of nucleosome structures provides an additional layer of regulation to ensure appropriate gene expression[73].

Initial DNase I hypersensitivity studies to examine nucleosome occupancy at *cis*-regulatory regions of inflammatory genes revealed stimulus-dependent alterations in promoter accessibility, providing a link between changes in chromatin structure and inducible transcription of pro-inflammatory genes[74–76]. However, the initial studies did not uncover whether chromatin structure acts as a general barrier of transcriptional activation or contributes to selective induction of inflammatory genes. Thus, subsequent efforts made to clarify the role of chromatin in selective gene activation during the innate immune response will be discussed.

B2. Chromatin Remodeling Factors

There are two general classes of chromatin remodeling factors that play roles in transcriptional changes. The first are ATP-dependent nucleosome remodeling complexes such as SWI/SNF, ISWI, and Mi-2β[77]. These are multi-protein complexes that, through their catalytic subunits, use energy gained from ATP hydrolysis to loosen the interaction between DNA and histones. Binding does not occur in a sequence specific manner, but is recruited to gene promoters through protein interaction. The second class of chromatin modifiers implicated in transcriptional activation contains intrinsic acetyltransferase activity, including CREB binding protein (CBP), p300, and GCN5/PCAF[73]. Members of this class associate at promoters through interaction with DNA binding proteins such as transcription factors, and binding of these proteins result in histone acetylation and subsequent loosening of chromatin. Both classes of chromatin modifiers promote loosening of DNA from nucleosomes to allow the general transcription machinery to bind and increase gene expression, and the role of the ATP-dependent nucleosome remodeling complexes in the innate immune response will be further examined.

One of the first studies of chromatin contributions in inducible gene activation during an immune response connected ATP-dependent nucleosome remodeling complex SWI/SNF recruitment to chromatin decondensation during T cell activation[78]. The SWI/SNF chromatin remodeling complex, through its catalytic subunit Brg1 or Brm, uses energy gained from ATP hydrolysis to loosen the interaction between DNA and histones[79]. The weakened contact between DNA and histones results in either sliding or eviction of nucleosomes. This exposes regulatory regions of DNA, allowing transcription factors or components of the general transcription machinery to interact with DNA and drive expression of target genes.

Subsequent studies focused on nucleosome remodeling in the context of selective gene activation[77,80]. Loss-of-function studies on Brg1 and Brm of the SWI/SNF complex demonstrated

that some but not all genes require nucleosome remodeling by SWI/SNF for activation upon stimulation of macrophages with lipopolysaccharide (LPS). Importantly, genes induced with rapid kinetics tend to be SWI/SNF-independent while the SWI/SNF-dependent genes are activated with late kinetics in response to LPS. The rapidly activated SWI/SNF-independent genes exhibit an open and accessible chromatin structure in their promoters during both resting and activated states while the promoters of SWI/SNF-dependent genes are only accessible after LPS stimulation, indicating that LPS induces nucleosome remodeling of selective genes by the SWI/SNF complex. For example, these studies demonstrate that IRF3-dependent gene promoters are assembled into stable nucleosomes prior to activation[80]. Activation of these genes requires IRF3-facilitated SWI/SNF nucleosome remodeling, as LPS-induced promoter accessibility was blocked in IRF3-deficient macrophages. This suggests that IRF3 mediates inducible chromatin remodeling for a subset of SWI/SNF-dependent genes in response to LPS.

Thus, upon TLR4 activation by LPS, select genes are activated with delayed kinetics due to their requirement on the SWI/SNF complex to displace nucleosomes that prevent accessibility to the promoter. Once remodeled, sequence-specific transcription factors, RNA polymerase II, and other components of the transcription machinery are recruited to the regulatory region of these genes to initiate transcription. Therefore, in addition to the variable requirement of genes for signaling pathways downstream of TLR4 or sequence-specific transcription factors, the availability of *cis*-elements for some genes and not others provides an additional regulatory layer that shapes selectivity of the innate immune response.

B3. CpG Islands

Loss-of-function studies on Brg1 and Brm subunits of the SWI/SNF nucleosome remodeling complex identified a select subset of rapidly induced genes that do not require chromatin restructuring for activation, as discussed above[77,80]. Promoter accessibility studies

demonstrated that these genes are not occluded by nucleosomes either before or after stimulation with LPS, indicating that their regulatory regions may have unique properties that contribute to their rapid kinetics of activation[80]. This would therefore provide another regulatory mechanism by which the innate immune system selectively upregulates transcription of specific subsets of genes.

Importantly, an additional finding from these studies was the observation that many SWI/SNF-independent genes contain CpG islands (CGIs) in their promoters[80]. CGIs are long stretches of DNA found in vertebrate genomes with a high number of CpG dinucleotides, rich in G and C nucleotide composition, and are typically hypomethylated despite the more common occurrence of methylated CpG dinucleotides dotted throughout the genome[81]. Although 70% of transcribed genes are associated with CGIs in their promoters and therefore are closely linked to transcription, the functional significance of their presence is still being refined. These promoter-CGI-containing genes do not assemble into stable nucleosomes compared to promoters with low CpG content[80]. This supports the knowledge that properly spaced A and T nucleotides, which CGIs are not equipped with, provide the necessary structure for DNA to wrap around nucleosomes[73]. Furthermore, the SWI/SNF-independent genes that contain CGIs in their promoters had high basal levels of H3K4 trimethylation and acetylation, covalent histone modifications that mark actively transcribed genes[80]. Therefore, the finding that the majority of SWI/SNF-independent genes contain promoter CGIs supports the observation that they also do not form stable nucleosome structures, and indicate that these genes made be transcribed during the basal state.

During the innate immune response, it is likely that SWI/SNF-independent genes are activated with rapid kinetics relative to the SWI/SNF-dependent genes due to their open chromatin formation in the resting state (Figure 1-2)[82]. Support for this comes from the finding that the promoters of these genes contain CpG islands. The presence of CGIs prevents

formation of stable nucleosomes, thereby circumventing the need for the SWI/SNF nucleosome remodeling complex to restructure chromatin prior to activation. The open chromatin formation may additionally allow transcription factors and the general transcription machinery to assemble at the promoter prior to activation, as evidenced by active histone marks in the basal state. Therefore, the chromatin status in gene promoters prior to its activation contributes to its activation kinetics, and represents an additional layer of regulation underlying selective activation of the innate immune response.

B4. Promoters and the General Transcription Machinery

Chromatin architecture plays an integral role in regulation of gene expression, as described above. Nucleosomal DNA prevents transcription factor binding and erroneous transcription by occlusion of *cis*-regulatory regions such as promoters and enhancers. The binding of transcriptional activators and co-activators initiates histone modification, chromatin reorganization and ultimately leads to transcription initiation through the general transcription machinery. Because enhancers can be found kilobases and megabases upstream or downstream from the transcription start site (TSS)[73,83], it can be challenging to connect an enhancer to its target gene. On the other hand, promoters are well defined because they lie immediately proximal to the TSS and contain the elements necessary for the general transcription machinery to assemble[73]. Although both promoters and enhancers have been shown to be critical in facilitating transcriptional activation[83–85], the function of promoters and its associated factors have been well documented due to its close proximity to the TSS and will be discussed in further detail.

Promoters are *cis*-regulatory regions of DNA immediately upstream (5') from the transcriptional start site (TSS), and can be divided into several elements: the core, proximal, and distal promoter[73]. The core promoter is defined as the region between +35 and -35 relative

to the TSS, and contains all of the elements necessary for the general transcription machinery to bind and initiate transcription[86]. The proximal promoter generally lies 250 base pairs upstream from the core promoter and contains most of the binding sites for transcription factors, while the distal promoter lies further upstream from the proximal and generally contains fewer transcription factor binding sites.

The core promoter comprises several elements that facilitate binding of the general transcription machinery and pre-initiation complex (PIC) formation. These elements include, but are not limited to, the TATA box, initiator element (Inr), downstream promoter element (DPE), and B recognition element (BRE)[73,86]. Importantly, it is not necessary for core promoters to contain all of these elements, and some elements may only be found in some subsets of genes. For example, genes with CpG-island promoters tend to be depleted of TATA boxes and can initiate transcription from multiple sites[81,86]. In contrast, genes containing TATA boxes tend to have a rigid TSS 25 base pairs downstream from the TATA box[86]. The major role of the core promoter is to provide binding sites for the general transcription machinery to initiate transcription. This includes RNA Polymerase II (PolII) and the general transcription factors (GTFs): TFIID, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH. Notably, TATA binding protein (TBP) is a subunit of the TFIID complex along with the numerous TATA binding protein-associated factors (TAFs) that binds to the TATA box, while TFIIB can bind to the BRE[86].

Initiation of transcription begins with formation of the pre-initiation complex (PIC)[73]. At core promoters containing TATA boxes, binding of the TBP subunit of TFIID recruits subsequent binding of TFIIA and TFIIB to stabilize the complex. This recruits PolII and the remaining GTFs to the PIC. TFIIH, one of the last GTFs to associate with the pre-initiation complex, plays a critical role by facilitating promoter melting through energy gained from ATP hydrolysis. Importantly, this stage of transcription can be defined by phosphorylation of Ser-5 in the carboxy-terminal domain (CTD) of PolII. Phosphorylation at this site facilitates clearance away

from the promoter and binding of RNA capping enzymes that stabilize newly synthesized transcripts. As PolII escapes from the promoter, it releases from the transcription machinery. Just downstream of the TSS, PolII stalls in a paused state due to negative elongation factors DSIF and NELF. Relief from the paused and into a productive elongation state is dependent on the positive transcription elongation factor P-Tefb. This protein is responsible for phosphorylating DSIF and NELF, causing them to disassociate from PolII. Additionally, P-Tefb phosphorylates Ser- 2 of the CTD to serve as the final transition from paused polymerase to full elongation of the transcript.

In the context of the innate immune response, recent studies have implicated an additional layer of regulation at the level of promoter pausing and unpausing that is signal-dependent, resulting in inflammatory genes activated with diverse kinetics[87]. Therefore, it is clear that selectivity of gene activation during the innate immune response is not simply governed by the transcription factors that are activated in a stimulus-specific manner. Recent efforts have demonstrated the essential role that chromatin and the basal transcription machinery plays in achieving specificity of gene activation in response to an innate immune stimulus[80,87]. For example, if a stimulus does not induce the IRF3 transcription factor, the SWI/SNF complex will be unable to remodel chromatin at the promoters of the IRF3-dependent genes. This prevents the IRF3-dependent genes from being transcribed since they are assembled into stable nucleosomes, while the genes that do not require nucleosome remodeling and remain in an open chromatin formation may be activated. Taken together, it is essential to uncover regulatory mechanisms at the signaling, transcription factor, and chromatin levels to understand the wiring of regulatory networks that confer gene selective responses.

C. Coordinated Control of Gene Expression: Interferon-β

Transcriptional regulation of gene expression during infection is essential to establish an effective innate immune response to a pathogen. This response is a double-edged sword: on one hand, upregulation of antimicrobial genes is necessary for defense against and resolution of infection; on the other hand, prolonged or excessive inflammation leads to development of host tissue damage, chronic inflammatory disorders, and pro-tumor microenvironments. Expression of innate immune responsive genes must therefore be tightly regulated. To achieve this, coordinated control between specific combinations of transcription factors and the sites at which they are recruited is necessary, in order to direct the basal transcriptional machinery to the promoter. One classic example of a gene under tight regulation is the gene encoding for Interferon-β (IFN-β), *IFNB1*.

The mechanisms of gene activation at the human *IFNB1* gene is perhaps the best characterized of all inducible inflammatory genes, due to its central role in antiviral immunity. Additionally, the multilayered contribution from numerous transcription factors and other regulatory proteins serves as a good model gene to study regulation by integrated signals. Termed the 'enhanceosome', this regulatory region of *IFNB1* is nearly 100% conserved across all mammalian genomes[88], suggesting a clear selective advantage throughout evolution for maintaining precise organization of the transcription factors involved. The enhanceosome spans a 55 base pair nucleosome-free region in the promoter, from -102 to -47 base pairs upstream of the transcription start site (TSS)[88,89]. Additionally, tightly clustered sequence-specific binding sites for ATF/c-Jun, interferon response factors IRF-3 and IRF-7, and NF-κB within the 55 base pair enhanceosome form a single composite binding element[88], suggesting that there is a high level of cooperative and combinatorial binding to achieve a response.

The *IFNB1* enhanceosome is divided into four positive regulatory domains (PRDs)[89,90], each containing a sequence-specific binding motif for one of the factors described above (Figure 1-3). Beginning at the 5' of the enhanceosome, ATF/c-Jun heterodimers bind to PRDIV[91]. ATF and c-Jun are members of the basic region leucine zipper (bZIP) transcription factor family that heterodimerize and bind to cyclic AMP recognition elements (CRE) 5'-TGAGCTCA-3'[92]. Notably, the CRE found in the IFNB1 enhanceosome is uniquely asymmetrical due to the noncanonical 3' half-site, 5'-TGACATAG-3'[92]. Structural studies demonstrate that the asymmetry of the site is critical for the binding orientation of ATF/c-Jun, as well as for cooperativity between the ATF/c-Jun dimers and IRF dimers at PRDIII[88,91]. IRF3 and IRF7 homo- and heterodimers bind to PRDIII and PRDI[91,93,94], and are characterized by their amino-terminal DNA binding domains that recognize the IRF binding element/interferon stimulated response element (ISRE) 5'-AANNGAAA-3'. NF-κB, specifically the p50:RelA heterodimer, binds to PRDII at the 3' end of the enhanceosome[88,95]. The NF-κB family of transcription factors are characterized by their highly conserved Rel Homology Region (RHR), and the canonical heterodimer p50:RelA recognizes the κB motif 5'-GGGAATTTCC-3'[40]. Although the enhanceosome itself is nucleosome free, two nucleosomes flank each end, with the 3' nucleosome lying over the TATA box to limit basal transcription[96].

Upon signal-dependent activation of these transcription factors, HMGA1a, an architectural protein that associates with the minor groove of DNA, is recruited to the enhancer to promote a favorable DNA conformation for enhanceosome assembly[97]. The 8 transcription factors, p50:RelA, two IRF dimers, and an ATF/c-Jun dimer, then cooperatively bind to the enhanceosome[88]. Strikingly, crystal structures of these transcription factors bound to the enhanceosome indicate that there are minimal protein-protein interactions between them[88,91]. This is unconventional because the majority of cooperative binding studies demonstrate that cooperativity is driven by protein-protein interactions[98]. This suggests that cooperativity is

largely dictated by the DNA sequence and architecture itself. Indeed, structural studies show that the asymmetry of the CRE allows IRF dimers to bind at the adjacent overlapping motif[91]. Together, these factors bind to the DNA to form a unique composite surface that first recruits GCN5/PCAF, followed by recruitment of co-activator CBP/p300 with high affinity to all transcription factors[96]. CBP/p300 binding considerably increases the potency of *IFNB1* activation by recruiting RNA polymerase II (PolII), components of the basal transcription machinery, and the ATP-dependent nucleosome remodeling complex SWI/SNF[99–101]. SWI/SNF remodels the nucleosome positioned over the TATA box, allowing TBP association and formation of the pre-initiation complex[102]. Biochemical studies have demonstrated that binding of single components to the enhanceosome is unable to promote robust transcription, and a functional response only occurs when all of the transcription factors are present[103]. Indeed, TLR3 and TLR4 are the only TLRs able to induce *IFNB1* because of their ability to signal through TRIF to activate IRF3[53]. Therefore, the combination of specific transcription factors, coactivators, as well as the highly conserved 55 base pair region all play indispensable roles in ensuring regulated expression of *IFNB1*.

The *IFNB1* enhanceosome is a classic example of combinatorial control by numerous transcription factors to promote gene expression. Due to the highly specialized role of this cytokine in initiating the antiviral response, expression needs to be tightly regulated. The precise positioning of sequence specific motifs and nucleosomes ensure that expression of this gene is tightly regulated to prevent unnecessary synthesis of IFN-β.

In summary, innate immunity is critical for rapid protection against invading pathogens and initiating adaptive immunity. Additionally, the innate immune system must be able to distinguish different pathogen types. Detection of a pathogen results in upregulation of a transcriptional program that is stimulus and cell type specific. This selectivity is presumably due

to an integration of signals from multiple signal transduction cascades, and combinatorial activity from different subsets of transcription factors that are engaged. A challenge in the field has been gaining a deeper and more precise understanding of the events that underlie specificity of the response. Although work has been done to address this, one drawback of these studies are the small numbers of genes that may be interrogated due to a limitation in quantitative methods. With the molecular biology field being revolutionized by highly quantitative high-throughput methods, we now have an opportunity to understand transcriptional networks at a genome-wide level. In chapter 2, I will describe genome-wide approaches to uncover how expression dynamics, promoter properties, and control by specific transcription factors converge to shape the regulatory framework for the TLR4-induced transcriptome. Chapter 3 of the dissertation will apply and interrogate the principles of TLR4-mediated gene activation described in chapter 2 in a physiological context. The findings in chapter 3 demonstrate that the framework can be applied to diverse settings to reveal what signaling pathways are relevant in various disease settings.

Figure Legends

Figure 1-1: The TLR Signaling Pathways

TLR signaling is initiated after engagement of the receptor by its ligand. In the case of TLR4, this initiates both the MyD88 and TRIF signaling cascades, ultimately leading to activation of transcription factors such as NF-κB, AP-1, and IRF3 and transcription of pro-inflammatory and antimicrobial genes.

Figure 1-2: Regulation of CpG-Island and Low CpG-Island Promoters at Inducible Genes

CpG-island and low CpG-island promoters exhibit distinct chromatin characteristics that contribute to selective induction of genes. CpG-island promoters have an open chromatin structure that allows for rapid activation and are promiscuously induced. Low CpG-island promoters are often assembled into nucleosomes that preclude basal transcription, and nucleosome remodeling by the SWI/SNF complex is required for their activation.

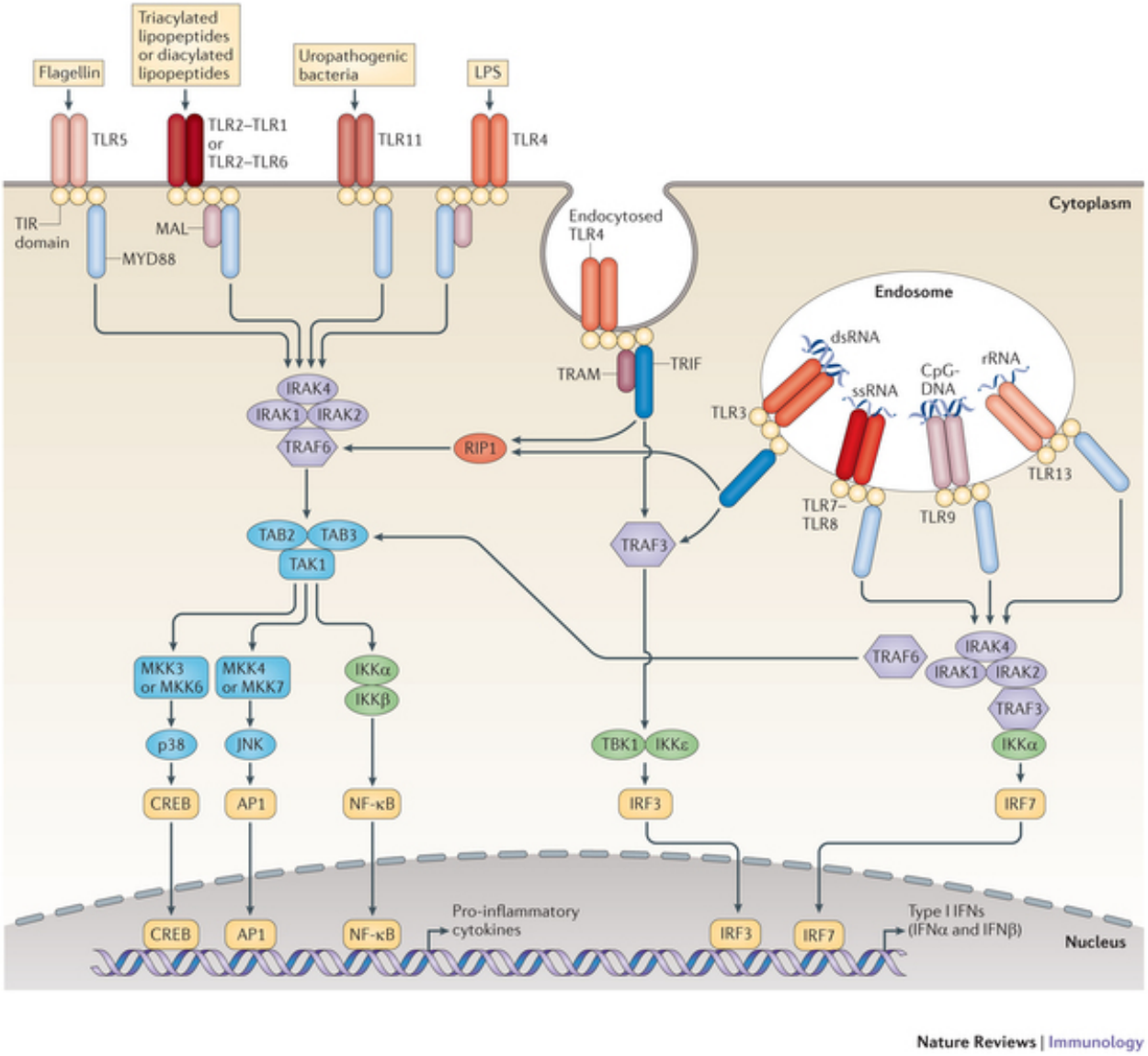
Figure 1-3: The IFN-β Enhanceosome

Type I interferon IFN-β is under tight regulation. The regulatory region of this gene lies between -102 and -47 base pairs relative to the transcription start site (TSS) and has been termed the 'enhanceosome'. Binding sites for ATF:c-Jun, IRF, and NF-κB are tightly packed into this region, and binding of all factors is necessary for cooperative activation of IFN-β.

Table 1-1: Pattern Recognition Receptors and their Ligands

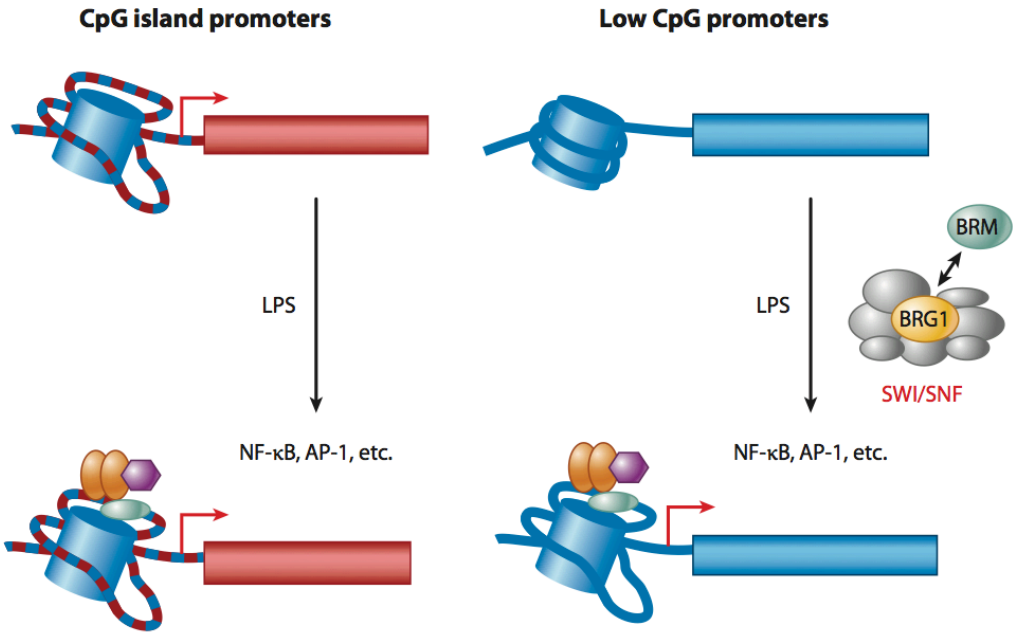
Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), NOD-like receptors (NLRs), and C-type-lectin receptors (CLRs) are broad classes of pattern recognition receptors (PRRs). RLRs and NLRs are found in the cytoplasm while CLRs are found on the plasma membrane. TLRs can be found either on the plasma membrane or endolysosomes. The ligands for PRRs are diverse, and are of bacterial, viral, fungal, or protozoan origin.

Figure 1-1: The TLR Signaling Pathways



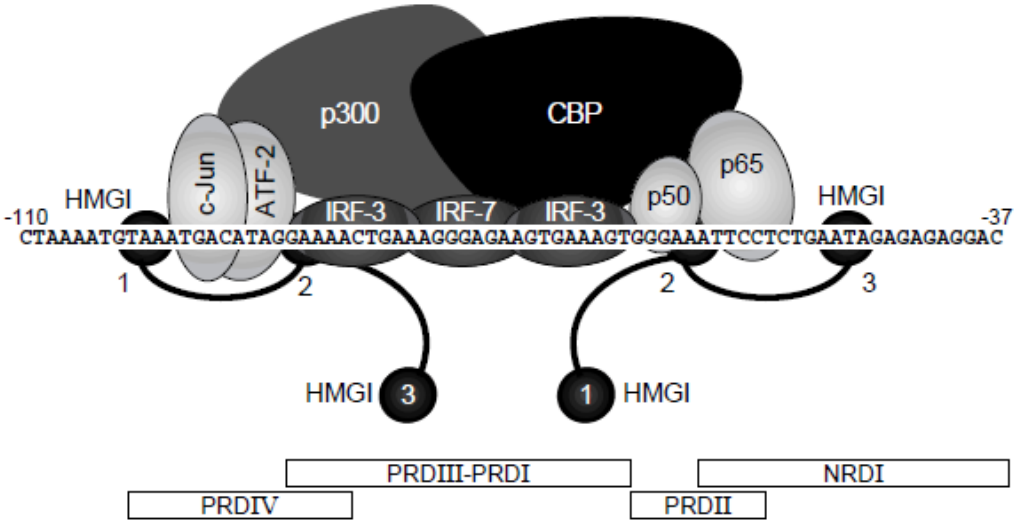
(O'Neill et al., 2013)

Figure 1-2: Regulation of CpG-Island and Low CpG-Island Promoters at Inducible Genes



(Smale et al., 2014)

Figure 1-3: The IFN-β Enhanceosome



(Maniatis et al., 1998)

Table 1-1: Pattern Recognition Receptors and their Ligands

Table 1. PRRs and Their Ligands

PRRs	Localization	Ligand	Origin of the Ligand
TLR			
TLR1	Plasma membrane	Triacyl lipoprotein	Bacteria
TLR2	Plasma membrane	Lipoprotein	Bacteria, viruses, parasites, self
TLR3	Endolysosome	dsRNA	Virus
TLR4	Plasma membrane	LPS	Bacteria, viruses, self
TLR5	Plasma membrane	Flagellin	Bacteria
TLR6	Plasma membrane	Diacyl lipoprotein	Bacteria, viruses
TLR7 (human TLR8)	Endolysosome	ssRNA	Virus, bacteria, self
TLR9	Endolysosome	CpG-DNA	Virus, bacteria, protozoa, self
TLR10	Endolysosome	Unknown	Unknown
TLR11	Plasma membrane	Profilin-like molecule	Protozoa
RLR			
RIG-I	Cytoplasm	Short dsRNA, 5′triphosphate dsRNA	RNA viruses, DNA virus
MDA5	Cytoplasm	Long dsRNA	RNA viruses (Picornaviridae)
LGP2	Cytoplasm	Unknown	RNA viruses
NLR			
NOD1	Cytoplasm	iE-DAP	Bacteria
NOD2	Cytoplasm	MDP	Bacteria
CLR			
Dectin-1	Plasma membrane	β-Glucan	Fungi
Dectin-2	Plasma membrane	β-Glucan	Fungi
MINCLE	Plasma membrane	SAP130	Self, fungi

(Takeuchi and Akira, 2010)

Works Cited

1. Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140,** 805–20 (2010).

2. Kawai, T. & Akira, S. TLR signaling. *Cell Death Differ.* **13,** 816–25 (2006).

3. Kumar, H., Kawai, T. & Akira, S. Pathogen recognition by the innate immune system. *Int. Rev. Immunol.* **30,** 16–34 (2011).

4. Rakoff-Nahoum, S. & Medzhitov, R. Toll-like receptors and cancer. *Nat. Rev. Cancer* **9,** 57–63 (2009).

5. Palm, N. W. & Medzhitov, R. Pattern recognition receptors and control of adaptive immunity. *Immunol. Rev.* **227,** 221–33 (2009).

6. Kawai, T. & Akira, S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **34,** 637–50 (2011).

7. Janeway, C. A. Approaching the Asymptote? Evolution and Revolution in Immunology. *Cold Spring Harb. Symp. Quant. Biol.* 1–3 (1989). at <http://symposium.cshlp.org/content/54/1.full.pdf>

8. Negishi, H. *et al.* Cross-interference of RLR and TLR signaling pathways modulates antibacterial T cell responses. *Nat. Immunol.* **13,** 659–66 (2012).

9. O'Neill, L. A. J., Golenbock, D. & Bowie, A. G. The history of Toll-like receptors - redefining innate immunity. *Nat. Rev. Immunol.* **13,** 453–60 (2013).

10. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.-M. & Hoffmann, J. A. The Dorsoventral Regulatory Gene Cassette spätzle/Toll/cactus Controls the Potent Antifungal Response in Drosophila Adults. *Cell* 973–983 (1996). doi:80172-5

11. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. . A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. **388,** 394–397 (1997).

12. Gerard, C. Bacterial infection. For whom the bell tolls. *Nature* **395,** 217, 219 (1998).

13. Poltorak, A. Defective LPS Signaling in C3H/HeJ and C57BL/10ScCr Mice: Mutations in Tlr4 Gene. *Science (80-.).* **282,** 2085–2088 (1998).

14. Qureshi, S. T. *et al.* Endotoxin-tolerant Mice Have Mutations in Toll-like Receptor 4 (Tlr4). *J. Exp. Med.* **189,** 615–625 (1999).

15. Hoshino, K. *et al.* Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J. Immunol.* **162,** 3749–52 (1999).

16. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458,** 1191–5 (2009).

17. Gay, N. J. & Gangloff, M. Structure and function of Toll receptors and their ligands. *Annu. Rev. Biochem.* **76,** 141–65 (2007).

18. Lu, Y.-C., Yeh, W.-C. & Ohashi, P. S. LPS/TLR4 signal transduction pathway. *Cytokine* **42,** 145–51 (2008).

19. Troutman, T. D., Bazan, J. F. & Pasare, C. Toll-like receptors, signaling adapters and regulation of the pro-inflammatory response by PI3K. *Cell Cycle* **11,** 3559–67 (2012).

20. Troutman, T. D. *et al.* Role for B-cell adapter for PI3K (BCAP) as a signaling adapter linking Toll-like receptors (TLRs) to serine/threonine kinases PI3K/Akt. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 273–8 (2012).

21. Kagan, J. C. *et al.* TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. *Nat. Immunol.* **9,** 361–8 (2008).

22. Zanoni, I. *et al.* CD14 controls the LPS-induced endocytosis of Toll-like receptor 4. *Cell* **147,** 868–80 (2011).

23. Fitzgerald, K. A. *et al.* LPS-TLR4 signaling to IRF-3/7 and NF-kappaB involves the toll adapters TRAM and TRIF. *J. Exp. Med.* **198,** 1043–55 (2003).

24. Sen, R. & Baltimore, D. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* **46,** 705–716 (1986).

25. Werner, S. L., Barken, D. & Hoffmann, A. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science* **309,** 1857–61 (2005).

26. Liou, H.-C. & Baltimore, D. Regulation of the NF-ηB/rel transcription factor and IηB inhibitor system. *Curr. Opin. Cell Biol.* **5,** 477–487 (1993).

27. Wang, D. & Baldwin, A. S. Activation of Nuclear Factor- B-dependent Transcription by Tumor Necrosis Factor- Is Mediated through Phosphorylation of RelA/p65 on Serine 529. *J. Biol. Chem.* **273,** 29411–29416 (1998).

28. Kahn-Perles, B., Lipcey, C., Lecine, P., Olive, D. & Imbert, J. Temporal and Subunit-specific Modulations of the Rel/NF- B Transcription Factors Through CD28 Costimulation. *J. Biol. Chem.* **272,** 21774–21783 (1997).

29. Pimentel-Muinos, F., Mazana, J. & Fresno, M. Regulation of interleukin-2 receptor alpha chain expression and nuclear factor.kappa B activation by protein kinase C in T lymphocytes. Autocrine role of tumor necrosis factor alpha. *J. Biol. Chem.* **269,** 24424–24429 (1994).

30. Kaltschmidt, C., Kaltschmidt, B., Neumann, H., Wekerle, H. & Baeuerle, P. A. Constitutive NF-kappa B activity in neurons. *Mol. Cell. Biol.* **14,** 3981–3992 (1994).

31. Barkett, M. & Gilmore, T. D. Control of apoptosis by Rel/NF-kappaB transcription factors. *Oncogene* **18,** 6910–24 (1999).

32. Krishnamoorthy, R. R. *et al.* Photo-oxidative Stress Down-modulates the Activity of Nuclear Factor- B via Involvement of Caspase-1, Leading to Apoptosis of Photoreceptor Cells. *J. Biol. Chem.* **274,** 3734–3743 (1999).

33. Hayden, M. S. & Ghosh, S. Shared principles in NF-kappaB signaling. *Cell* **132,** 344–62 (2008).

34. Hayden, M. S., West, A. P. & Ghosh, S. NF-kappaB and the immune response. *Oncogene* **25,** 6758–80 (2006).

35. Rao, P. *et al.* IkappaBbeta acts to inhibit and activate gene expression during the inflammatory response. *Nature* **466,** 1115–9 (2010).

36. Hayden, M. S. & Ghosh, S. Signaling to NF-kappaB. *Genes Dev.* **18,** 2195–224 (2004).

37. Zhong, H., Voll, R. E. & Ghosh, S. Phosphorylation of NF-κB p65 by PKA Stimulates Transcriptional Activity by Promoting a Novel Bivalent Interaction with the Coactivator CBP/p300. *Mol. Cell* **1,** 661–671 (1998).

38. Chen, L. -f. Acetylation of RelA at discrete sites regulates distinct nuclear functions of NF-kappaB. *EMBO J.* **21,** 6539–6548 (2002).

39. Vermeulen, L., De Wilde, G., Van Damme, P., Vanden Berghe, W. & Haegeman, G. Transcriptional activation of the NF-kappaB p65 subunit by mitogen- and stress-activated protein kinase-1 (MSK1). *EMBO J.* **22,** 1313–24 (2003).

40. Siggers, T. *et al.* Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. *Nat. Immunol.* **13,** 95–102 (2012).

41. Taniguchi, T., Ogasawara, K., Takaoka, A. & Tanaka, N. IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.* **19,** 623–55 (2001).

42. Honda, K. & Taniguchi, T. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat. Rev. Immunol.* **6,** 644–58 (2006).

43. Levy, D. E., Kessler, D. S., Pine, R., Reich, N. & Darnell, J. E. Interferon-induced nuclear factors that bind a shared promoter element correlate with positive and negative transcriptional control. *Genes Dev.* **2,** 383–393 (1988).

44. Matsuyama, T. *et al.* Targeted disruption of IRF-1 or IRF-2 results in abnormal type I IFN gene induction and aberrant lymphocyte development. *Cell* **75,** 83–97 (1993).

45. Sato, M. *et al.* Distinct and Essential Roles of Transcription Factors IRF-3 and IRF-7 in Response to Viruses for IFN-α/β Gene Induction. *Immunity* **13,** 539–548 (2000).

46. Miyamoto, M. *et al.* Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN-β gene regulatory elements. *Cell* **54,** 903–913 (1988).

47. Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434,** 772–7 (2005).

48. Lin, R., Heylbroeck, C., Pitha, P. M. & Hiscott, J. Virus-Dependent Phosphorylation of the IRF-3 Transcription Factor Regulates Nuclear Translocation, Transactivation Potential, and Proteasome-Mediated Degradation. *Mol. Cell. Biol.* **18,** 2986–2996 (1998).

49. Sato, M., Tanaka, N., Hata, N., Oda, E. & Taniguchi, T. Involvement of the IRF family transcription factor IRF-3 in virus-induced activation of the IFN-β gene. *FEBS Lett.* **425,** 112–116 (1998).

50. Yoneyama, M. *et al.* Direct triggering of the type I interferon system by virus infection: Activation of a transcription factor complex containing IRF-3 and CBP/p300. *EMBO J.* **17,** 1087–1095 (1998).

51. Sato, M. *et al.* Positive feedback regulation of type I IFN genes by the IFN-inducible transcription factor IRF-7. *FEBS Lett.* **441,** 106–110 (1998).

52. Marié, I., Durbin, J. E. & Levy, D. E. Differential viral induction of distinct interferon-alpha genes by positive feedback through interferon regulatory factor-7. *EMBO J.* **17,** 6660–6669 (1998).

53. Doyle, S. E. *et al.* IRF3 Mediates a TLR3/TLR4-Specific Antiviral Gene Program. *Immunity* **17,** 251–263 (2002).

54. Parekh, B. S. & Maniatis, T. Virus Infection Leads to Localized Hyperacetylation of Histones H3 and H4 at the IFN-β Promoter. *Mol. Cell* **3,** 125–129 (1999).

55. Lin, R., Genin, P., Mamane, Y. & Hiscott, J. Selective DNA Binding and Association with the CREB Binding Protein Coactivator Contribute to Differential Activation of Alpha/Beta Interferon Genes by Interferon Regulatory Factors 3 and 7. *Mol. Cell. Biol.* **20,** 6342–6353 (2000).

56. Pearson, G. *et al.* Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr. Rev.* **22,** 153–83 (2001).

57. Boulton, T. G. *et al.* ERKs: A family of protein-serine/threonine kinases that are activated and tyrosine phosphorylated in response to insulin and NGF. *Cell* **65,** 663–675 (1991).

58. Dong, C., Davis, R. J. & Flavell, R. A. MAP kinases in the immune response. *Annu. Rev. Immunol.* **20,** 55–72 (2002).

59. Arthur, J. S. C. & Ley, S. C. Mitogen-activated protein kinases in innate immunity. *Nat. Rev. Immunol.* **13,** 679–92 (2013).

60. Gupta, S., Weiss, A., Kumar, G., Wang, S. & Nel, A. The T-cell antigen receptor utilizes Lck, Raf-1, and MEK-1 for activating mitogen-activated protein kinase. Evidence for the existence of a second protein kinase C-dependent pathway in an Lck-negative Jurkat cell mutant. *J. Biol. Chem.* **269,** 17349–17357 (1994).

61. Risco, A. *et al.* p38γ and p38δ kinases regulate the Toll-like receptor 4 (TLR4)-induced cytokine production by controlling ERK1/2 protein kinase pathway activation. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 11200–5 (2012).

62. Yang, H.-T. *et al.* Coordinate regulation of TPL-2 and NF-κB signaling in macrophages by NF-κB1 p105. *Mol. Cell. Biol.* **32,** 3438–51 (2012).

63. Zou, H. *et al.* Differential requirement of MKK4 and MKK7 in JNK activation by distinct scaffold proteins. *FEBS Lett.* **581,** 196–202 (2007).

64. L, B. Mechanisms of MAPK signalling specificity. (2006). at <http://www.biochemsoctrans.org/bst/034/0837/bst0340837.htm>

65. Ajibade, A. A. *et al.* TAK1 negatively regulates NF-κB and p38 MAP kinase activation in Gr-1+CD11b+ neutrophils. *Immunity* **36,** 43–54 (2012).

66. Sato, S. *et al.* Essential function for the kinase TAK1 in innate and adaptive immune responses. *Nat. Immunol.* **6,** 1087–95 (2005).

67. Shim, J.-H. *et al.* TAK1, but not TAB1 or TAB2, plays an essential role in multiple signaling pathways in vivo. *Genes Dev.* **19,** 2668–81 (2005).

68. Robinson, M. J., Beinke, S., Kouroumalis, A., Tsichlis, P. N. & Ley, S. C. Phosphorylation of TPL-2 on serine 400 is essential for lipopolysaccharide activation of extracellular signal-regulated kinase in macrophages. *Mol. Cell. Biol.* **27,** 7355–64 (2007).

69. Beinke, S., Robinson, M. J., Hugunin, M. & Ley, S. C. Lipopolysaccharide activation of the TPL-2/MEK/extracellular signal-regulated kinase mitogen-activated protein kinase

cascade is regulated by IkappaB kinase-induced proteolysis of NF-kappaB1 p105. *Mol. Cell. Biol.* **24,** 9658–67 (2004).

70. Mendelson, K. G., Contois, L.-R., Tevosian, S. G., Davis, R. J. & Paulson, K. E. Independent regulation of JNK/p38 mitogen-activated protein kinases by metabolic oxidative stress in the liver. *Proc. Natl. Acad. Sci.* **93,** 12908–12913 (1996).

71. Marais, R., Wynne, J. & Treisman, R. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell* **73,** 381–393 (1993).

72. Ananieva, O. *et al.* The kinases MSK1 and MSK2 act as negative regulators of Toll-like receptor signaling. *Nat. Immunol.* **9,** 1028–36 (2008).

73. Carey, M., Petersen, C. & Smale, S. T. *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*. (CSH Laboratory Press, 2008).

74. Ward, S. B. *et al.* Chromatin remodeling of the interleukin-2 gene: Distinct alterations in the proximal versus distal enhancer regions. *Nucleic Acids Res.* **26,** 2923–2934 (1998).

75. Siebenlist, U. *et al.* Promoter region of interleukin-2 gene undergoes chromatin structure changes and confers inducibility on chloramphenicol acetyltransferase gene during activation of T cells. *Mol. Cell. Biol.* **6,** 3042–3049 (1986).

76. Agarwal, S. & Rao, A. Modulation of Chromatin Structure Regulates Cytokine Gene Expression during T Cell Differentiation. *Immunity* **9,** 765–775 (1998).

77. Ramirez-Carrozzi, V. R. *et al.* Selective and antagonistic functions of SWI/SNF and Mi-2beta nucleosome remodeling complexes during an inflammatory response. *Genes Dev.* **20,** 282–96 (2006).

78. Zhao, K. *et al.* Rapid and Phosphoinositol-Dependent Binding of the SWI/SNF-like BAF Complex to Chromatin after T Lymphocyte Receptor Signaling. *Cell* **95,** 625–636 (1998).

79. Peterson, C. L. & Workman, J. L. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr. Opin. Genet. Dev.* **10,** 187–192 (2000).

80. Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138,** 114–28 (2009).

81. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–22 (2011).

82. Smale, S. T., Tarakhovsky, A. & Natoli, G. Chromatin contributions to the regulation of innate immunity. *Annu. Rev. Immunol.* **32,** 489–511 (2014).

83. Ghisletti, S. *et al.* Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32,** 317–28 (2010).

84. De Santa, F. *et al.* A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS Biol.* **8,** e1000384 (2010).

85. Sullivan, A. L. *et al.* Serum response factor utilizes distinct promoter- and enhancer-based mechanisms to regulate cytoskeletal gene expression in macrophages. *Mol. Cell. Biol.* **31,** 861–75 (2011).

86. Butler, J. E. F. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16,** 2583–92 (2002).

87. Foster, S. L. & Medzhitov, R. Gene-specific control of the TLR-induced inflammatory response. *Clin. Immunol.* **130,** 7–15 (2009).

88. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129,** 1111–23 (2007).

89. Thanos, D. & Maniatis, T. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell* **83,** 1091–1100 (1995).

90. Maniatis, T. *et al.* Structure and Function of the Interferon- Enhanceosome. *Cold Spring Harb. Symp. Quant. Biol.* **63,** 609–620 (1998).

91. Panne, D., Maniatis, T. & Harrison, S. C. Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J.* **23,** 4384–93 (2004).

92. Falvo, J. V., Parekh, B. S., Lin, C. H., Fraenkel, E. & Maniatis, T. Assembly of a Functional Beta Interferon Enhanceosome Is Dependent on ATF-2-c-jun Heterodimer Orientation. *Mol. Cell. Biol.* **20,** 4814–4825 (2000).

93. Chen, W. *et al.* Insights into interferon regulatory factor activation from the crystal structure of dimeric IRF5. *Nat. Struct. Mol. Biol.* **15,** 1213–20 (2008).

94. Escalante, C. R., Nistal-Villán, E., Shen, L., García-Sastre, A. & Aggarwal, A. K. Structure of IRF-3 bound to the PRDIII-I regulatory element of the human interferon-beta enhancer. *Mol. Cell* **26,** 703–16 (2007).

95. Berkowitz, B., Huang, D.-B., Chen-Park, F. E., Sigler, P. B. & Ghosh, G. The x-ray crystal structure of the NF-kappa B p50.p65 heterodimer bound to the interferon beta -kappa B site. *J. Biol. Chem.* **277,** 24694–700 (2002).

96. Agalioti, T. *et al.* Ordered Recruitment of Chromatin Modifying and General Transcription Factors to the IFN-β Promoter. *Cell* **103,** 667–678 (2000).

97. Thanos, D., Du, W. & Maniatis, T. The High Mobility Group Protein HMG I(Y) Is an Essential Structural Component of a Virus-inducible Enhancer Complex. *Cold Spring Harb. Symp. Quant. Biol.* **58,** 73–81 (1993).

98. Merika, M. & Thanos, D. Enhanceosomes. *Curr. Opin. Genet. Dev.* **11,** 205–208 (2001).

99. Merika, M., Williams, A. J., Chen, G., Collins, T. & Thanos, D. Recruitment of CBP/p300 by the IFNβ Enhanceosome Is Required for Synergistic Activation of Transcription. *Mol. Cell* **1,** 277–287 (1998).

100. Kim, T. K., Kim, T. H. & Maniatis, T. Efficient recruitment of TFIIB and CBP-RNA polymerase II holoenzyme by an interferon-beta enhanceosome in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 12191–6 (1998).

101. Yie, J., Senger, K. & Thanos, D. Mechanism by which the IFN-beta enhanceosome activates transcription. *Proc. Natl. Acad. Sci.* **96,** 13108–13113 (1999).

102. Lomvardas, S. & Thanos, D. Nucleosome Sliding via TBP DNA Binding In Vivo. *Cell* **106,** 685–696 (2001).

103. Wathelet, M. Virus Infection Induces the Assembly of Coordinately Activated Transcription Factors on the IFN-? Enhancer In Vivo. *Mol. Cell* **1,** 507–518 (1998).

CHAPTER 2

Quantitative Dissection of a Stimulus-Induced Transcriptional Cascade Reveals Common and Gene-Specific Regulatory Strategies

Abstract

Much has been learned about the regulation of stimulus-induced transcriptional cascades from large-scale systems analyses of hundreds or thousands of genes that exceed a minimal induction threshold and are grouped into large co-expression cluster. For this study, we hypothesized that new insights into the regulatory logic would emerge from an approach that separates strongly and weakly induced genes and relies heavily of quantitative aspects of high-throughput data sets. To this end, we examined 226 genes whose primary transcripts are induced by a relatively large magnitude in mouse macrophages stimulated with lipid A. Our data suggest that these genes are regulated by mechanisms that generally differ from those used to regulate weakly induced genes. By combining quantitative consideration of induction magnitudes and kinetics with similarly quantitative analysis of loss-of-function, ChIP-seq, and binding motif data sets, we obtained insights into the relationships between NF–κB binding motifs and in vivo NF–κB binding and function. The results further suggest that the transcription factors IRF3 and SRF are dedicated to the regulation of surprisingly small numbers of strongly induced genes, with considerable gene-specific variation that reveals the extent to which each inducible gene is uniquely regulated. Together, our results reveal the value of focused analyses of limited numbers of genes for a mechanistic understanding of regulatory cascades.

Introduction

The molecular biology revolution of the mid-1970s was followed by a period of approximately 20 years during which gene regulation was studied at the level of individual model genes. The emergence of DNA microarrays in the late 1990s combined with the availability of whole-genome sequences opened new avenues toward the study of gene regulation at a global scale during development, in response to environmental stimuli, and in the context of disease. DNA

microarrays made it possible to identify groups of genes that characterize a cell type, developmental stage, environmental response, or disease state. This central method also made it possible to uncover transcriptional networks and examine the mechanisms by which these networks regulate physiological states and events.

Despite the great value of microarrays, a limitation is that they often compress the magnitude with which mRNA levels are modulated, and the degree of compression can vary from gene to gene[1]. Because of the limited dynamic range and quantitative value of microarray data sets, a low threshold – often 1.5- to 2-fold – has generally been used to define a group of modulated genes of interest, and these genes are then considered equally in studies of the regulatory mechanisms. One benefit of this approach is that hundreds or thousands of modulated genes can be grouped together, providing considerable statistical power for the subsequent analysis. With this approach, microarray-based studied resulted in numerous major discoveries and they continue to be of great value. However, the quantitative limitations have presented a barrier for many efforts to dissect regulatory mechanisms.

More recently, high-throughput sequencing technologies have led to the development of methods, including RNA sequencing (RNA-seq), that allow transcript levels to be evaluated at a global scale more accurately and quantitatively[2]. Although RNA-seq methods remain imperfect, their accuracy and larger dynamic range have opened new opportunities for meaningful mechanistic analyses of transcriptional cascades and networks.

One physiological event that has been studied extensively at a global scale is the response of cells within the mouse innate immune system to inflammatory stimuli, with most studies focusing on the stimulation of macrophages or dendritic cells to lipopolysaccharide (LPS) or lipid A (the active component of LPS). LPS and lipid A are recognized by Toll-like receptor 4 (TLR4), which activates several signaling pathways to induce a robust transcriptional

cascade. Early studies used DNA microarrays to evaluate the kinetics with which genes modulated in response to the stimulus are activated and inactivated[3–6]. The promoters of co-regulated genes exhibited enrichment for the binding sites of key transcription factors implicated in the inflammatory cascade, thereby providing a starting point toward a global understanding of the regulatory logic underlying the cascade. These early studies have been followed by large-scale studies in which gene expression profiles obtained using microarrays or RNA-seq were combined with ChIP-seq data sets and siRNA knockdowns to evaluate the roles of specific signaling pathways, transcription factors, and chromatin-related events in the cascade[7,8]. RNA-seq methods that monitor nascent transcripts rather than mRNA have also been employed to separate transcriptional regulation from the regulation of mRNA stability[9–12].

The results of these studies, performed by our laboratory and others, have provided considerable insight into mechanisms regulating the transcriptional cascade. However, a limitation of all of these genomics-based studies is that the results demonstrate trends and relatively loose relationships, while lacking precision and confidence in the direct or indirect roles of specific signaling pathways, transcription factors, and chromatin proteins in the regulation of individual genes within the cascade. One fundamental reason for the relatively low confidence of the results is that the functional relevance of transcription factor ChIP-seq peaks is difficult to evaluate, due to substantial evidence that some and perhaps most transcription factors bind opportunistically to many sites in addition to their functionally important interactions[13–18]. Even when ChIP-seq data sets are evaluated in the context of microarray and RNA-seq data sets, it is difficult to identify with confidence those genes that represent direct targets of a factor[8,18–20].

It could be argued that the next frontier is the development of experimental and bioinformatic strategies that allow the field to move from trends and loose relationships to

precise assignments of signaling molecules, transcription factors, and chromatin events to their direct targets.

Here, we describe an effort to dissect the lipid A-induced transcriptional cascade by incorporating the quantitative value of nascent RNA-seq data sets, as well as ChIP-seq and transcription factor binding motif data sets. A fundamental aspect of this approach was the separation of genes induced by a large magnitude from those induced by smaller magnitudes, which are far more prevalent. The emphasis on potently induced genes created barriers because of the reduced statistical power of the subsequent analysis, thereby requiring careful consideration of strategies that may provide meaningful advances. The results presented provide critical insights into the logic through which the cascade is regulated and a framework on which a complete understanding of the cascade can be built.

Results

Basic Properties of the Lipid A-Induced Transcriptional Cascade

To uncover principles regulating the lipid A-induced transcriptional cascade, we began by performing RNA-seq with RNA from mouse C57BL/6 bone marrow-derived macrophages (BMDMs) treated with lipid A for 0, 15, 30, 60, and 120 minutes. To separate transcriptional kinetics and the magnitudes of transcriptional activation and inactivation from influences of mRNA stability, the analysis was performed with nascent transcripts isolated from the chromatin of BMDMs. Although our long-term goal is to understand how both transcriptional and post-transcriptional modes of regulation contribute to the gene expression cascade, a preferred approach is to evaluate the contributions of each regulatory mode in isolation, with the resulting insights then combined.

An initial analysis of three biological replicates of the nascent transcript RNA-seq experiment indicated that 3,863 (14.1%) of the 27,384 annotated Refseq genes (NCBI37/mm9, prior to removal of duplicate isoforms) reached an expression level of at least 3 RPKM in at least one of the time points. We focused on genes that reached a relatively high expression level of 3 RPKM because our subsequent analysis emphasized induction magnitude, which can be evaluated only if both the basal and induced transcript levels can be measured with confidence.

Of the 3,863 genes expressed at 3 RPKM or greater, 1,340 (34.7%) were induced by at least 2-fold ($p<0.01$) (Figure 2-1A). Importantly, however, 57.8% of these genes were induced by less than 5-fold and 79.5% were induced by less than 10-fold, with only 14.7% induced by 10-50-fold and 5.8% induced by greater than 50-fold (Figure 2-1A). If all genes induced by 2-fold or greater were evaluated together, the analysis would be dominated by genes induced by less than 10-fold, which would likely mask key events involved in the regulation of the strongly induced genes. Notably, most induced genes encoding key cytokines, chemokines, effector molecules, and transcription factors known to play critical roles in immune regulation were induced by greater than 10-fold. For this reason, we chose to focus our analysis on the potently induced genes, with the resulting insights then examined in the context of the weakly induced genes (see below). It is noteworthy that the basal transcript levels of the weakly induced genes were generally higher than those of the strongly induced genes (Figure 2-1B).

With the above considerations in mind, we focused our analysis on 226 inducible genes. 215 of these genes were significantly ($p<0.01$) induced by at least 10-fold during the 2-hr induction period. The remaining 11 genes were transiently induced by 5-10-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. Notably, although the analysis focuses on only 226 potently induced genes, their basal and peak transcript levels, as well as their fold-induction

values, were distributed over more than two orders of magnitude (Figure 2-1C); these broad distributions suggest that, despite the focus on a limited number of genes, the genes analyzed are likely to be regulated by diverse mechanisms. It is also noteworthy that all 226 genes are protein-coding genes. Detection of short RNAs were would require a different RNA-seq method; long non-coding RNAs were captured by this approach, yet no non-coding RNAs were induced by 10-fold while reaching the 3 RPKM expression threshold.

Separation of Primary and Secondary Response Genes

Among the many parameters that can be used to dissect the lipid A-induced transcriptional cascade, we chose to first separate primary and secondary response genes. Toward this goal, we performed RNA-seq with nascent transcripts from BMDMs stimulated with lipid A for 0, 30, 60, and 120 min in the presence or absence of cycloheximide (CHX), an inhibitor of protein synthesis. (Please note that at least two complete biological replicates were performed for all RNA-seq experiments, with average RPKM and fold-induction values used for the bioinformatic analyses.) Analysis of maximum RPKM values for the 226 induced genes revealed 83 genes that were expressed at a level in CHX-treated cells that was <33% of the expression level in untreated cells (Figure 2-1D). These 83 genes were tentatively included in the secondary response group.

IFN-β expression is induced by lipid A and is known to activate a Type I IFN gene program[21,22]. RNA-seq analysis of nascent transcripts from Type I IFN receptor (IFNAR)-deficient (*Ifnar-/-*) BMDMs stimulated with lipid A revealed that 62 of the 226 inducible genes were expressed at less than 30% of wild-type (WT) (Figure 2-1D). Interestingly, 11 of these 62 IFNAR-dependent genes were classified as primary response genes in the CHX analysis because they exhibited expression levels in the presence of CHX that placed them just above

the threshold used for classification as secondary response genes. Nevertheless, an analysis of their induction kinetics revealed greater similarity to the other IFNAR-dependent secondary response genes than to the primary response genes (data not shown; see Figure 2-3). Because of their strong IFNAR-dependence and their kinetic profiles, and because CHX may not completely inhibit protein synthesis and may have indirect effects on gene expression, these 11 genes were added to the secondary response category (Figure 2-1D). Thus, 132 and 94 genes, respectively, were defined as primary and secondary response genes for the current analysis.

It is important to emphasize that the thresholds and criteria used to separate genes into primary and secondary response groups are subject to debate. Furthermore, some genes appear to possess both primary and secondary response components (data not shown). Thus, the classification assignments will need to be re-evaluated frequently as our knowledge of the regulatory cascade increases.

Separation of IFNAR-Dependent and -Independent Secondary Response Genes

As described above, a central feature of the secondary response to lipid A stimulation is the activation of Type I IFN signaling via IFNAR. Therefore, as the next broad step toward characterizing the lipid A cascade, we separated secondary response genes into IFNAR-dependent and IFNAR-independent groups. The assignment of genes was dictated primarily by the magnitude by which the expression of each secondary response gene was decreased in *Ifnar-/-* BMDMs in comparison to WT BMDMs. Strikingly, 42 of the 94 secondary response genes were expressed <10% of WT in *Ifnar-/-* BMDMs, with an additional 22 expressed between 10 and 33% (Figure 2-2A,B). Kinetic analysis revealed that 41 of the 42 genes that were 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 2-2C), indicating that a robust

transcription response to IFNAR signaling begins between 60 and 120 min post-stimulation. In contrast, 22 of the 23 secondary response genes 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 of stimulation (Figure 2-2C). This finding suggests that the CHX-sensitive events needed for activation of the IFNAR-independent secondary response genes generally occur much more rapidly than the autocrine/paracrine loop needed to activate the IFNAR-dependent genes.

To separate IFNAR-dependent and -independent genes more carefully, we further examined the RNA-seq data sets from lipid A-stimulated *Ifnar*$^{-/-}$ macrophages, as well as additional RNA-seq data sets we generated from WT macrophages stimulated for 0, 15, 30, 60, and 120 min with Pam3CSK4 (PAM), a TLR2 ligand that does not induce strong IFNAR signaling[22]. Twenty-nine secondary response genes were identified that remained inducible in the absence of IFNAR signaling (Figure 2-2D, top).

Interestingly, although these 29 IFNAR-independent secondary response genes were induced in PAM-stimulated WT BMDMs or in lipid A-stimulated *Ifnar-/-* BMDMs, a subset, including the critical T cell polarizing cytokines *Il12b*, *Il6*, *Lif*, and *Il27,* were induced much less potently by PAM than by lipid A (Figure 2-2D, bottom). This finding suggests that the TRIF pathway that is 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 possibility, a direct comparison of WT BMDMs to *Trif-/-* BMDMs revealed strong TRIF-dependence of these genes (Figure 2-2D, bottom). In fact, *Il12b*, *Il6*, *Lif*, and *Il27* exhibited greater TRIF-dependence than any other primary or secondary response gene (Figure 2-2E). 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 regulatory strategies described above, we performed gene ontology analysis with our 132 primary response genes, the 65 IFNAR-dependent secondary response genes, and the 29 IFNAR-independent secondary response genes (Figure 2-2F). The primary response gene analysis suggested broad roles in regulating inflammation and the quantities and functions of leukocytes and blood cells. As expected, the IFNAR-dependent secondary response genes were implicated in anti-viral responses. Most interestingly, the small group of IFNAR-independent secondary response class exhibited highly significant enrichment for genes that help regulate T cell proliferation, differentiation, and activation. Notably, the gene ontology program suggested that 14 of the 29 IFNAR-independent secondary response genes may be involved in the regulation of T cell responses. Eleven of these 14 genes are among the 13 IFNAR-independent secondary response genes that are most potently induced by lipid A (average induction of 376-fold for these 11 genes). Thus, these results reveal common regulatory features of a prominent group of genes that help bridge the innate and adaptive immune systems. Nevertheless, a careful examination reveals that the induction kinetics for each of these genes is unique (Figure 2-3), 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 Primary Response Genes

Shifting our attention to the 132 primary response genes, we first examined their expression kinetics in greater detail by nascent transcript RNA-seq with lipid A-stimulated BMDMs collected every five min during the first hour of activation, with an additional 120-min time point. We also performed nascent transcript RNA-seq with BMDMs from mutant mice lacking signaling molecules or transcription factors known to play central roles in the lipid A response.

Specifically, WT BMDMs were compared to BMDMs from *Myd88-/-, Trif-/-, Myd88-/-Trif-/-,* and *Irf3-/-* mice, with samples collected 0, 30, 60, and 120 min after stimulation. We also performed RNA-seq with WT BMDMs stimulated with lipid A for 0, 15, 30, 60, and 120 min in the presence of ERK and p38 MAPK inhibitors; the two inhibitors were analyzed together in this analysis because little effect was observed in pilot experiments with each inhibitor alone. Two biological replicates were performed for each experiment, and each experiment analyzed nascent transcripts. The results consider the maximum induced RPKM observed in WT cells for each gene to be 100% and the basal RPKM observed 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 2-4A shows that each perturbation resulted in a continuum of effects on the 132 primary response genes. For the purposes of this study, genes expressed <33% of WT were considered to be dependent on the factor that was eliminated or inhibited, but with the recognition that this solution is imperfect and will require continual refinement as our knowledge of the transcriptional cascade advances. By combining the data sets indicating the dependencies of each of the 132 genes with k-means cluster analysis of expression kinetics, an initial classification of the 132 primary response genes emerged (Figure 2-4D; see Figure 2-5 for gene names). Cluster 1 includes nine genes that exhibited reduced expression (<33% of WT) in both *Trif-/-* and *Irf3-/-* macrophages. Clusters 2-5 included 28 genes that exhibited reduced expression (<33% of WT) in *Trif-/-* macrophages but not in *Irf3-/-* macrophages (Figure 2-4B,D); these genes were subdivided by k-means analysis on the basis of their expression kinetics. Clusters 6-9 include 38 genes that exhibited reduced expression (<33%) in WT macrophages treated with both p38 and ERK MAPK inhibitors, but without strongly reduced expression in the *Trif-/-* macrophages; as above, the genes were subdivided by k-means analysis of expression kinetics (Figure 2-4D). Finally, Clusters 10-16 include the remaining 57 genes that did not

exhibit reduced expression in the presence of MAPK inhibitors or in the *Trif-/-* or *Irf3-/-* cells; these genes were divided into seven kinetic clusters by k-means analysis. It is noteworthy that only five of the 132 primary response genes exhibited reduced expression (<33%) in Myd88-/- cells (Figure 2-4C,D), in contrast to the eight of 29 IFNAR-independent secondary response genes. This finding is consistent with our previous analysis that implicated MyD88 more strongly in the regulation of secondary response genes than primary response genes in LPS-stimulated macrophages[23]. Importantly, no genes were induced in *Myd88-/-Trif-/-* mutant cells (data not shown).

In addition to the degree of dependence of each primary response gene on MyD88, TRIF, IRF3, and MAPKs, Figure 2-4D indicates the basal transcript and fold-induction values for each gene, reflecting the broad ranges documented in Figure 2-1C. Furthermore, Figure 2-4D indicates which genes contain CpG-island promoters or low CpG (LCG) promoters. Consistent with our previous studies (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 among the other clusters according to rules that remain to be elucidated.

Initial Transcription Factor Binding Motif and ChIP-Seq Analyses

The above studies provide a framework for mechanistic dissection of the primary response to lipid A stimulation by 1. focusing attention on genes that are reduced relatively strongly, 2. separating primary response and secondary response genes, 3. using loss-of-function and inhibitor studies to identify genes that exhibit requirements for key signaling pathways and factors known to participate in the primary response, 4. clustering the genes on the basis of their

induction kinetics, and 5. displaying induction magnitudes and basal transcript levels along with promoter type (CpG-island vs. LCG).

To extend this foundation toward a meaningful mechanistic understanding of the transcriptional cascade, we first evaluated the over-representation of transcription factor binding motifs within the promoters of genes within each of the 16 clusters in Figure 2-4D. The motif analysis results (Figure 2-6) provided initial insight into transcription factors that may regulate genes within each cluster. However, toward the goal of understanding the molecular logic through which the transcriptional cascade is regulated, the statistical enrichments revealed by this analysis were largely unsatisfying. For example, although NF-κB motifs are statistically enriched in the promoters of genes in several clusters, a closer analysis revealed considerable heterogeneity within each of these clusters, with only a subset of promoters in a cluster generally containing a strong NF-κB motif (data not shown). Statistical but imprecise enrichments were also apparent when examining published ChIP-seq data sets for NF-κB and other transcription factors known to participate in the response (data not shown). Additional strategies are therefore needed to move beyond statistical enrichments toward more precise and meaningful insights into the logical control of a stimulus-induced transcriptional cascade.

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

Because of its common role in regulating inducible transcription in response to inflammatory stimuli, we first focused on NF-κB. In an effort to address the uncertainties associated with the interpretation of both ChIP-seq and binding motif data sets, we envisioned that quantitative and simultaneous consideration of both types of data may be of value. As a first step, NF-κB ChIP-seq peak scores and binding motif scores were plotted for the promoters (-500 to +150 relative to the transcription start site [TSS]) of each of the 132 primary response genes (Figure 2-7A).

For NF-κB motifs, we took advantage of recent protein binding microarray (PBM) analyses in which relative binding of different NF-κB dimeric species to a broad range of DNA sequence motifs was examined experimentally[24]. ChIP-seq experiments were then performed with antibodies targeting RelA, a subunit of the most abundant NF-κB dimer (RelA:p50) thought to be involved in transcriptional activation in response to TLR4 signaling. ChIP-seq experiments were performed with BMDBs 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 (see Experimental Procedures). Analysis of the RelA ChIP-seq data sets revealed 8,458 total peaks, with 942 promoter peaks among 21,168 annotated Refseq genes.

When focusing on the promoters of the 132 strongly induced primary response genes, an interesting relationship between ChIP-seq peak scores and binding motif Z scores emerged. Specifically, a motif Z score threshold emerged that resulted in a high probability of a strong ChIP-seq peak; 37 of 44 promoters (84%) containing an NF−κB 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-κB motif was below this same threshold supported similarly strong binding (Figure 2-7A, left, 4B, left). On the basis of this initial observation, promoters were separated into six distinct classes for further analysis, including three ChIP-seq categories (no binding, peak strength <19, and peak strength >19) and two motif categories (Z score <6.4 and >6.4).

To evaluate the significance of these classifications, we examined promoters for all other annotated genes. A visual examination of the graph in Figure 2-7A (right) revealed that the vast majority of RelA ChIP-seq promoter peaks in genes that do not represent strongly induced primary response genes were associated with promoters with relatively weak motifs (<6.4, Figure 2-7A, right). For a closer examination, annotated genes were separated into five groups:

the 132 strongly induced primary response genes, the 94 strongly induced secondary response genes, 732 genes induced between 2 and 10 fold, 1732 genes that were expressed at a nascent transcript level >3 RPKM but without induction, and the remaining 18,487 annotated genes expressed at a transcript level <3 RPKM.

An examination of the ChIP-seq/motif categories for genes in each of these five classes revealed extensive enrichment of genes whose promoters combined strong ChIP-seq peaks and strong NF-κB motifs among the strongly induced primary response gene class. Specifically, whereas 28% (37/132) of the strongly induced primary response genes combined strong ChIP-seq peaks and motifs, only 1.6% (27/1723) of expressed but uninduced genes combined strong peaks and motifs. In contrast, little or no enrichment of strongly induced primary response genes 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 enrichment in the primary response gene 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 primary response genes 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 strongly induced primary response genes suggests that most or all of the 37 primary response genes possessing these properties are directly activated by RelA-containing dimers via direct promoter binding. The ability to define a motif threshold (Z score=6.4) above which 84% of promoters supported strong NF-κB binding is interesting to consider in light of previous models suggesting that NF-κB may usually interact functionally with weak motifs[25,26]. Instead, the data suggest that a strong NF-κB motif as defined by PBM analysis is usually sufficient to support strong NF-κB binding to an inducible promoter.

The significant but lesser enrichment of promoters with strong NF-κB peaks and weak binding motifs among the strongly induced primary response genes is also of interest. In these promoters, NF-κB may bind directly to weak motifs, perhaps via cooperative binding with other transcription factors. Alternatively, NF-κB may be recruited to these promoters by other transcription factors, or the NF-κB ChIP-seq signal could be due to looping of an NF-κB-bound enhancer to the promoter. Importantly, fewer primary response genes were found in this strong ChIP-seq peak/weak motif category than in the strong ChIP-seq/strong motif category, suggesting that NF-κB usually associates with promoters via direct binding to strong motifs.

An examination of the 732 genes induced by 2-10-fold provides additional insights. A higher percentage of genes in this weakly induced class (5.9%) contain strong NF-κB 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-κB binding to strong motifs. However, a much smaller percentage of genes in this 2-10-fold induced class (5.9%) combine strong NF-κB peaks and motifs than in the strongly induced primary response gene class (28%), suggesting that a much smaller fraction of the weakly induced genes is regulated by direct binding of NF-κB to strong promoter motifs.

Examination of NF-κB regulated genes

A major goal of this study is to elucidate the logic through which the lipid A-induced transcriptional cascade is regulated. This issue not only concerns the question of how a diverse array of factors and pathways contribute to the cascade, but also the question of whether specific factors arbitrarily regulate individual genes or whether an underlying logic exists. An examination of the identities of the 37 strongly induced primary response genes that combine

strong ChIP-seq peaks and strong motifs provides compelling evidence of an underlying logic. Specifically, over a third of these genes (13 of 37, see Figure 2-7C) encode NF-κB or IκB family members or key regulators of NF-κB activation, including three NF-κB family members (*Nfkb1*, *Nfkb2*, and *RelB*), five IκB family members (*Nfkbia, Nfkbib, Nfkbid, Nfkbie,* and *Nfikbiz*), two NF-κB-inducing receptors (*Tlr2* and *Cd40*), and three regulators of NF-κB activation (*Tnfaip3*, *Tnip3*, and *Traf1*). Strikingly, these 13 genes represent the only NF-κB/IκB family members or direct regulators of NF-κB signaling in the entire group of 132 primary response genes. It is noteworthy that the promoters of genes encoding the two NF-κB family members (*Rela* and *Rel*) and one IκB family member (*Bcl3*) missing from this list also combine a strong RelA ChIP-seq peak with a strong NF-κB motif (Figure 2-7F); these three genes were not among the 132 strongly induced primary response genes because their magnitudes of induction did not reach the 10-fold threshold.

Figure 2-7C shows the complete list of genes that combine strong ChIP-seq peaks and motifs, along with the strongest NF-κB binding motif found in each promoter on the basis of the PBM data. The strong binding observed at the 37 primary response genes in this category can be accounted for by only 21 different motifs, as some motifs are found in two or more of the promoters. An examination of these motifs shows that each adheres 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) (4 motifs).

The evidence that a high level of similarity to the optimal NF-κB consensus may be needed for consistent NF-κB binding in the RelA ChIP-seq experiments was initially surprising, given the formal possibility that NF-κB might often bind to weak motifs via cooperative binding with other factors. However, additional support for the significance of this finding is provided by an examination of binding motif enrichment at the 132 primary response genes in comparison to

the 1,723 expressed but uninduced, without any consideration of ChIP-seq peaks. Specifically, motifs with Z scores above 8.0 were strongly enriched among in the promoters of the 132 primary response genes. 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 2-7G).

One remaining question is the reason seven promoters with motifs exceeding the threshold of 6.4 did not support detectable RelA binding in the ChIP-seq experiments. Three of these motifs possess high Z scores (8.4-8.6) and their sequences clearly conform to the consensus that would be expected to support strong NF-κB binding (Figure 2-7D). It is noteworthy, however, that two of these three motifs are at a distance upstream of their corresponding TSS (-310 and -395) that exceeds the distance of all but five of the 37 promoters that support NF-κB binding (Figure 2-7D). We hypothesize that these two motifs do not support binding in vivo because they are occluded by nucleosomes flanking the promoter. Similarly, the third strong motif is located farther downstream from the TSS (+137) than the motifs found in the 37 promoters that support strong NF-κB binding, suggesting that this motif may also be masked by a nucleosome. The four remaining motifs that fail to support binding possess Z scores between 6.7 and 7.4. On the basis of an examination of these motifs, we speculate that their Z scores may be defined inaccurately due to limitations of the PBM method. It is noteworthy that one of these motifs is found in two different primary response promoters, and none of these motifs match the 21 motifs in Figure 2-7C that coincided with RelA ChIP-seq peaks. We tentatively conclude that the Z score threshold of 6.4, although a fairly consistent indicator of capacity for NF-κB binding, is imperfect, perhaps due to imperfections in the PBM values or because other subtle features of the promoter environment can influence NF-κB binding to motifs in the borderline Z score range.

To summarize, we found evidence that considerable insight can be provided by classification of promoters on the basis of quantitative and simultaneous consideration of ChIP-seq peak scores and PBM Z scores. Notably, a parallel analysis using Transfac position weight matrix (PWM) scores rather than PBM Z scores was also of value, but with slightly less predictive accuracy (data not shown), presumably because PBM scores, unlike Transfac scores, are based on a direct experimental evaluation of protein binding to each motif sequence. We found that promoters combining strong ChIP-seq peaks and binding motifs are greatly enriched in a quantitatively defined group of strongly induced primary response genes. This combination of properties does not appear to have evolved randomly, as all NF-κB and IκB family members and regulators of NF-κB signaling within the group of primary response genes fall into this category. The results strongly suggest that a single strong motif in a nucleosome-depleted promoter region is sufficient for strong NF-κB binding, with evidence that a protein-DNA interaction affinity threshold usually must be exceeded to support binding. Our data also suggest that only a small percentage of weakly induced genes (2-10-fold) are regulated by direct binding of NF-κB to a strong promoter motif. A major question that remains to be answered is the significance of NF-κB ChIP-seq peaks at a limited number of promoters that lack strong motifs, given that the vast majority of promoters containing similarly weak motifs do not support binding.

Kinetic and Functional Analysis of Putative NF-κB Targets

The above results suggest that the 37 strongly induced primary response genes combining strong ChIP-seq peaks and motifs are strong candidates for genes that are directly regulated by canonical RelA:p50 NF-κB dimers. To test this prediction, we examined the activation kinetics for these genes, as well as their dependence on RelA. An examination of the activation kinetics

revealed that most of these genes are first upregulated between 10 and 20 minutes after lipid A stimulation. This is evident from the third graph in Figure 2-9A, in which the fold-increase in RPKM relative to the preceding time point is highlighted. It is noteworthy that, although most of these genes are initially upregulated at approximately the same time, their expression kinetics are quite diverse, implicating other factors in their regulation. It is noteworthy that putative NF-κB target genes that are also dependent on MAPK signaling were, on average, induced slightly earlier than the other putative target genes (Figure 2-9C).

We also examined RelA dependence by comparing WT and *Rela-/-* fetal liver-derived macrophages. A wide range of dependencies was observed (Figure 2-9A, right), possibly due to redundancy with c-Rel. However, most of the genes with strong NF-κB ChIP-seq peaks and binding motifs in their promoters exhibited considerable RelA dependence. It is noteworthy that 3 of these 37 genes did not exhibit significant RelA dependence; at these genes, NF-κB promoter binding may not regulate induction or greater redundancy may exist between different NF-κB family members.

Since most genes containing strong NF-κB ChIP-seq peaks and binding motifs in their promoters are induced with similar activation kinetics and contain considerable RelA dependence, we asked whether these functional properties are restricted to genes whose promoters contain strong ChIP-seq peaks and motifs. Interestingly, several other primary response genes exhibited similar activation kinetics and/or degrees of RelA-dependence (Figure 2-9B,D). A subset of these genes contains RelA ChIP-seq peaks in their promoters (Figure 2-9B, right), but most do not. The significance of this finding is difficult to determine from these data, but we speculate that NF-κB directly regulates these genes by binding to distant enhancers. Consistent with this possibility, RelA ChIP-seq peaks are found at variable distances

from many primary response genes (Figure 2-8), although it is difficult to determine the significance of these peaks and the specific genes these binding events might regulate.

Gene-Specific Regulation of IRF3-Dependent Genes

Although most studies of transcriptional cascades and networks focus on large clusters of co-regulated genes, our quantitative analysis of a limited number of potently induced genes provides a glimpse of the fact that many genes are subject to unique modes of regulation, whereas others fall into very small groups of co-regulated genes. This concept is exemplified by an examination of primary response genes that exhibit strong dependence on the transcription factor, IRF3. As shown in Figure 2-4, only 9 of the 132 strongly reduced primary response genes exhibited expression levels in both *Irf3-/-* and *Trif-/-* macrophages that fail to reach 33% of the expression level observed in WT macrophages. As shown in Figure 2-9, five of these genes are contained within the group of 37 primary response genes containing strong NF-κB ChIP-seq peaks and motifs; the other four lack RelA binding and NF-κB motifs. The kinetic profiles of these 9 genes are shown in Figure 2-11A. One notable difference between the five genes that appear to be regulated by both NF-κB and IRF3 versus the four genes regulated by IRF3 alone is that the induction magnitude of the former group is much higher than that of the latter group, with average induction magnitudes of 643-fold and 40-fold, respectively (Figure 2-11A,B).

A careful examination of the five genes that appear to be regulated by both NF-κB and IRF3 is especially revealing with respect to the extent to which genes have evolved unique regulatory strategies. Among this group, the expression kinetics of *Ccl5* and *Ifnb1* are unique, whereas *Cxcl10*, *Gbp5*, and *Irg1* exhibit greater similarity to each other (Figure 2-11A). These latter three genes are initially induced 10-15 min. post-stimulation, the time at which most NF-

κB-dependent genes are induced. Consistent with the hypothesis that NF-κB contributes to this early induction, RelA ChIP-seq peaks were observed at these genes by 15 min after lipid A stimulation (Figure 2-11A, right). Furthermore, their early induction is unaltered in *Irf3-/-* macrophages (data not shown). The IRF3-dependence of these genes is observed only at later time points, when IRF3 is activated and appears to synergize with NF-κB to support potent induction.

Interestingly, *Ccl5* is unique in comparison to these three genes and all other primary response genes. *Ccl5* transcriptional induction was not observed until the 25 min time point (Figure 2-11A). Furthermore, RelA binding was not observed at this gene until the 30 min time point (Figure 2-11A, right). We previously used a restriction enzyme accessibility assay to document inducible nucleosome remodeling at the *Ccl5* promoter and we showed that this nucleosome remodeling event was not observed in *Irf3-/-* macrophages[27]. Thus, *Ccl5* appears to employ a unique mode of regulation in which transcriptional activation requires an IRF3-dependent nucleosome remodeling event to provide NF-κB with access to its promoter binding site, with NF-κB and IRF3 then presumably synergizing to support the potent 1,700-fold induction of this gene. Notably, *Ccl5* was the only primary response gene bound by RelA that lacked a called RelA ChIP-seq peak at the 15-min time point (Figure 2-9A, right; Figure 2-11A, right).

Ifnb1 is one of the best-studied inducible genes, yet its regulatory mechanisms also appear to be unique. Transcriptional induction of this gene was not observed until the 35-min time point. However, in contrast to the properties of *Ccl5*, RelA binding to the *Ifnb1* promoter was observed at all induced time points examined (Figure 2-11A, right). This early binding is consistent with extensive prior evidence that the promoter region containing the NF-κB and IRF3 binding sites lacks a nucleosome in unstimulated cells[28]. Nevertheless, despite the

observed binding of NF-κB upon its early activation, transcriptional induction of *Ifnb1* was not observed until 35 min post-stimulation, consistent with prior evidence that activation is strongly dependent on synergy between NF-κB, IRF3, and ATF-2/c-Jun complex[29].

Although high-quality ChIP-seq data with IRF3 antibodies in mouse cells have not been obtained, an examination of NF-κB and IRF3 binding motifs revealed that consensus IRF3 motifs accompany the strong NF-κB motifs in all five of these promoters (Figure 2-11C). The distances between the IRF3 and NF-κB motifs range from 2-bp in the *Ifnb1* promoter to 55 bp in the Ccl5 promoter (Figure 2-11D). Interestingly, of the four IRF3-dependent genes that do not contain strong NF-κB motifs, only one gene (*Isg15*) contained an IRF3 motif of similar strength to those found in the genes containing strong NF-κB motifs (Figure 2-11C,D). Thus, IRF3 may regulate the other three genes by binding a more distant site or via an indirect mechanism.

Together, these results begin to 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 studies raised the hypothesis that IRF3 and NF-κB cooperatively activate hundreds or thousands of genes[30], the results presented here demonstrate that only five primary response genes induced by greater than 10-fold combine strong NF-κB binding, strong IRF3-dependence, and a strong IRF3 motif. We cannot rule out the possibility that these two factors collaborate at distant enhancers, but if such collaboration occurs at the enhancers of strongly induced primary response genes, it does not appear to result in strong IRF3-dependence. Thus, functionally significant collaboration between these two key factors may occur at a much smaller number of inducible genes than expected.

Regulation of Transiently Transcribed Genes by Serum Response Factor (SRF)

From a direct visualization of the initial cluster analysis in Figure 2-4, the most distinctive cluster is arguably Cluster 6. Genes in this MAPK-dependent cluster exhibit rapid upregulation within 5 min of lipid A stimulation, with downregulation of nascent transcripts by the 30 or 35 minute time points. This cluster contains only three genes, *Egr1*, *Fos*, and *Nr4a1*, yet the initial binding motif analysis suggests enrichment of binding sites for SRF in their promoters. Because SRF has been implicated in the induction of these genes in response to a broad range of stimuli, we examined SRF binding by ChIP-seq at a genome-wide scale in macrophages stimulated with lipid A for 0, 15, 30, 60, and 120 minutes. Because we did not observe kinetic changes in SRF binding, the analysis was followed by focusing on reproducible called peaks across all five time points.

Interestingly, the SRF ChIP-seq data sets yielded the strongest ChIP-seq signals we have detected to date with any transcription factor we have examined, as well as the greatest specificity of binding, with only a small number of peaks. A simultaneous examination of SRF ChIP-seq peaks and Transfac Position Weight Matrix (PWM)-defined motifs at the promoters of the 132 strongly induced primary response genes revealed that only seven promoters contain strong ChIP-seq peaks (ChIP-seq score >10) and all seven promoters contain strong motifs (Transfac score >90) (Figure 2-11E). No strong ChIP-seq peaks were observed at this group of 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 quite far from their TSS (-306 and -331), raising the possibility that they are occluded by nucleosomes. Thus, as with NF–κB, strong binding of SRF is closely correlated with peak strength, suggesting that weaker SRF motifs do not support binding via cooperative interactions with other factors.

Surprisingly, an examination of the remaining 21,036 annotated genes revealed only 39 additional genes that reach the same peak and motif thresholds achieved by the 7 binding events observed within the small primary response group (Figure 2-11E,F). Instead, the vast majority of binding events observed in other gene classes combined either strong or weak ChIP-seq peaks with a motif of a strength that supports SRF binding in only a small fraction of instances (Figure 2-11E,F). Thus, although only 7 of the 132 strongly induced primary response genes contain strong SRF ChIP-seq peaks and motifs, this represents a great enrichment relative to all other gene classes; notably, only 5 of 732 genes in the 2-10-fold class (*Ralgapa1, Filip1l, Actg1, Lima1, Glipr1*) contain strong SRF peaks and motifs in their promoters.

A closer examination of the seven genes that contain strong SRF ChIP-seq peaks and motifs supports the hypothesis that six of these genes are functional targets of SRF. This group includes the three genes found within cluster 6 of Figure 2-4A (*Egr1*, *Fos*, and *Nr4a1*) along with four additional genes (*Egr2, Dusp5, Zfp36*, and *Rnd3*). In an analysis of the magnitude of induction at each time point relative to the previous time point ((X_n/X_{n-1}; Figure 2-11G, third panel), we found that all but *Rnd3* are initially upregulated during the first 5 min of lipid A stimulation, and all but *Rnd3* are dependent on MAPK signaling for their induction. The similar initial induction kinetics is apparent from an examination of the fold-change in signal at each time point relative to the previous time point (Figure 2-11G). MAPKs are known to be responsible for activation of the ternary complex factors (TCFs) that serve as critical co-activators for SRF[31–33]. The fact that *Rnd3* is activated with different kinetics and does not exhibit MAPK dependence suggests that the strong binding of SRF to an SRF consensus sequence in its promoter may not have functional consequences, or activation of this gene may require a second class of SRF co-activator proteins that are not activated by MAPK signaling[34].

Interestingly, an examination of the overall induction kinetics for this group of seven genes explains why only three were placed in the same kinetic cluster in Figure 2-4A: these

three genes exhibit relatively uniform kinetics of transcriptional induction and repression, whereas *Egr2*, *Dusp5*, and *Zfp36*, although initially induced at the same 5-min time point, are either further upregulated at later time points (*Egr2* and *Dusp5*) or are upregulated less potently and downregulated more slowly (*Zfp36*). These results therefore illustrate the necessity for careful evaluation of multiple types of data and multiple parameters within a data set to gain insight into the direct functional targets of transcription factors.

Lastly, an analysis of the 132 primary response genes led to the identification of only two additional genes that exhibit similarly rapid induction kinetics as the six genes discussed above: *Btg2* and *Ier2*. These two genes lack promoter ChIP-seq peaks and motifs for SRF, but instead were among the group of genes containing strong NF-κB ChIP-seq peaks in their promoters. 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 were found to contain strong SRF ChIP-seq peaks at upstream regions that coincide with CpG islands and are conserved through evolution (Figure 2-12B). The SRF peak at the Btg2 locus is 10 kb upstream of the TSS, whereas the SRF peak at the Ier2 locus is 1 kb upstream of the TSS. Remarkably, only three other primary response genes contain strong SRF ChIP-seq peaks within 10 kb of their TSS (upstream of the promoter), indicating that this property is rare. These results support a hypothesis in which SRF contributes to the early transient induction of these genes by cooperating with NF-κB bound to the promoters. We further speculate that this arrangement allows these two genes to be induced most potently by stimuli that induce both NF−κB and SRF, whereas cooperation between these two factors may not occur at the other SRF targets. It is noteworthy, however, that neither gene exhibited dependence on RelA, suggesting that SRF is the dominant regulator of transcription in cells that induce both NF−κB and SRF, or that RelA acts redundantly with other NF−κB family members (e.g. c-Rel) to regulate these genes.

Discussion

Through integration of quantitative information from chromatin RNA-seq, ChIP-seq, and transcription factor binding motif datasets, we have gained valuable insight into the regulatory mechanisms that govern transcriptional cascades in response to an inflammatory stimulus. Furthermore, the focused analysis on a limited number of potently induced genes provided us with insight into regulatory differences between the potently induced and the more prevalent weakly induced genes.

By assessing gene expression in both CHX-treated and *Ifnar-/-* RNA-seq datasets, we classified 132 and 94 potently induced genes as primary and secondary response, respectively. Based on SRF ChIP-seq analysis in combination with SRF binding motif analysis and MAPK sensitivity in MAPK inhibitor-treated macrophages, we found that the SRF targets were among the most rapidly and transiently induced genes in response to lipid A (Figure 2-13A). Of the remaining primary response genes, two exhibited similar rapid and transient kinetics. Interestingly, not only did both genes have SRF binding in evolutionary conserved regions between 1-10 kb upstream from the TSS, but both also displayed strong promoter RelA binding, suggesting that NF-κB and SRF collaborate in the regulation of a subset of genes.

Similarly, by taking into consideration the quantitative information from RelA ChIP-seq and p50:RelA binding motif datasets, along with RelA dependence and kinetic expression profiles, we identified 37 putative NF-κB target genes. The majority of these genes exhibited similar changes in expression at the 10-15 minute time point (Figure 2-13A, second panel). Although the majority of the NF-κB target genes exhibited a characteristic change in expression between 10 and 15 minutes, it is important to note that their activation dynamics were not entirely homogeneous. For example, the NF-κB targets were sub-classified into those that also require MAPK based on their expression in MAPK-deficient macrophages (NF-κB/MAPK). This

subset of genes was activated earlier than the NF-κB targets that also required IRF3 (NF-κB/IRF3). The NF-κB/IRF3 targets, which were sub-classified based on their low expression in TRIF- and IRF3- deficient macrophages, were activated with delayed kinetics compared to other NF-κB target genes, suggesting that IRF3 contributed to their delayed expression. This supports the idea that the distinct expression kinetics within the NF-κB target genes was due to collaboration of NF-κB with other transcription factors or signaling pathways.

Strikingly, all of the key regulators of NF-κB signaling and regulation contained in the potently induced primary response were targets of NF-κB itself, indicating that there is an underlying logic that may explain the framework for inducible transcription. Examination of the remaining primary response genes identified a number of genes that exhibited similar activation kinetics and/or RelA-dependence despite the absence of strong κB motifs or RelA binding peaks. While most of these genes did not have significant RelA peaks in the promoter, many of them had RelA peaks within 10kb from the TSS. This suggests that the subset of genes that do not have strong RelA binding in the promoters but exhibit similar activation kinetics and/or RelA dependence could be regulated by NF-κB at enhancers. Taken together, we hypothesize that the promoter-binding RelA targets are inducible in diverse cell types, while the subset of proposed enhancer-regulated NF-κB target genes are activated in a cell type specific manner.

In addition to a subset of NF-κB/IRF3 regulated genes, we identified a relatively small number of IRF3 (but not NF-κB) target genes based on their low expression in *Trif-/-* and *Irf3-/-* macrophages. Comparison of these two subsets of genes revealed that the IRF3/NF-κB targets were activated slightly earlier and with much higher potency than the IRF3 targets. We speculate that this is due to NF-κB translocation into the nucleus soon after lipid A treatment, followed by delayed activation of IRF3 to enhance transcription of the target genes. It is important to note that these are not generalizable rules governing activation of this subset of

genes. Although similar, the activation kinetics within the small subset of IRF3/NF-κB target genes was not homogeneous, indicating that each gene is likely to have unique mechanisms that regulate their activation. For example, previous studies have shown that *Ccl5* requires promoter nucleosome remodeling in an IRF3-dependent manner for its activation. The finding that *Ccl5* is the only IRF3/NF-κB target gene with delayed RelA binding at the promoter, and that this delay corresponds with delayed activation of *Ccl5* (discussed in Figure 2-11A) suggests a model in which RelA binding at the *Ccl5* promoter occurs only after an IRF3-mediated nucleosome remodeling event.

The secondary response genes were categorized based on their dependence on IFNAR. Using RNA-seq datasets derived from PAM-stimulated wildtype macrophages and lipid A-stimulated *Ifnar-/-* macrophages, 29 and 65 genes were classified as IFNAR-independent and IFNAR-dependent, respectively (Figure 2-13B). Examination of their activation profiles revealed distinct expression kinetics between the two subsets of genes, where the IFNAR-independent genes were activated with earlier kinetics relative to the IFNAR-dependent genes. Additionally, we identified key T cell polarizing cytokines such as *Il12b, Il6, Il27,* and *Lif* as IFNAR-independent. Further gene ontology analysis revealed that all of the most potently induced IFNAR-independent genes played roles in communicating with and initiating the adaptive immune response, while the IFNAR-dependent genes were critical for antimicrobial responses. Interestingly, prior studies have demonstrated that IFNAR-independent genes such as *Il12b, Il6*, and *Nos2* require nucleosome remodeling at the promoter for their activation. Together, this suggests that the IFNAR-independent genes likely have unique mechanisms regulating their activation.

Lastly, there were small subsets of MAPK-dependent and TRIF-dependent genes (Figure 2-13A). There was also a smaller subset of genes that could not be classified based on the criteria examined in this study. Further investigation is needed to understand what

transcription factors regulate these subsets of genes. However, the mechanistic framework established by this quantitative analysis has provided important insight towards understanding the regulatory logic underlying selective transcription in response to external stimuli and in different cell types.

Experimental Procedures

Cell Culture and Reagents

Bone marrow-derived macrophages were prepared from 6-week-old C57BL/6, *Myd88*-/-, *Trif*-/-, *Irf3*-/-, or *Ifnar*-/- male mice. Fetal liver macrophages were prepared from D14.5 C57BL/6 or *p65*-/- embryos. Macrophages were activated on day 6 with lipid A (100 ng/ml) (Sigma) or Pam3CSK4 (100 ng/ml) (InvivoGen). When indicated, macrophages were preincubated for 15 minutes with cycloheximide (10 mg/ml) or 1 hour with PD0325901 (10 mM) (Sigma) and BIRB0796 (1 mM) (AXON Medchem) prior to activation.

RNA and RNA-seq Library Preparation

Subcellular fractions of macrophages were prepared as described[12]. Chromatin purity was confirmed by immunoblot analysis of SNRP70, b-Tubulin (Sigma), and H3 (Abcam). Total RNA and chromatin RNA were isolated using TRI-reagent (MRC) followed by RNeasy columns (Qiagen). RNA was treated with RNase-free DNase I (Qiagen) prior to elution into RNase-free water. Chromatin RNA was depleted of ribosomal RNA using the Ribominus Eukaryote kit (Life Technologies). Strand-specific libraries were generated by using 60 ng chromatin RNA or 400 ng total RNA according to manufacturers instruction from the TruSeq RNA Sample Preparation Kit v2 (Illumina), with the following modifications: second strand cDNA was synthesized in the

presence of deoxyuridine triphosphate (dUTP) according to the dUTP method[35]. cDNA libraries were single-end sequenced (50bp) on an Illumina HiSeq 2000.

RNA-seq Read Mapping and Processing

Reads were aligned to the mouse genome (NCBI37/mm9 build) with TopHat v1.3.3 and allowing reads to be aligned once with up to two mismatches per read permitted. RPKM values were calculated as described[2]. Since chromatin transcripts are largely unspliced, RPKM values were calculated by counting all mapped reads within the transcription unit and divided by the length of the whole locus. mRNA RPKM values were calculated by counting all reads mapped to exons and divided by the length of the spliced product.

All RPKMs represent an average from two biological replicates. Genes were included in the analysis if they met all of the following requirements: The maximum RPKM value must reach 3 at any of the lipid A-stimulated time points, the fold induction level relative to the basal RPKM reached 10-fold, and the induced expression level was significantly different from the basal expression level (P<0.01), as determined by the EdgeR package in R Bioconductor[36]. Additionally, a gene was also included for further analysis if the fold induction relative to the basal RPKM reached 5-fold at the 15-minute time point of lipid A treatment.

Co-expressed gene classes were classified based on their dependence for the following factors, in the following order: IRF3, TRIF, MAPK, and unclassified. The genes in each class were then sub-clustered into k-means co-expression clusters using Cluster3[37] with log2 normalized RPKM values.

To determine the dependence of a gene for a factor, the percent expression relative to wildtype C57BL/6 expression was used. The basal RPKM value for each gene in wildtype

samples was set to 0% and the maximum RPKM value for that gene was scaled to 100%. The RPKMs in the knockout samples were converted to percent expression using this scale. For *Myd88-/-, Trif-/-, Irf3-/-, Ifnar-/-,* cycloheximide-treated, and MAPK inhibited samples, the maximum percent expression was used to represent the dependence of a gene for the perturbation. For *Rela-/-,* the percent expression in *Rela-/-* fetal liver macrophages at the earliest time point that wildtype samples were induced at least 3-fold was used to represent the dependence of a gene on RelA.

Chromatin Immunoprecipitation and Library Preparation

Chromatin immunoprecipitation of RelA was performed as previously described[38]. Briefly, bone marrow-derived macrophages were activated with 100 ng/ml of lipid A. Following stimulation, cells were fixed, nuclei were purified and lysed in buffer containing 1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, supplemented with protease inhibitors. Chromatin was sheared with a Diagenode Bioruptor to fragments ranging between 200 – 1000 bps. Chromatin was immunoprecipitated with anti-RelA antibody (Abcam). ChIP-seq libraries were made using the Illumina TruSeq ChIP Sample Prep Kit according to manufacturer's instruction.

Chromatin immunoprecipitation of SRF was performed as previously described[39], with modifications. After fixing activated bone marrow-derived macrophages, nuclei were purified and lysed in buffer containing 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, supplemented with protease inhibitors. Chromatin was sheared with a Misonix 3000 microtip. The fragmented chromatin was immunoprecipitated with anti-SRF antibody (Santa Cruz). ChIP-seq libraries were made using the LTP Library Preparation Kit (Kapa) according to manufacturer's instruction.

ChIP-seq Read Mapping and Processing

Reads were aligned to the mouse genome (NCBI37/mm9 build) with Bowtie2 restricting to uniquely mapped reads. Uniquely mapped reads were used for peak calling and annotation using HOMER, a software suite for next-generation sequencing analysis[40]. Peaks were called if they passed a false discovery rate of 0.01, and were enriched over input samples. Called peaks were considered for downstream analysis if peaks from at least 4 of 7 replicates were overlapping within 200 bp for RelA and 5 of 5 replicates were overlapping within 300 bp for SRF using the mergePeaks function. Peaks were annotated to Refseq genes based on the closest transcription start site (TSS).

Motif Analysis

The promoters of genes were used for motif analysis unless otherwise indicated. The promoter was defined as the region spanning -500 bp to +150 bp relative to the TSS. The strongest p50:RelA binding site within each promoter was identified against the consensus motif determined through a protein binding microarray (PBM) dataset of NF-κB dimers[24]. The strength of the motifs is represented as Z scores. Transfac position weight matrices (PWMs) were used to identify the best matching SRF and IRF3 binding sites in promoters using Pscan[41]. The strength of each motif is represented by a numerical value on a scale of 0-100, with 100 being a perfect match to the consensus motif.

Accession Numbers

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE67357.

Figure Legends

Figure 2-1. Properties of the Lipid A-Induced Transcriptional Cascade

Chromatin-associated transcripts were isolated from 0, 15, 30, 60, and 120 minute lipid A activated BMDMs and analyzed by RNA-seq. (A) The distribution of maximum fold inductions relative to unstimulated over the 2-hour stimulation period for the 1,340 significantly induced (2-fold, $p < 0.01$) and expressed (3 RPKM) genes is shown. The dashed gray lines represent 5-fold, 10-fold, and 50-fold induction thresholds. (B) The 1,340 induced genes were grouped into the following fold induction bins: 2-5-fold, 5-10-fold, 10-50-fold, and >50-fold. The basal RPKM values are shown, and the horizontal red lines indicate median RPKM values within each bin. (C) The distribution of maximum fold inductions relative to unstimulated (left), the peak RPKM value (top right), and the basal RPKM value (bottom right) for each of the 226 10-fold induced genes is shown. (D) The 226 10-fold induced genes were separated into primary and secondary response based on their expression in cycloheximide-treated (CHX) and *Ifnar-/-* BMDMs stimulated with lipid A. Genes were classified as secondary response if they were expressed less than 33% in CHX or 30% in *Ifnar-/-* samples. Of the 83 genes expressed at less than 33% in CHX-treated macrophages relative to wildtype, 74 were differentially expressed between CHX-treated and wildtype samples ($p<0.01$). The venn diagram indicates the number of genes affected by CHX treatment, the absence of IFNAR, or both.

Figure 2-2. Analysis of IFNAR-Independent and IFNAR-Dependent Secondary Response Genes

(A) The activation kinetics of secondary response genes from BMDMs stimulated at 5-minute intervals from 0-1 hour, and 2 hours is shown. The shades of blue indicate percentile values. The lipid A-induced secondary response genes were sorted based on their maximum percent

expression in *Ifnar-/-* BMDMs relative to C57Bl/6 BMDMs (purple column). The maximum percent expressions in *Myd88-/-, Trif-/-, and Irf3-/-* are shown in the three columns to the right. (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 of the secondary response genes in the IFN-dependence bins reach at least 10% of their maximum expression are indicated in the table. (D) The maximum fold induction of the IFNAR-independent genes in Pam3CSK4-stimulated (black) and lipid A-stimulated *Ifnar-/-* (purple) BMDMs is shown on the top panel, and the percent expression of the IFNAR-independent genes in Pam3CSK4-stimulated (black), lipid A-stimulated *Ifnar-/-* (purple), and lipid A-stimulated *Trif-/-* (orange) BMDMs relative to wildtype BMDMs stimulated with lipid A is shown on the bottom panel. The IFNAR-independent genes were defined as those that are induced 10-fold and expressed greater than 3 RPKM in the absence of IFNAR signaling, or expressed at greater than 50% of wildtype in *Ifnar-/-* BMDMs stimulated with lipid A or wildtype BMDMs stimulated with Pam3CSK4. (E) A scatterplot comparing the maximum RPKM values in Pam3CSK4-stimulated BMDMs (y-axis) and the maximum RPKM values in lipid A-stimulated BMDMs (x-axis) for the primary response (blue) and the IFNAR-independent secondary response (red) genes are shown. (F) Ingenuity Pathway Analysis was performed to identify the top functional annotations for the primary response, IFNAR-dependent secondary response, and IFNAR-independent secondary response. (G) The IFNAR-independent genes that are involved in the proliferation, differentiation, and activation of T lymphocytes, are colored based on their fold induction in the absence of IFNAR signaling.

Figure 2-3. IFNAR-Independent and IFNAR-Dependent Secondary Response Genes

An expanded version of Figure 2-2A 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.

Figure 2-4. Properties of Primary Response Genes

(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 relative to wildtype lipid A-stimulated BMDMs for the 135 primary response genes are shown. The horizontal dashed grey line indicates the 33% expression threshold used to call a gene as dependent or independent. The percent expression in (B) *Trif-/-* versus *Irf3-/-* or (C) *Trif-/-* versus *Myd88-/-* for the primary response genes are shown. TRIF lo (< 33% expression relative to wildtype) IRF3 hi (> 33% expression relative to wildtype) genes are shown in orange, and the TRIF lo (< 33% expression relative to wildtype) IRF3 lo (< 33% expression relative to wildtype) genes are shown in green. (D) The activation kinetics of the primary response genes from BMDMs stimulated at 5-minute intervals between 0-60 minutes, and 120 minutes are represented as log2 normalized and mean-centered RPKM values. The primary response genes 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 (cluster 2-5; < 33% in *Trif-/-* only), and MAPK-dependent (cluster 6-9; < 33% in MAPK inhibitor-treated samples). The remaining primary response genes were not dependent on any perturbation examined (cluster 10-16; > 33% in all perturbed datasets). The genes in each broad class were subclustered (k-means) based on their expression kinetics using Cluster3. To the right of the expression kinetic

heatmap shows the following properties for each gene: basal expression value (grey), fold induction magnitude (blue), promoter CpG-island (beige), and the maxiumum percent expression in *Myd88-/-* (red), *Trif-/-* (orange), *Irf3-/-* (green), and MAP kinase inhibitor-treated (light blue) BMDMs.

Figure 2-5. Highly Induced Primary Response Genes Ordered by Their Dependence on Various Signaling Pathways

An expanded version of Figure 2-4D is shown to include gene names for each 10-fold significantly induced primary response gene.

Figure 2-6. Promoter Motif Analysis of Primary Response Gene Clusters

Overrepresented transcription factor binding motifs are shown for each cluster, 1-16. The genes were clustered as described in Figure 2-4D. The transcription factor families are shown to the left, in alphabetical order. The color intensity is proportional to the negative log (p-value).

Figure 2-7. Identification of Putative NF–κB Target Genes Through Transcription Factor Binding Motif and ChIP-seq Analysis

(A) The protein binding microarray (PBM) z scores of p50:RelA (y-axis) and RelA ChIP-seq peak scores (x-axis) in the promoter (-500 to +150) for the primary response genes (right) and all remaining genes in the genome (left) were plotted. The genes were subcategorized into 2-10 fold induced (blue), not induced (red), secondary response (green), and low expression (grey). The horizontal dashed grey line indicates the PBM z score threshold (6.4), and the vertical

dashed grey line indicates the ChIP-seq peak score threshold (19). (B) Tables indicating the distribution of genes as shown in (A), as number of genes (left) and percent of genes (right) within each gene class. (C-F) Tables indicating the best matching κB motif (column 1), gene name (column 2), PBM p50:RelA z score (column 3), location of the motif from the TSS (column 4), RelA ChIP-seq peak score (column 5), and either the function or fold induction (column 6) for the primary response genes with strong κB motifs and strong RelA binding (C), strong kB motifs that do not support strong RelA binding (D), weak kB motifs at support strong RelA binding (E), and other NF-κB and IκB family members (F). (G) A line graph of the p50:RelA motif Z score enrichment for the primary response genes relative to the expressed but not induced genes.

Figure 2-8. The Position of RelA Peaks Relative to the Transcriptional Start Sites of All Genes

(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 2-9. Kinetic and Functional Analysis of Putative NF−κB Target Genes

(A) The 37 putative NF−κB target genes were grouped based on their role in mediating NF-κB signaling, MAPK dependence, or IRF3 dependence. The normalized expression values from 0-25 minutes (left panel) and 0-120 minutes (middle panel), and the fold change relative to the previous time point (right panel) are shown. To the right of the heatmaps, the basal expression, fold induction magnitude, promoter-CpG content, expression in *Rela-/-*, *Trif-/-*, *Irf3-/-*, and MAPK-inhibited BMDMs are shown from left to right. The presence of a p50:RelA motif based on PBM datasets and the RelA ChIP-seq binding peak scores at 0, 15, 30, 60, and 120 minutes of lipid A stimulation are indicated in the far right panels. (B) The primary response genes that exhibited similar activation kinetics and/or RelA dependence to the 37 putative NF−κB target genes are shown in the same layout as in (A). (C) The average activation kinetics of the NF-κB subgroups is shown as log2 fold inductions relative to basal during the 120-minute lipid A treatment period. (D) The average activation kinetics of the two additional clusters from Figure 2-9B (Cluster 5 and 6) are shown as log2 fold inductions relative to basal during the 120-minute lipid A treatment period.

Figure 2-10. Putative NF−κB Target Genes and the Genes That Exhibit Similar Kinetics and/or RelA-Dependence

An expanded version of Figure 2-9A and Figure 2-9B 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 2-11. Analysis of IRF3 and SRF Target Genes

(A) The primary response genes exhibiting dependence on IRF3 (< 33% expression in both *Irf3-/-* and *Trif-/-* macrophages) were separated based on their additional requirement for NF-κB for their activation. The colors indicate the percentile of the relative expression. To the right of the heatmap are columns indicating the basal RPKM level, fold induction magnitude relative to unstimulated, and promoter-CpG content for each gene. The rightmost heatmap indicates the RelA ChIP-seq binding peak scores for the indicated lipid A treatment time points. (B) The fold induction for each IRF3-dependent gene is shown over the 2 hour stimulation time period, grouped based on their additional requirement for NF-κB. (C) For each primary response gene, the higher maximum percent expression from either *Trif-/-* or *Irf3-/-* BMDMs (y-axis) was assessed against the best scoring IRF3 motif (x-axis) within the promoter based on the IRF Transfac position weight matrix (PWM). The five IRF3/NF-κB genes are highlighted in blue, and the four IRF3 genes are highlighted in green. The horizontal dashed grey line indicates the 33% expression threshold, and the vertical dashed grey line indicates the 90% Transfac score threshold. (D) For each IRF3-dependent gene, we identified the IRF3 and p50:RelA binding sites (for the IRF3/NF-κB groups of genes) as well as the position of each motif relative to the transcriptional start site. The spacing between the NF-κB and IRF3 motifs is indicated in the right column. (E) Scatterplot comparing the Transfac PWM scores of SRF binding motifs (y-axis) versus the SRF ChIP-seq peak scores (x-axis) in the promoter (-500 to +150) for the primary response genes (right) and all remaining genes in the genome (left) is shown. These genes in the right hand graph were subcategorized into 2-10 fold induced (blue), not induced (red), secondary response (green), and low expression (grey). The horizontal dashed grey line indicates the SRF motif threshold (90%), and the vertical dashed grey line indicates the SRF ChIP-seq peak score threshold (10). (F) Tables of the distribution of all annotated genes, shown as number of genes (left) and percent of genes (right) within each class. (G) The log2

normalized expression values from 0-25 minutes (first panel), 0-120 minutes (second panel), and the fold induction relative to the expression level at the previous time point (third panel) for the seven putative SRF target genes are shown. To the right of the heatmaps are columns indicating the basal expression level, fold induction magnitude, promoter-CpG content, and MAPK dependence for each gene. (H) Genes that exhibited similar activation kinetics as the putative SRF target genes are shown, with the same layout as in Figure 2-11G.

Figure 2-12. Regulation of Transiently Transcribed Genes by SRF

(A) The activation kinetics of the seven putative SRF target genes is shown. The values represent fold induction magnitudes relative to unstimulated. (B) The two genes that exhibited similar activation kinetics as the seven putative SRF targets were examined on UCSC Genome Browser to identify distal SRF binding peaks. RelA binding peaks were also examined for these two genes. The transcription start sites of the genes are indicated as red arrows, and the green rectangles indicate CpG-islands. (C) The log2 fold induction relative to unstimulated samples in control (black) and MAPK inhibitor-treated (blue) BMDMs for the seven putative SRF targets are shown. (D) For each annotated gene in each gene category (primary, secondary, 2-10 fold induced, not induced but expressed, and unexpressed), SRF 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 upstream and downstream from the TSS. The SRF peaks were then grouped based on their ChIP-seq peak score (>10 or < 10). If a gene did not have a peak in the indicated region, a score of 0 was assigned to that gene for that region. 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.

Figure 2-13. Classification of Lipid A-Induced Genes

(A) 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, *Trif-/-*, and *Irf3-/-* BMDMs. (B) The 94 secondary response genes were grouped based on their dependence on IFNAR. 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: expression in *Ifnar-/-* BMDMs, and expression in Pam3CSK4-stimulated WT BMDMs.

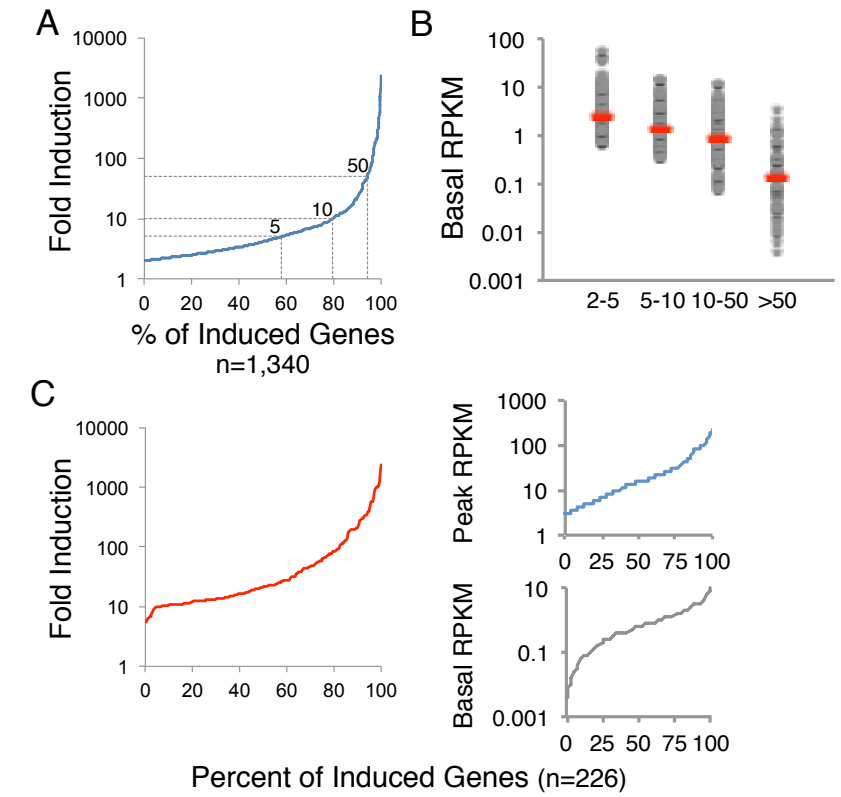
Figure 2-14. Final Classification of the Primary Response Genes

An expanded version of Figure 2-13A is shown to include gene names within each class of primary response genes.

Figure 2-15. Final Classification of the Secondary Response Genes

An expanded version of Figure 2-13B is shown to include gene names within each class of secondary response genes.

Figure 2-1. Properties of the Lipid A-Induced Transcriptional Cascade



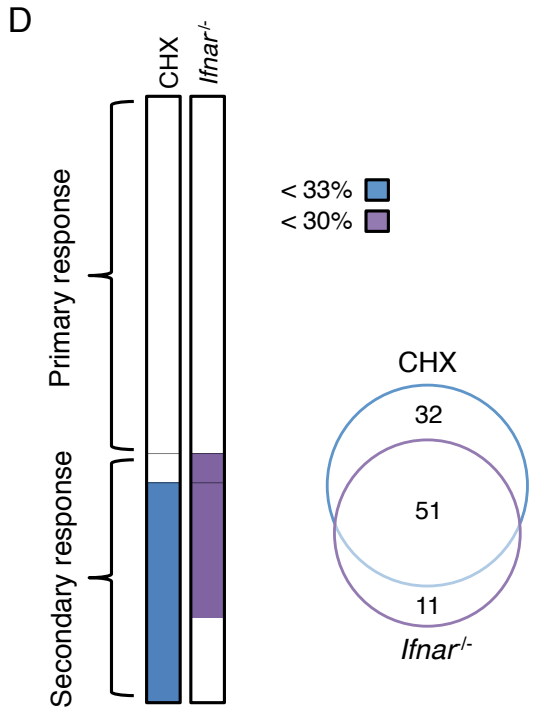
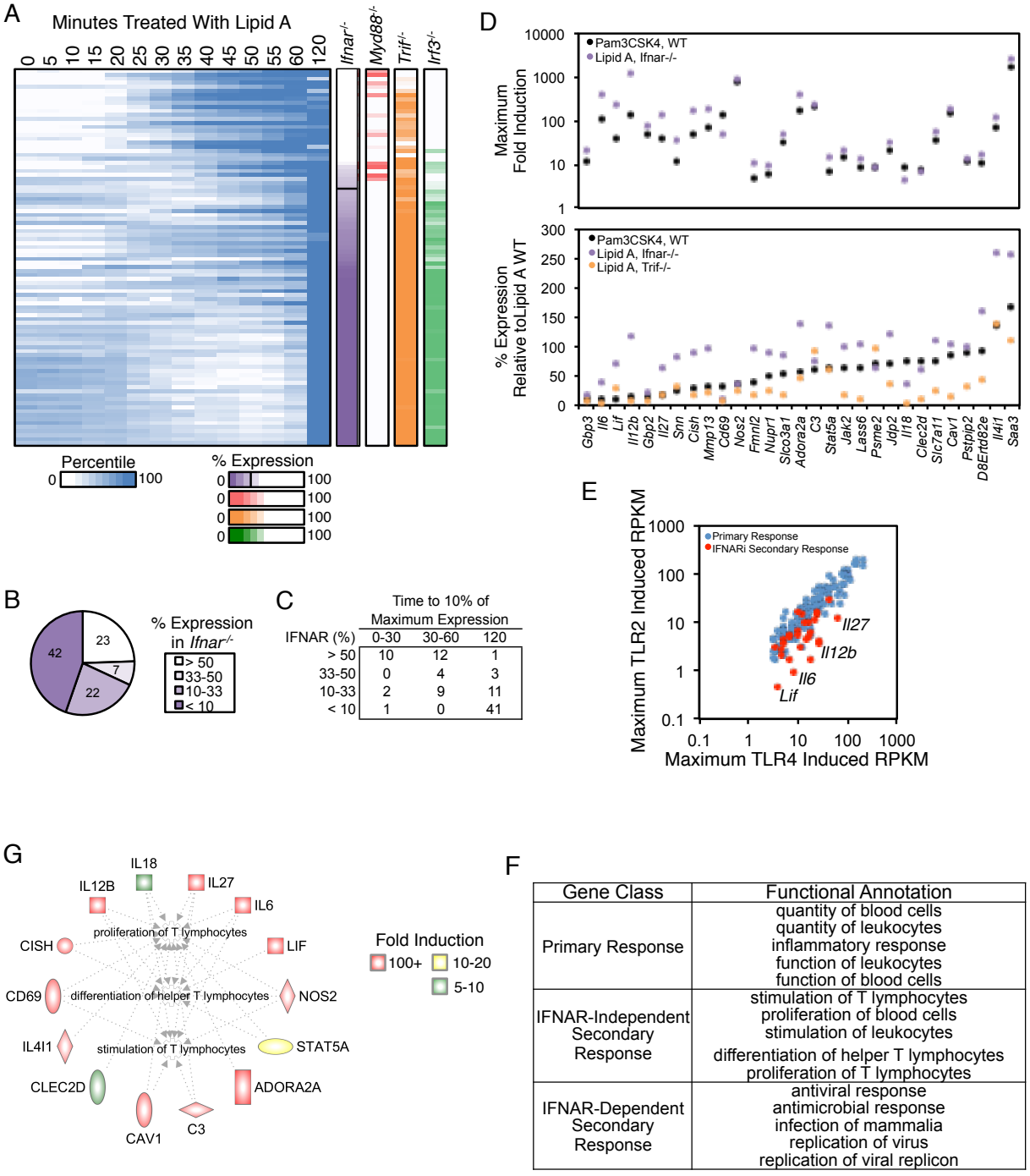


Figure 2-2. Analysis of IFNAR-Independent and IFNAR-Dependent Secondary Response Genes



F

Gene Class	Functional Annotation
Primary Response	quantity of blood cells quantity of leukocytes inflammatory response function of leukocytes function of blood cells
IFNAR-Independent Secondary Response	stimulation of T lymphocytes proliferation of blood cells stimulation of leukocytes differentiation of helper T lymphocytes proliferation of T lymphocytes
IFNAR-Dependent Secondary Response	antiviral response antimicrobial response infection of mammalia replication of virus replication of viral replicon

Figure 2-3. IFNAR-Independent and IFNAR-Dependent Secondary Response Genes

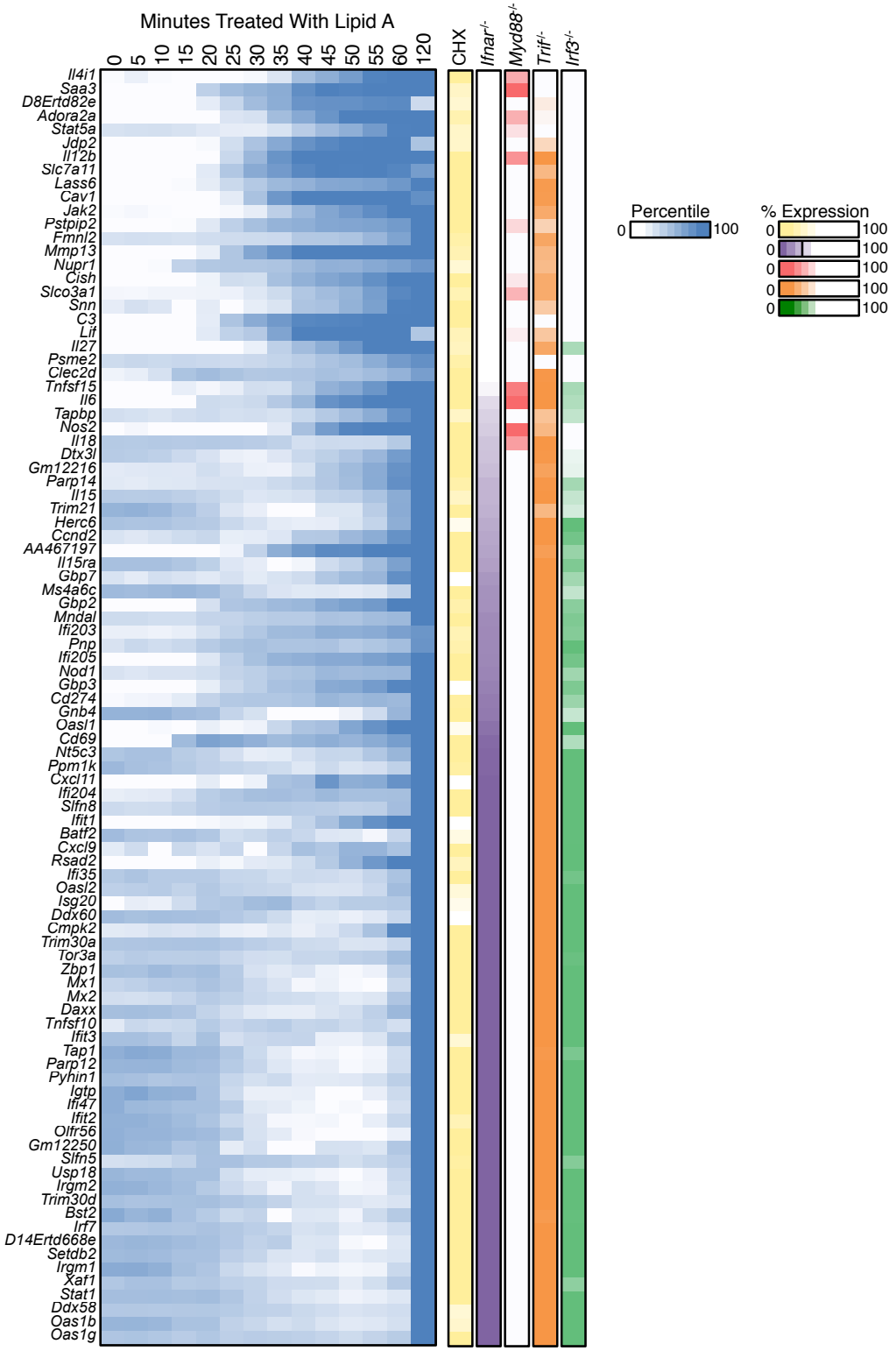


Figure 2-4. Properties of Primary Response Genes

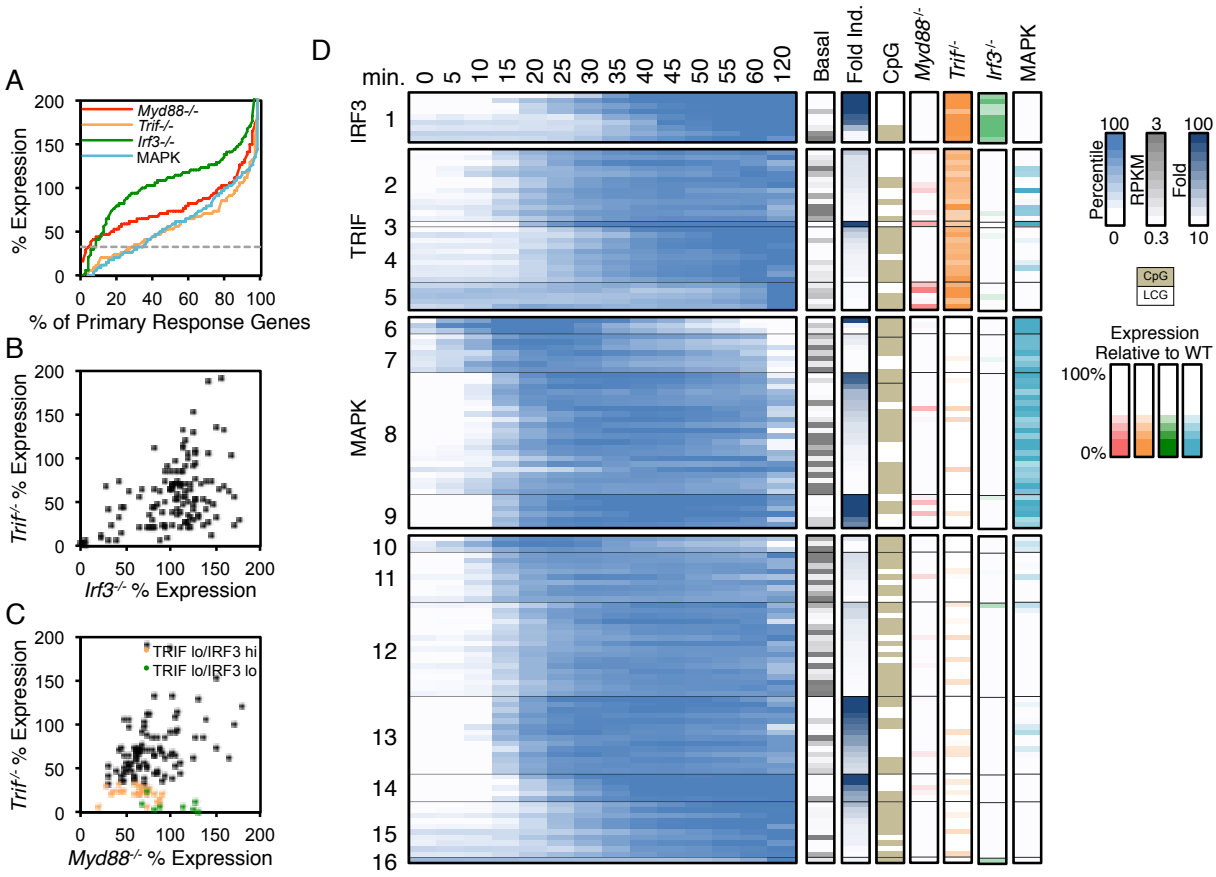


Figure 2-5. Highly Induced Primary Response Genes Ordered by Their Dependence on Various Signaling Pathways

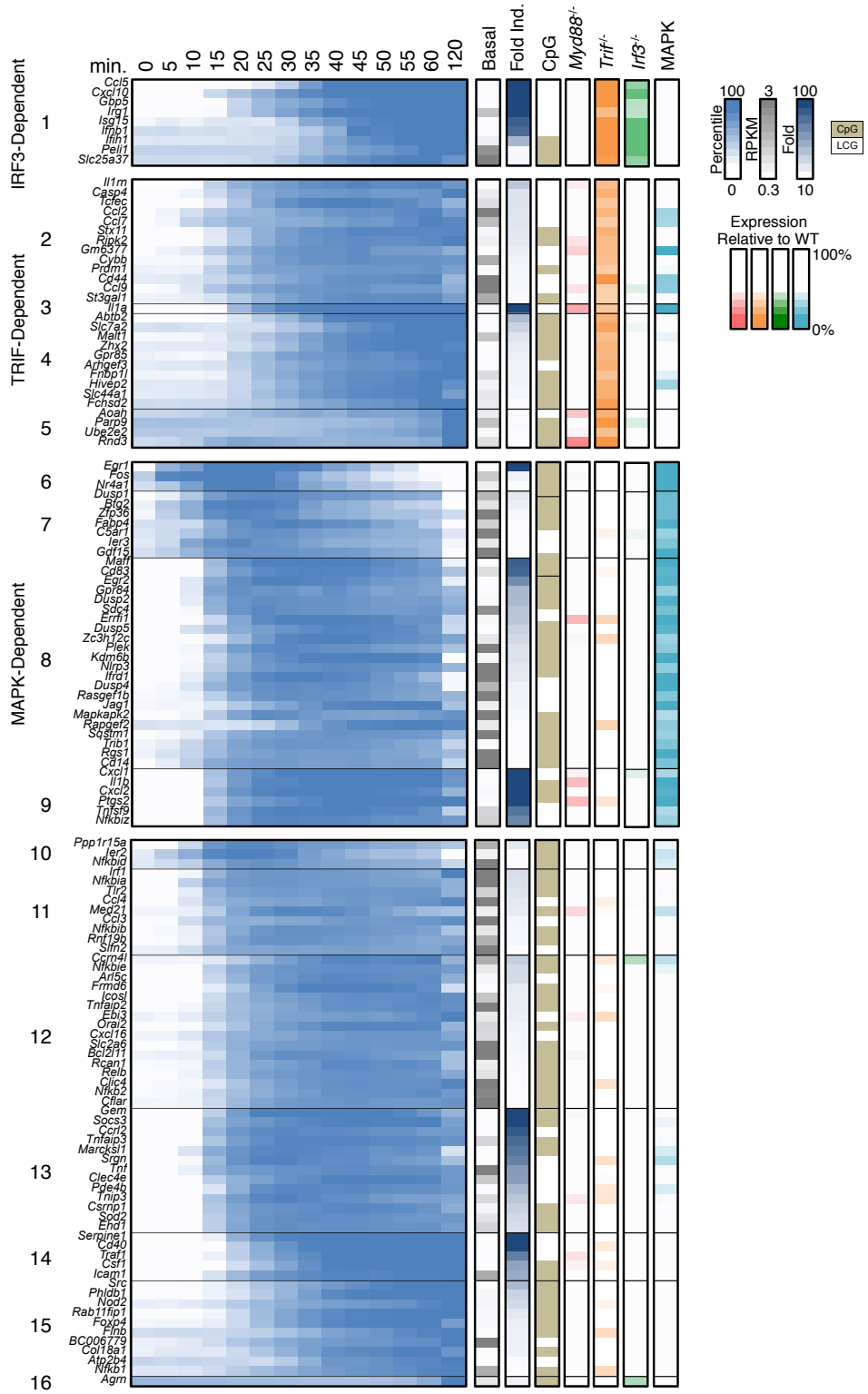


Figure 2-6. Promoter Motif Analysis of Primary Response Gene Clusters

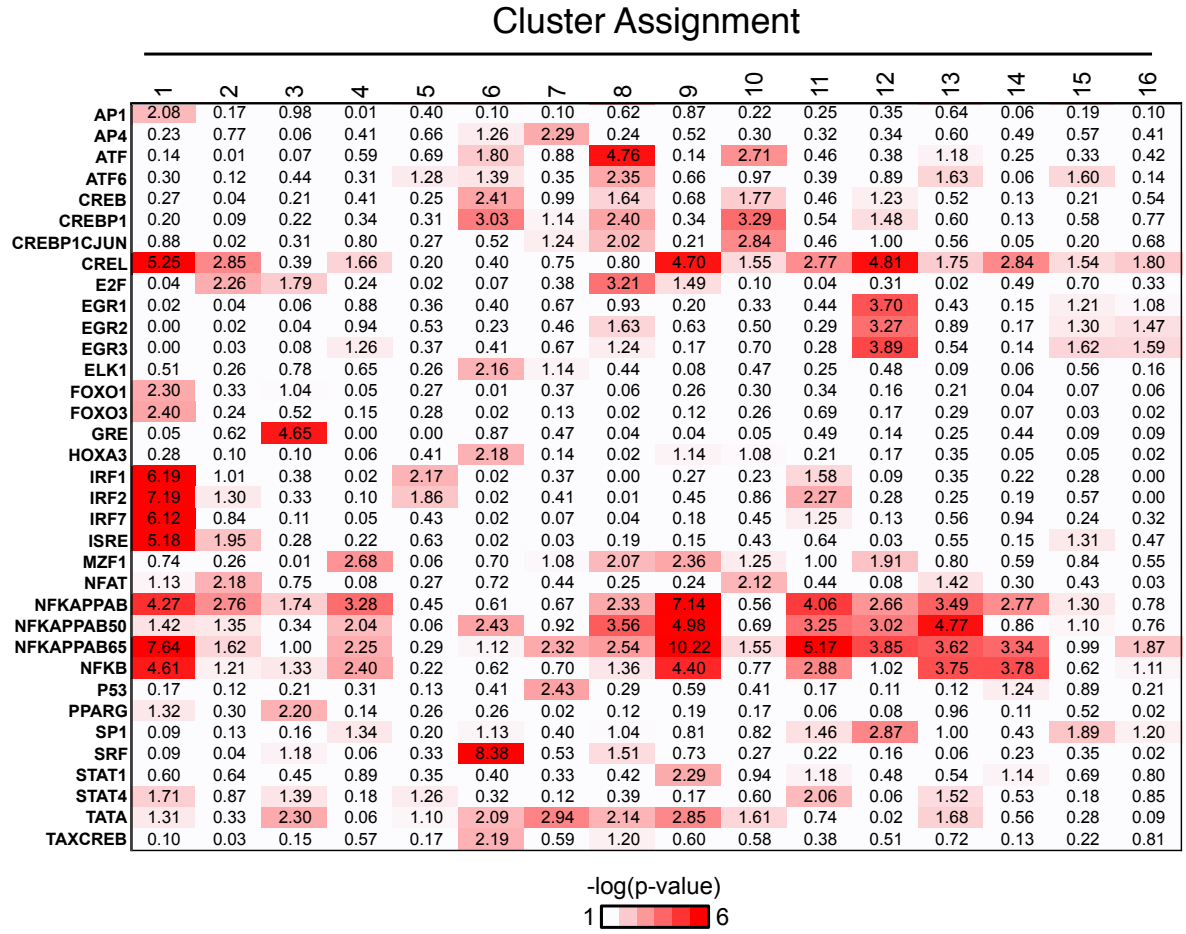
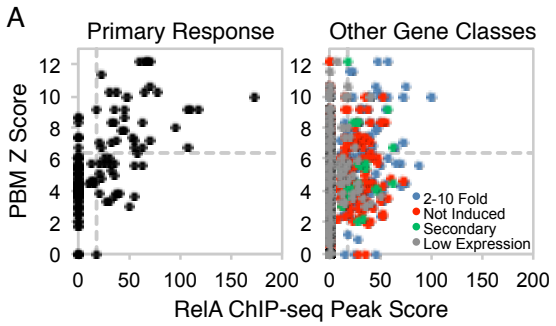


Figure 2-7. Identification of Putative NF-κB Target Genes Through Transcription Factor Binding Motif and ChIP-seq Analysis

A



B

Number of Genes

PBM	ChIP-seq 0	ChIP-seq < 19	ChIP-seq > 19
Primary (n=132)			
> 6.4	7	0	37
< 6.4	62	6	20
Secondary (n=94)			
> 6.4	7	1	5
< 6.4	65	4	12
2-10 Fold (n=732)			
> 6.4	27	6	43
< 6.4	571	21	64
Not Induced (n=1723)			
> 6.4	87	9	27
< 6.4	1497	38	65
Low Expression (n=18487)			
> 6.4	1028	56	100
< 6.4	16875	165	263

Percent of Genes

PBM	ChIP-seq 0	ChIP-seq < 19	ChIP-seq > 19
Primary (n=132)			
> 6.4	5.3	0.0	28.0
< 6.4	47.0	4.5	15.2
Secondary (n=94)			
> 6.4	7.4	1.1	5.3
< 6.4	69.1	4.3	12.8
2-10 Fold (n=732)			
> 6.4	3.7	0.8	5.9
< 6.4	78.0	2.9	8.7
Not Induced (n=1723)			
> 6.4	5.0	0.5	1.6
< 6.4	86.9	2.2	3.8
Low Expression (n=18487)			
> 6.4	5.6	0.3	0.5
< 6.4	91.3	0.9	1.4

C

Motifs With RelA Binding

Sequence	Gene	PBM z score	PBM TSS	ChIP-seq Score	Function
GGGGAATTCC	*Tlr2*	12.2	-173	70.9	Transmembrane receptor
	Nfkbib	12.2	8	67.7	**IκB family**
	Sdc4	12.2	-82	67.3	Signaling molecule
	Relb	12.2	-24	59.8	**NF-κB family**
	Nfkbid	12.2	-218	65.6	**IκB family**
GGGAATTCCC	*Bcl2l11*	11.5	-314	23.6	Apoptosis regulator
GGGGGTTTCC	*Ebi3*	10.6	-310	69.2	Cytokine
GGGAAATTCC	*Cxcl2*	10.3	-96	78.6	Chemokine
	Cxcl1	10.3	-70	29.6	Chemokine
	Cxcl10	10.3	-159	66.1	Chemokine
	Ifnb1	10.3	-116	38.3	Cytokine
GGGGATTTCC	*Tnfaip3*	10.0	-12	173.0	**NF-κB signaling**
	Casp4	10.0	-104	49.0	Inflammasome regulation
GGGAATTTCC	*Nfkb1*	9.2	31	117.3	**NF-κB family**
	Nfkbia	9.2	-29	109.7	**IκB family**
	Nfkbiz	9.2	-228	108.7	**IκB family**
	Fchsd2	9.2	-298	46.0	Signaling molecule
	Gbp5	9.2	-77	35.7	Inflammasome regulation
GGGGATTCCC	*Ptgs2*	9.1	-428	19.0	Enzyme
GGGGGATTCC	*Nfkbie*	8.6	-122	45.5	**IκB family**
GGGACTTTCC	*Nfkb2*	8.4	10	84.2	**NF-κB family**
	Csf1	8.4	-154	33.6	Cytokine
GGGGTTTTCC	*Stx11*	8.4	-284	62.7	Transporter
AGGGGGTTTT	*Rapgef2*	8.0	-292	96.4	Signaling molecule
TGGAATTTCC	*Ier3*	7.9	-140	44.3	Apoptosis regulator
	Irg1	7.9	-85	43.5	Enzyme
GGGGGATTTC	*Kdm6b*	7.4	-134	53.8	Transcription regulator
GGGATTTCCC	*Tnip3*	7.2	-54	71.3	**IκB regulator**
	Cd40	7.2	-83	55.2	Transmembrane receptor
TGGAAATTCC	*Icam1*	7.2	-174	38.0	Transmembrane receptor
AGGGGATTTT	*Traf1*	7.1	-106	20.9	Signaling molecule
GGGGAATCCC	*Irf1*	7.0	-13	33.8	Transcription regulator
GGGAGTTTCC	*Ccl5*	6.8	-90	54.7	Chemokine
AGGGGAATTT	*Gpr84*	6.8	-364	108.5	Signaling molecule
GGGAAAGTCC	*Btg2*	6.7	24	22.4	Transcription regulator
	Ccrl2	6.7	-316	63.9	Chemokine Receptor
GGGACTTCCC	*Cd83*	6.4	29	56.8	Transmembrane receptor

D

Motifs Without RelA Binding

Sequence	Gene	PBM z score	PBM TSS	ChIP-seq Score	Function
AGGGAATTCC	*Serpine1*	8.6	-310	0	Other
TGGGAATTCC	*Agrn*	8.6	-395	0	Signaling molecule
GGGACTTTCC	*Phldb1*	8.4	137	0	Other
GGGGCTTTCC	*Ccl9*	7.3	-35	0	Cytokine
	Hivep2	7.3	-221	0	Transcription regulator
GGGGCGCCCC	*Src*	7.3	-218	0	Kinase
GGGGTTCCCC	*Sod2*	6.7	-83	0	Enzyme

E

Weak Motifs With RelA Binding

Sequence	Gene	PBM z score	PBM TSS	ChIP-seq Score	Function
AGGGGGAAAA	*Zhx2*	6.1	36	25.3	Transcription regulator
TGGGGAAAAA	*Tcfec*	6	-197	29.9	Transcription regulator
CGGGGATTTC	*Cxcl16*	5.8	128	13.8	Chemokine
GGGTATTTCC	*Dusp5*	5.7	-263	24.7	Phosphatase
GGGTTTTCCC	*Ier2*	5.7	-290	35.6	Other
	Cd44	5.7	-450	21.5	Enzyme
	Rasgef1b	5.7	109	29.9	Signaling molecule
TGGGGAAATT	*Il1rn*	5.6	-160	62.3	Cytokine
AGGGGGAAAT	*Il1b*	5.6	-275	55.4	Cytokine
GGGTTTCCCC	*Slc2a6*	5.2	-143	32.5	Transporter
TGGGGAAAAC	*Srgn*	5.1	-274	20.8	Other
GGGAAAATTT	*Ccl3*	4.9	-63	35.0	Chemokine
GGGGAGGGAA	*Tnfsf9*	4.8	-32	22.6	Cytokine
GGGGACGGGG	*Nr4a1*	4.7	-159	15.1	Nuclear receptor
TGGGGTTTTC	*Cflar*	4.6	-116	14.1	Apoptosis regulator
GGGTGGGAAA	*Gm6377*	4.5	-156	11.6	Other
GGGGCTTTTC	*Gem*	4.5	-62	16.4	Enzyme
GGGGGAAGAC	*Ccl4*	3.9	-165	19.4	Chemokine
	Cybb	3.9	-403	41.4	Enzyme
GGGATTCCCA	*Slfn2*	3.9	-278	37.6	Other
TGGGGAAATG	*Malt1*	3.7	-251	55.9	Peptidase
GGGGGAGAAC	*Tnf*	3.6	-451	35.8	Cytokine
GGGGAAAAGG	*C5ar1*	3.4	106	26.9	G-protein coupled receptor
	Plek	3.4	-32	29.4	Signaling molecule
GGAAATTCTG	*Clec4e*	3	-54	49.9	Transmembrane receptor
-	*Aoah*	0	-	17.4	Enzyme

F

Other NF-κB and IκB Members

Sequence	Gene	PBM z score	PBM TSS	ChIP-seq Score	Fold Induction
AGGGGATTTC	*Rela*	6.5	-46	33.0	3.6
GGGGATTTCC	*Rel*	10	-90	72.2	7.9
GGGACTTCCC	*Bcl3*	6.4	2	39.9	7.5

G

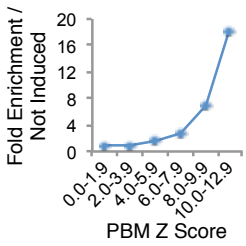


Figure 2-8. The Position of RelA Peaks Relative to the Transcriptional Start Sites of All Genes

A

Peak Locations Relative to Annotated Genes

Number of Genes

Gene Category	Promoter	10kb	20kb	100kb	> 100kb	ChIP-seq Peak Score
Primary (n=132)	69	56	102	84	121	0
	6	9	7	15	4	< 19
	57	67	23	33	7	> 19
Secondary (n=94)	72	64	82	75	91	0
	5	10	4	5	1	< 19
	17	20	8	14	2	> 19
2 to 10 Fold (n=732)	598	428	577	490	681	0
	26	85	59	88	20	< 19
	108	219	96	154	31	> 19
Not Induced (n=1723)	1584	1333	1559	1450	1673	0
	47	156	77	113	19	< 19
	92	234	87	160	31	> 19
Low Expression (n=18487)	17903	17379	17866	17313	18151	0
	221	569	325	578	173	< 19
	363	539	296	596	163	> 19

Percent of Genes

Gene Category	Promoter	10kb	20kb	100kb	> 100kb	ChIP-seq Peak Score
Primary (n=132)	52.3	42.4	77.3	63.6	91.7	0
	4.5	6.8	5.3	11.4	3.0	< 19
	43.2	50.8	17.4	25.0	5.3	> 19
Secondary (n=94)	76.6	68.1	87.2	79.8	96.8	0
	5.3	10.6	4.3	5.3	1.1	< 19
	18.1	21.3	8.5	14.9	2.1	> 19
2 to 10 Fold (n=732)	81.7	58.5	78.8	66.9	93.0	0
	3.6	11.6	8.1	12.0	2.7	< 19
	14.8	29.9	13.1	21.0	4.2	> 19
Not Induced (n=1723)	91.9	77.4	90.5	84.2	97.1	0
	2.7	9.1	4.5	6.6	1.1	< 19
	5.3	13.6	5.0	9.3	1.8	> 19
Low Expression (n=18487)	96.8	94.0	96.6	93.6	98.2	0
	1.2	3.1	1.8	3.1	0.9	< 19
	2.0	2.9	1.6	3.2	0.9	> 19

B

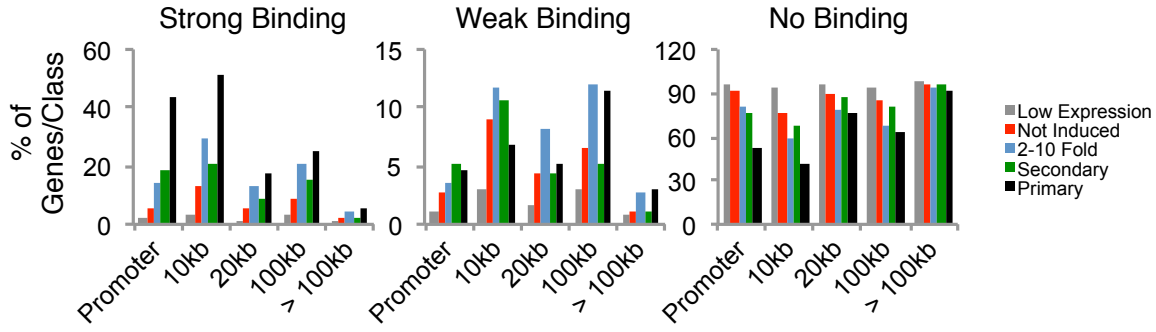


Figure 2-9. Kinetic and Functional Analysis of Putative NF-κB Target Genes

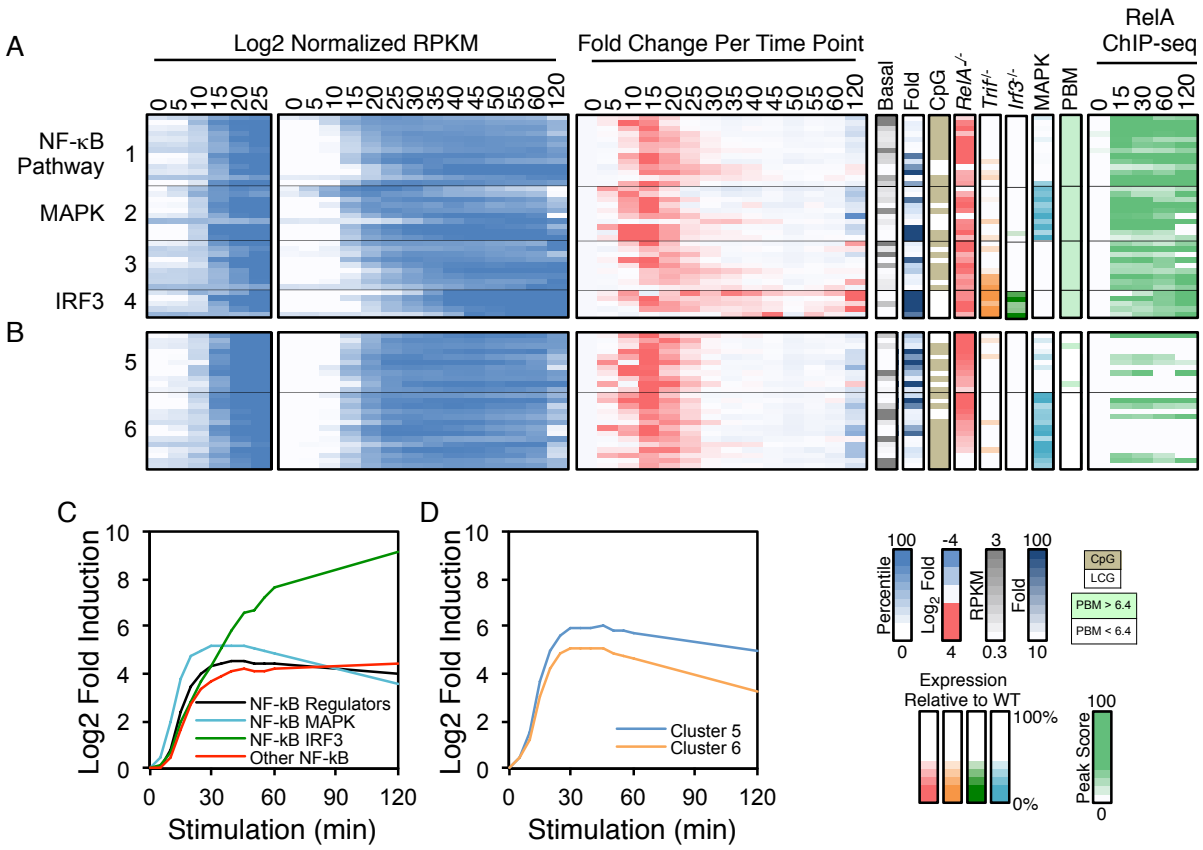


Figure 2-10. Putative NF-κB Target Genes and the Genes That Exhibit Similar Kinetics and/or RelA-Dependence

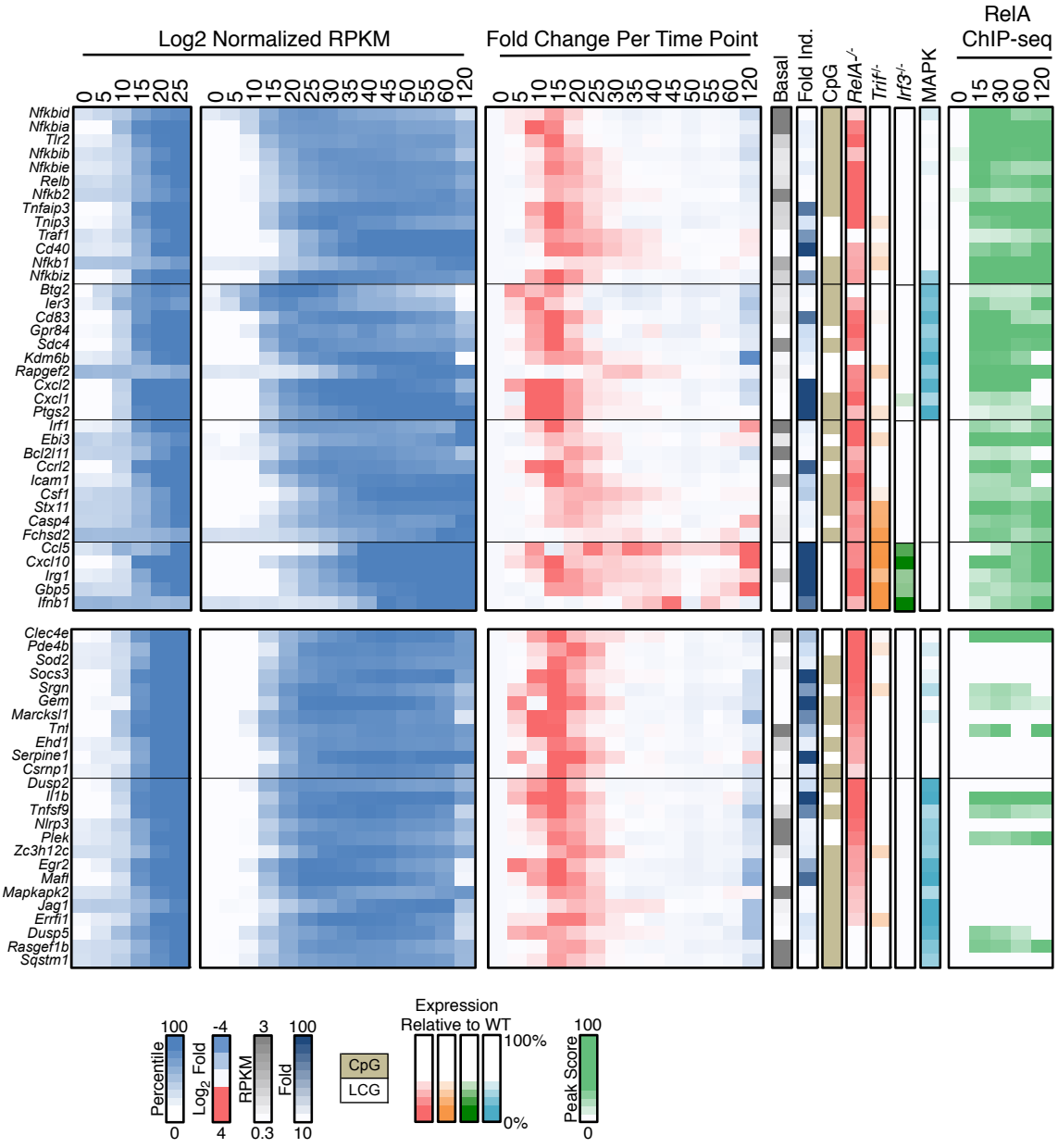
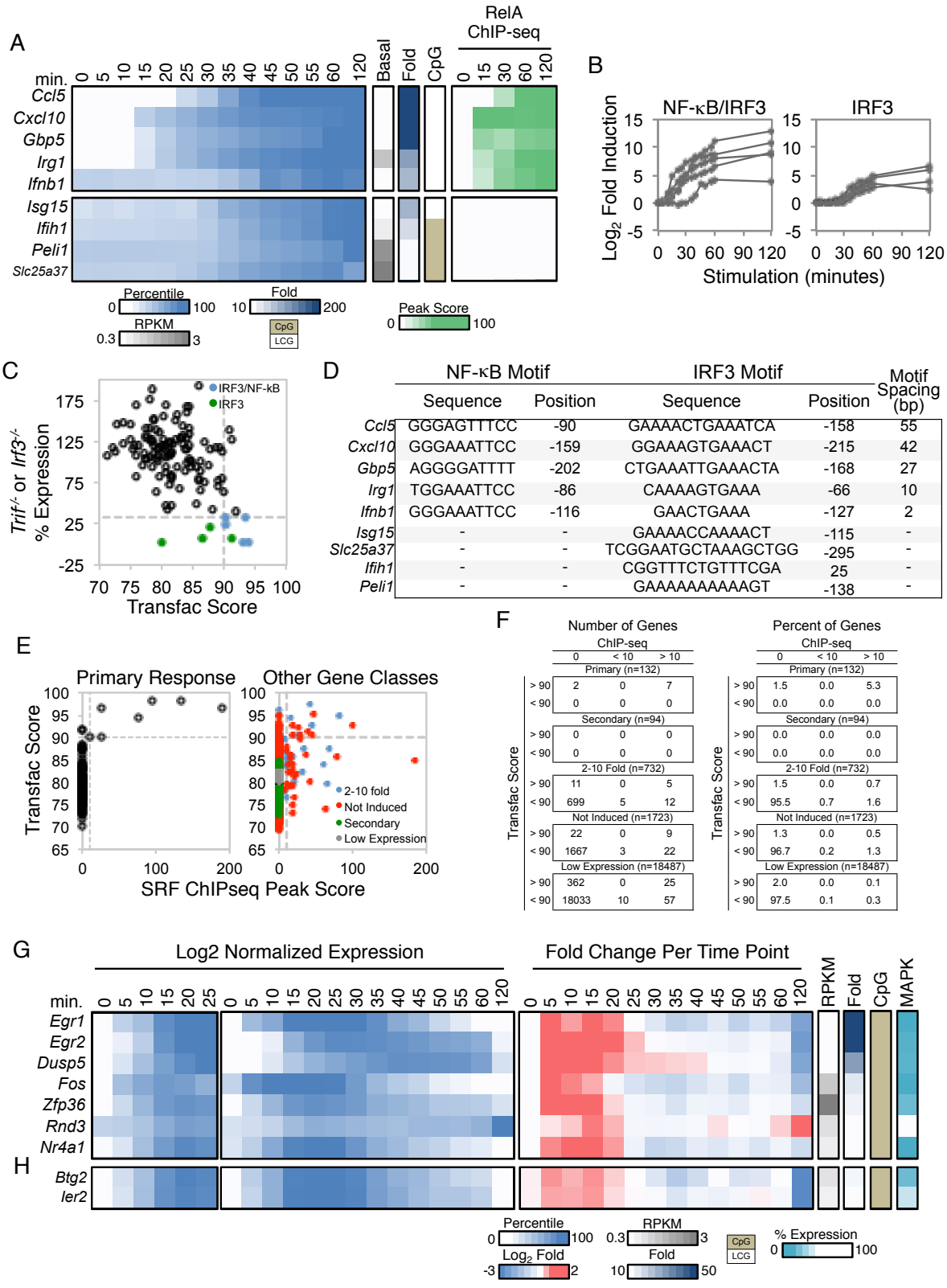
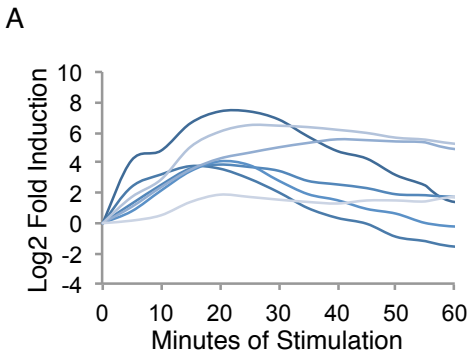


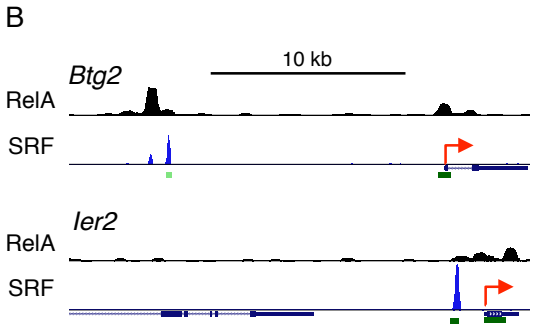
Figure 2-11. Analysis of IRF3 and SRF Target Genes

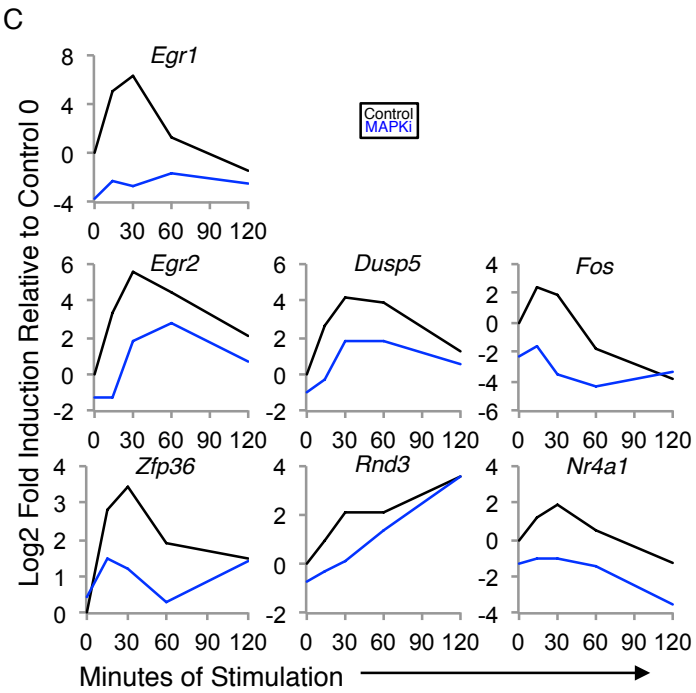


	NF-κB Motif Sequence	NF-κB Motif Position	IRF3 Motif Sequence	IRF3 Motif Position	Motif Spacing (bp)
Ccl5	GGGAGTTTCC	-90	GAAAACTGAAATCA	-158	55
Cxcl10	GGGAAATTCC	-159	GGAAAGTGAAACT	-215	42
Gbp5	AGGGGATTTT	-202	CTGAAATTGAAACTA	-168	27
Irg1	TGGAAATTCC	-86	CAAAAGTGAAA	-66	10
Ifnb1	GGGAAATTCC	-116	GAACTGAAA	-127	2
Isg15	-	-	GAAAACCAAAACT	-115	-
Slc25a37	-	-	TCGGAATGCTAAAGCTGG	-295	-
Ifih1	-	-	CGGTTTCTGTTTCGA	25	-
Peli1	-	-	GAAAAAAAAAAGT	-138	-

Figure 2-12. Regulation of Transiently Transcribed Genes by SRF







D

Number of Genes

Gene Category	Promoter	10kb	20kb	100kb	> 100kb	
Primary (n=132)	125	126	126	125	132	0
	0	1	0	0	0	< 10
	7	5	6	7	0	> 10
Secondary (n=94)	94	92	94	93	94	0
	0	0	0	0	0	< 10
	0	2	0	1	0	> 10
2 to 10 Fold (n=732)	710	704	721	707	725	0
	5	3	1	1	0	< 10
	17	25	10	24	7	> 10
Not Induced (n=1723)	1689	1664	1709	1689	1720	0
	3	1	1	1	0	< 10
	31	58	13	33	3	> 10
Low Expression (n=18487)	18395	18372	18449	18378	18467	0
	10	7	2	2	0	< 10
	82	108	36	107	20	> 10

Percent of Genes

Gene Category	Promoter	10kb	20kb	100kb	> 100kb	
Primary (n=132)	94.7	95.5	95.5	94.7	100.0	0
	0.0	0.8	0.0	0.0	0.0	< 10
	5.3	3.8	4.5	5.3	0.0	> 10
Secondary (n=94)	100.0	97.9	100.0	98.9	100.0	0
	0.0	0.0	0.0	0.0	0.0	< 10
	0.0	2.1	0.0	1.1	0.0	> 10
2 to 10 Fold (n=732)	97.0	96.2	98.5	96.6	99.0	0
	0.7	0.4	0.1	0.1	0.0	< 10
	2.3	3.4	1.4	3.3	1.0	> 10
Not Induced (n=1723)	98.0	96.6	99.2	98.0	99.8	0
	0.2	0.1	0.1	0.1	0.0	< 10
	1.8	3.4	0.8	1.9	0.2	> 10
Low Expression (n=18487)	99.5	99.4	99.8	99.4	99.9	0
	0.1	0.0	0.0	0.0	0.0	< 10
	0.4	0.6	0.2	0.6	0.1	> 10

Figure 2-13. Classification of Lipid A-Induced Genes

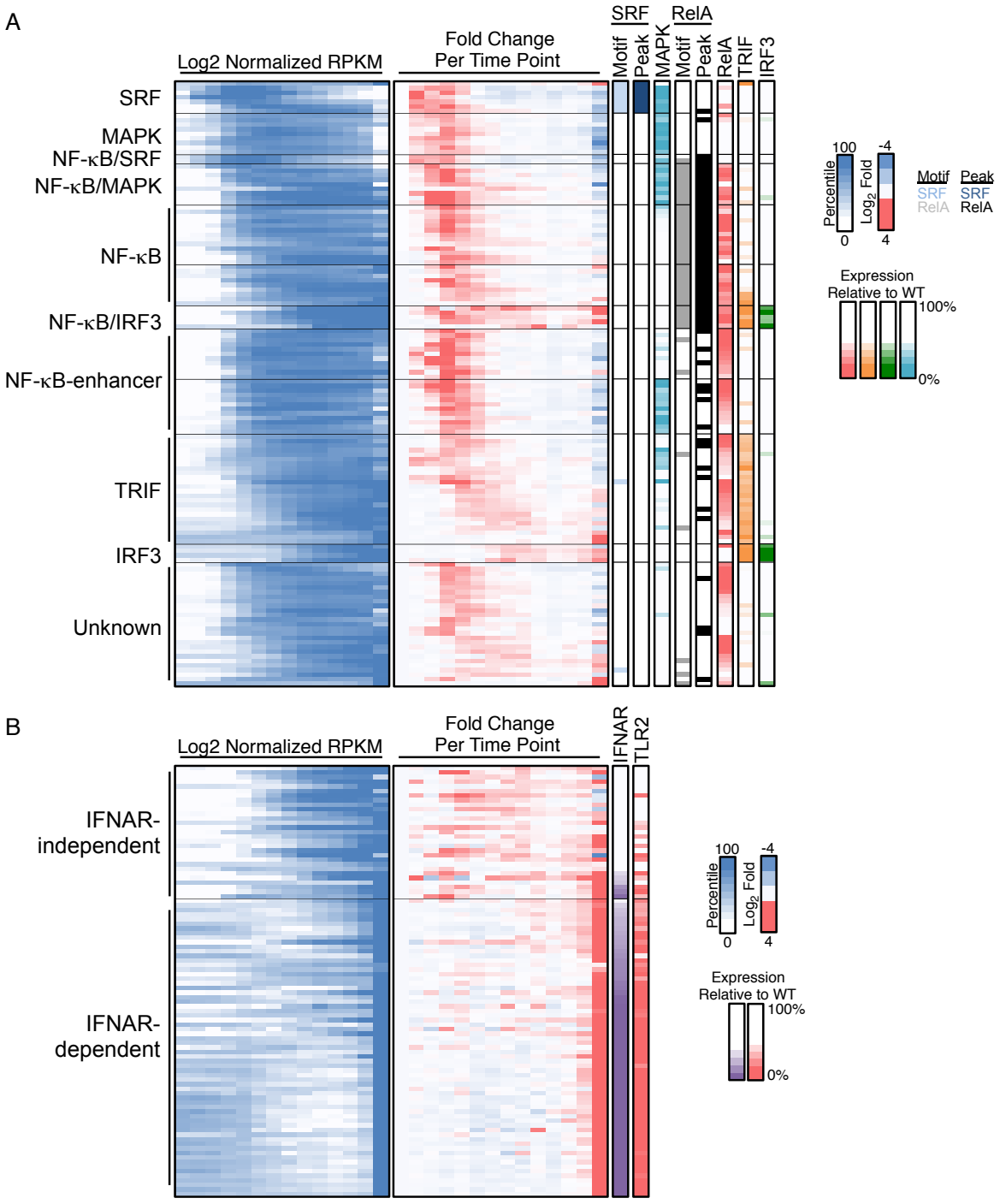


Figure 2-14. Final Classification of the Primary Response Genes

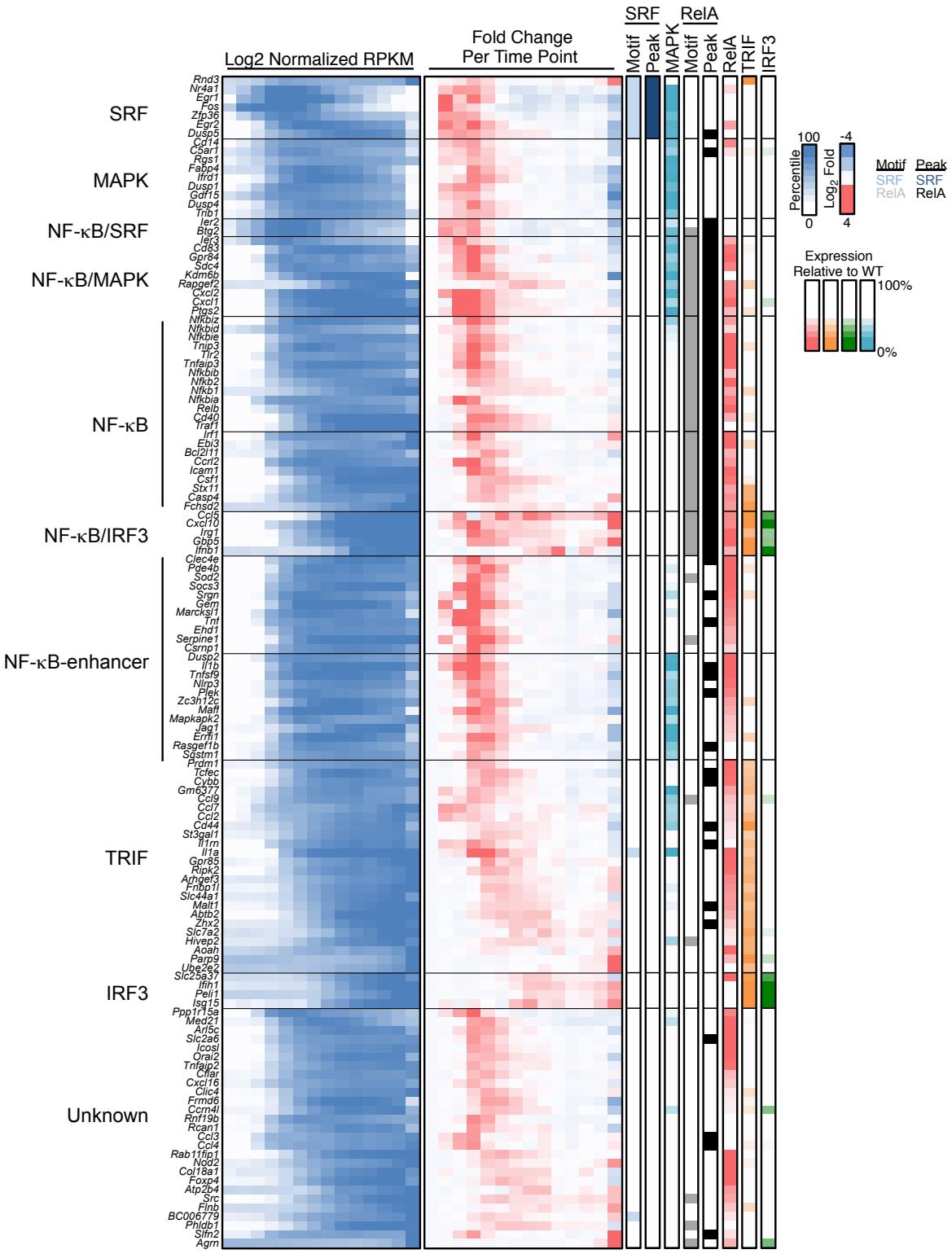
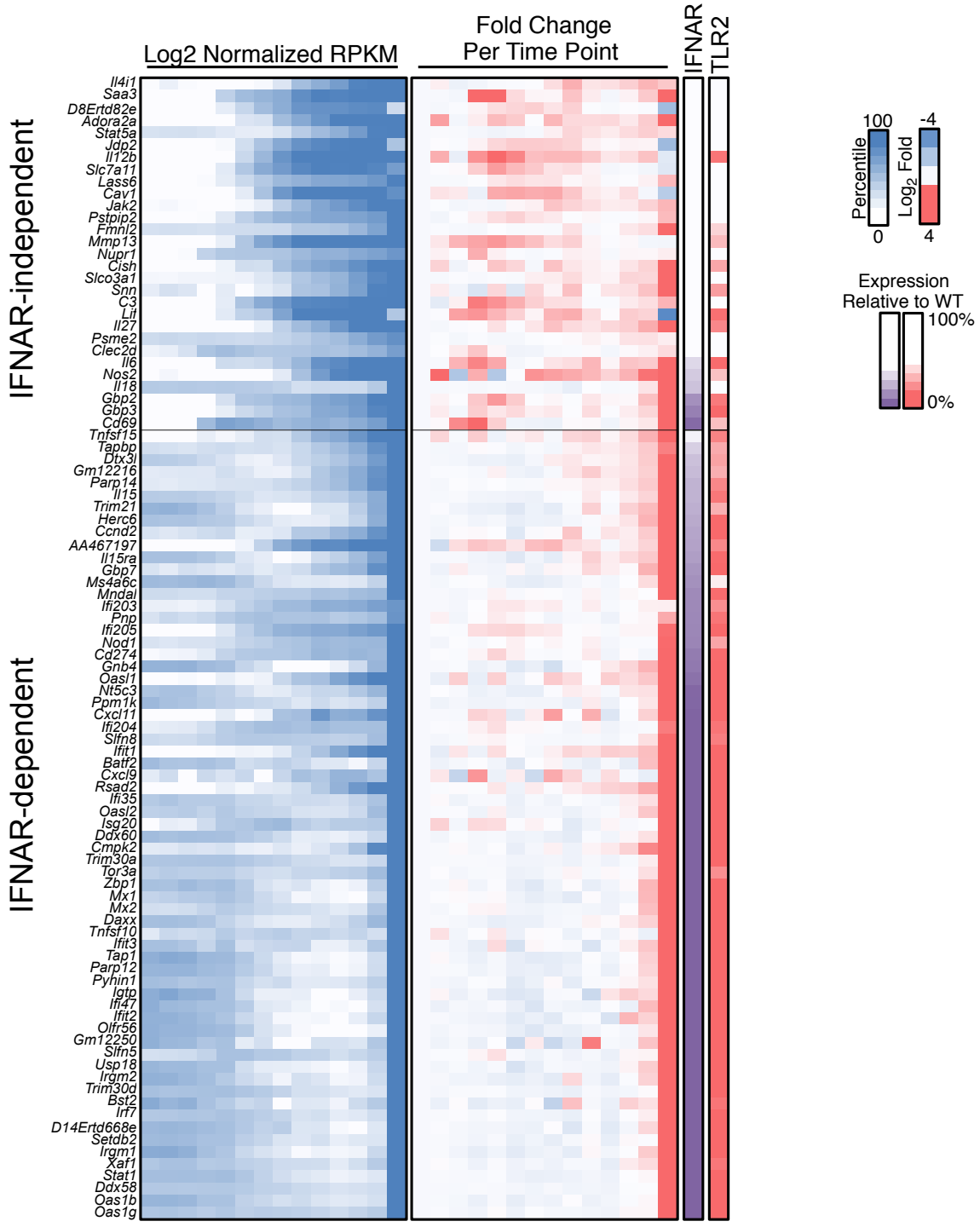


Figure 2-15. Final Classification of the Secondary Response Genes



Works Cited

1. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18,** 1509–1517 (2008).

2. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–8 (2008).

3. Jenner, R. G. & Young, R. A. Insights into host responses against pathogens from transcriptional profiling. *Nat. Rev. Microbiol.* **3,** 281–294 (2005).

4. Nilsson, R. *et al.* Transcriptional network dynamics in macrophage activation. *Genomics* **88,** 133–142 (2006).

5. Elkon, R., Linhart, C., Halperin, Y., Shiloh, Y. & Shamir, R. Functional genomic delineation of TLR-induced transcriptional networks. *BMC Genomics* **8,** 394 (2007).

6. Ramsey, S. A. *et al.* Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS Comput. Biol.* **4,** (2008).

7. Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326,** 257–63 (2009).

8. Garber, M. *et al.* A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals. *Mol. Cell* **47,** 810–822 (2012).

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

10. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469,** 368–373 (2011).

11. Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* **29,** 436–442 (2011).

12. Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150,** 279–290 (2012).

13. Cao, Y. *et al.* Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell* **18,** 662–74 (2010).

14. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13,** 613–26 (2012).

15. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43,** 264–8 (2011).

16. Li, X.-Y. *et al.* The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol.* **12,** R34 (2011).

17. MacArthur, S. *et al.* Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* **10,** R80 (2009).

18. Cusanovich, D. A., Pavlovic, B., Pritchard, J. K. & Gilad, Y. The functional consequences of variation in transcription factor binding. *PLoS Genet.* **10,** e1004226 (2014).

19. Suzuki, H. *et al.* The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.* **41,** 553–62 (2009).

20. Xu, X. *et al.* A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Res.* **17,** 1550–1561 (2007).

21. Kagan, J. C. *et al.* TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. *Nat. Immunol.* **9,** 361–8 (2008).

22. Toshchakov, V. *et al.* TLR4, but not TLR2, mediates IFN-beta-induced STAT1alpha/beta-dependent gene expression in macrophages. *Nat. Immunol.* **3,** 392–8 (2002).

23. Kayama, H. *et al.* Class-specific regulation of pro-inflammatory genes by MyD88 pathways and I??B?? *J. Biol. Chem.* **283,** 12468–12477 (2008).

24. Siggers, T. *et al.* Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. *Nat. Immunol.* **13,** 95–102 (2012).

25. Wong, D. *et al.* Extensive characterization of NF-κB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. *Genome Biol.* **12,** R70 (2011).

26. Wang, V. Y.-F. *et al.* The transcriptional specificity of NF-κB dimers is coded within the κB DNA response elements. *Cell Rep.* **2,** 824–39 (2012).

27. Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138,** 114–28 (2009).

28. Agalioti, T. *et al.* Ordered Recruitment of Chromatin Modifying and General Transcription Factors to the IFN-β Promoter. *Cell* **103,** 667–678 (2000).

29. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129,** 1111–23 (2007).

30. Freaney, J., Kim, R., Mandhana, R. & Horvath, C. Extensive cooperation of immune master regulators IRF3 and NFκB in RNA Pol II recruitment and pause release in human innate antiviral transcription. *Cell Rep.* **4,** 959–973 (2013).

31. Price, M. A., Rogers, A. E. & Treisman, R. Comparative analysis of the ternary complex factors Elk-1, SAP-1a and SAP-2 (ERP/NET). *EMBO J.* **14,** 2589–601 (1995).

32. Whitmarsh, A., Shore, P., Sharrocks, A. & Davis, R. Integration of MAP kinase signal transduction pathways at the serum response element. *Science* **269,** 403–407 (1995).

33. Cahill, M. A., Janknecht, R. & Nordheim, A. Signalling pathways: Jack of all cascades. *Curr. Biol.* **6,** 16–19 (1996).

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

35. Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7,** 709–15 (2010).

36. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–40 (2010).

37. De Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20,** 1453–4 (2004).

38. Barish, G. D. *et al.* Bcl-6 and NF-κB cistromes mediate opposing regulation of the innate immune response. *Genes Dev.* **24,** 2760–2765 (2010).

39. Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* **1,** 729–748 (2006).

40. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–89 (2010).

41. Zambelli, F., Pesole, G. & Pavesi, G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.* **37,** W247–52 (2009).

CHAPTER 3

Global Analysis of LPS Tolerance

Abstract

Recognition of pathogen-associated molecular patterns such as lipopolysaccharide (LPS) through Toll-like receptor 4 (TLR4) results in an immense transcriptional response that includes the upregulation of genes encoding for proinflammatory cytokines. While necessary to counteract environmental dangers, excessive responses can be devastating to host tissues and organs. LPS tolerance is a well-studied phenomenon in which cells or organisms become refractory to repeated exposures of endotoxin to avoid hyperinflammation and host damage. Although long appreciated as a mechanism of immunosuppression, there is still no consensus as to the mechanisms underlying this phenomenon. In this study, we took a global approach to analyze the transcriptome of LPS tolerance. The use of transcripts isolated from chromatin represented a novel approach to study LPS tolerance. The analysis of gene induction in the tolerant state in the context of co-regulated gene classes from the TLR4 transcriptional activation network analysis provided us with insight into the mechanisms underlying LPS tolerance. Additionally, we have uncovered differences in the prolonged expression of subsets of genes, which complicated efforts to determine whether specific subsets of genes are susceptible or resistant to tolerance. Numerous Interferon-stimulated genes (ISGs) remained highly expressed after removal of the tolerizing LPS, which may be due to the persistence of IFN-β production. Furthermore, the overall hyporesponsiveness of the majority of inducible genes appeared to diminish over time. Together, the data suggest a broad downregulation in expression of genes upon repeated exposure to LPS, but that the immunosuppression may be transient. The findings should help to uncover the multiple mechanisms that contribute to establishment of the hyporesponsive state in cells after prolonged exposure to endotoxin.

Introduction

The recognition of microbial dangers by the cells of the innate immune system is critical to initiate responses that counteract infection. However, innate immune responses to pathogens

are a double-edged sword: Failure to respond to such dangers can be fatal to host organisms, yet prolonged or hyperactive responses can lead to chronic or oftentimes fatal inflammation. A tightly regulated response that allows for successful antimicrobial responses while simultaneously limiting excessive inflammation is therefore necessary for effective immunity. Because of these reasons, a better understanding of the underlying mechanisms that balance innate immune responses is required.

One well-studied mechanism of preventing excessive inflammation is endotoxin or LPS tolerance, defined as an overall hyporesponsiveness of organisms to repeated exposure of LPS. The initial observation took place in 1947 by Paul Beeson[1,2], where he observed that rabbits given repeated intravenous doses of pyrogenic substances derived from bacterium progressively diminished in their reaction over time. Since this initial observation, numerous studies to understand the mechanisms underlying LPS tolerance both *in vivo* and *in vitro* have been performed. One of the most well characterized effects of LPS tolerance is the decrease in cytokine production and release after stimulation. Upon repeated exposures to LPS, studies have found severe reduction in TNF-α production. At the transcriptional level, genes encoding proinflammatory cytokines such as *Il12b, Il6,* and *Il1b* are drastically reduced in expression after LPS tolerance has been established compared to naïve cells encountering LPS for the first time.

Despite the large amount of research to understand LPS tolerance, there is still no consensus with respect to the underlying mechanisms. The proposed mechanisms that establish LPS tolerance can be broadly classified into several groups based on the cellular compartment of hyporesponsiveness. First, it has been proposed that tolerance occurs at the level of the receptor. One study recently demonstrated that surface expression of the TLR4/MD2 complex begins diminishing shortly after the first exposure to LPS, causing the cells to become refractory to subsequent doses of LPS[3].

Other groups suggest that cells become tolerant to repeated LPS stimulations due to alterations in key signaling molecules that mediate the TLR4 response. In tolerized cells, IRAK1,

a key mediator of the MyD88 arm of the TLR4 signaling pathway, has been shown to be downregulated[3], and the association between IRAK1 and MyD88 is impaired[4]. Additionally, IRAK-M, a negative regulator of TLR4 proximal signaling, is upregulated after activation of TLR4[5–8]. Studies of tolerance further downstream from the proximal signaling machinery have demonstrated defects in both NF-κB and MAPK signaling. For example, IκB-α degradation does not occur when cells are exposed to repeated doses of LPS[9–11], while studies have demonstrated that activating NF-κB dimers, such as p50:RelA, are suppressed and inhibitory NF-κB dimers, such as p50 homodimers, are increased[12–14]. Further, activation of ERK1/2 and JNK members of the MAPK signaling pathways is impaired following the initial LPS exposure[5,11,15].

The third broad class of mechanisms of tolerance implicates suppressive autocrine mediators secreted out of cells after initial exposure to LPS. Studies have demonstrated a role for IL-10[16,17], a potent anti-inflammatory cytokine that suppresses activity of cytokines, such as IL-12 and IL-1β, as well as roles for other secreted factors, such as TGF-β[18], soluble TNF receptor[19] and prostaglandin E2 (PGE_2)[20] in establishing a tolerant state.

Most recently, chromatin has been implicated to play a role in LPS tolerance. In one particular study, the genes induced by LPS could be categorized into two broad classes: antimicrobial genes that remain inducible (not tolerizable), and inflammatory genes that are not inducible (tolerizable) when challenged with a second dose of LPS[21]. Importantly, both classes of genes acquire histone methylation marks after treatment with the first dose of LPS. Upon stimulation of cells with a second dose of LPS, however, only the antimicrobial genes acquire histone acetylation marks, RNA polymerase II recruitment, and transcription. Moreover, the inflammatory genes that require nucleosome remodeling for activation return to an inaccessible state after stimulation and nucleosome remodeling is not observed during the second stimulation; in contrast, antimicrobial genes return to a closed chromatin conformation, but

nucleosome remodeling occurs upon re-stimulation with LPS. The findings from this study imply that chromatin is an essential mechanism that establishes tolerance of selective subsets of genes: Antimicrobial genes retain their inducibility due to their critical roles in maintaining immunity against infection, while inflammatory genes become silent following the first exposure to LPS in order to prevent hyperinflammation and tissue damage.

Taken together, the great interest in understanding the etiology of endotoxin tolerance is an indication of its importance in human disease settings, such as sepsis and surgery. Patients who have previously encountered pathogen are at risk of developing secondary infections due to their refractory state towards pathogens[22]. Additionally, the lack of consensus of the mechanisms underlying LPS tolerance supports the complexity of the phenomenon, and demonstrates that unanswered questions remain. Importantly, a global analysis of gene expression changes during LPS tolerance is necessary to provide insight into the functional consequences of prolonged hyporesponsiveness. Previous studies have focused on readouts from model genes such as *Il6* and *Tnf*, but whether the findings apply to all LPS-inducible genes remain unclear.

RNA-sequencing technologies have emerged as a quantitative and highly accurate method to measure gene expression[23]. In this study, we took a global approach to identify genes that may or may not be inducible after the tolerization period. Initial studies revealed that global analysis of tolerance using mRNA did not provide an accurate measurement of tolerance at the transcriptional level, and that global analysis of primary transcripts was necessary to circumvent these challenges. We also examined the duration of the hyporesponsive state to determine whether tolerance diminishes over time. Using these findings in conjunction with our understanding of the TLR4 transcriptional activation network (Chapter 2) has revealed insight into the extent of LPS responsiveness after tolerance, as well as the activation dynamics of co-regulated genes. These observations should help to clarify the global extent of LPS tolerance,

understand what signaling pathways might be involved in the hyporesponsive state, and may provide useful tools in the future to help understand immunosuppressive pathologies.

Results

Primary and Mature Transcripts Exhibit Distinct Gene Expression Patterns

In initial studies to study the impact of LPS tolerance on gene expression, bone marrow-derived macrophages (BMDMs) were either not tolerized or tolerized with lipid A (the active component of LPS) for 24 hours (henceforth called "tolerizing LPS"; Figure 3-1A). After 24 hours, lipid A-containing media was replaced with lipid A-free media and rested for 0.25, 4, 8, or 24 hours to determine how long the effects of LPS tolerance last. After the rest period, cells were stimulated for 2 hours with a low (10 ng/ml) or high (100 ng/ml) dose of lipid A ("stimulating LPS"). Additionally, the cells that did not receive the tolerizing LPS dose were treated with the stimulating LPS dose to measure the transcriptional activation potential in non-tolerized macrophages. We assessed the effect of tolerance by monitoring gene expression for genes reported to be "not tolerizable" (inducible in tolerized cells given the stimulating LPS dose) such as *Ccl5* and *Saa3,* and "tolerizable" (not inducible in tolerized cells given the stimulating LPS dose) such as *Il12b* and *Il6*[21].

Consistent with previous findings, analysis of *Ccl5* and *Saa*3 mRNA revealed that these genes were expressed at a high level after the 24-hr initial lipid A treatment and rest period, with only a modest decline in mRNA levels as the rest period increased from 0.25 hr to 24 hrs. Moreover, the *Ccl5* and *Saa3* mRNA levels did not increase when the rested cells were re-stimulated with lipid A. The absence of induction during this second stimulation period at first glance suggests that the *Ccl5* and *Saa3* genes may be tolerized. However, the fact that the *Ccl5* and *Saa3* mRNA levels remain very high relative to the levels observed in unstimulated non-tolerized BMDMs could be interpreted as evidence that the cells are not subject to tolerance.

Thus, it is difficult to determine from these mRNA data whether the *Ccl5* and *Saa3* genes should be placed in the tolerized or not tolerized classes.

To evaluate more carefully the extent to which the *Ccl5* and *Saa3* genes are tolerized, it was necessary to analyze transcription independently of influences of mRNA stability. This was accomplished by examining nascent unspliced transcripts through the use of qRT-PCR primers that amplified fragments spanning exon-intron junctions. Strikingly, the nascent transcript level for the *Ccl5* gene was reduced nearly to background in the tolerized and rested cells (Fig. 3-1B, bottom), demonstrating that the high *Ccl5* mRNA level in these cells was not due to continued transcription, but rather to the existence of a stable mRNA pool. Importantly, when the tolerized and rested cells were re-stimulated with lipid A, induction of the *Ccl5* gene was observed with a high dose of lipid A, but not with a low lipid A dose. This result demonstrates that the *Ccl5* gene was partially tolerized, in that it was resistant to low-dose lipid A re-stimulation, but was not fully tolerized because it was potently induced by a high dose of lipid A.

We next analyzed the *Il12b* and *Il6* genes. An analysis of *Il12b* nascent transcripts (Fig. 3-1B, bottom) revealed that this gene was efficiently tolerized after the 24-hr initial stimulation period and short rest times of either 0.25 or 4 hrs, as nascent transcript induction was not observed after these short rest periods with either low or high doses of lipid A. However, potent *Il12b* induction was observed after rest periods of 8 or 24 hrs, but only by high-dose lipid A. An analysis of *Il6* nascent transcripts revealed strong tolerance to low-dose lipid A after a rest period of 0.25 hrs, with loss of tolerance to low-dose lipid A after longer rest periods of 4, 8, and 24 hrs. With high-dose lipid A, *Il6* tolerance was not observed, even with the 0.25-hr rest period, as potent induction was consistently observed. The results obtained with *Il12b* and *Il6* mRNAs were consistent with those obtained with their nascent transcripts, but these results were more challenging to interpret due to the apparent presence of stable pools of the *Il12b* and *Il6* mRNAs, which made it difficult to determine the extent to which gene transcription was tolerized.

To summarize, these results provide evidence that an examination of nascent transcripts is likely to be preferable for determining the extent to which genes are sensitive or resistant to endotoxin tolerance, due to confounding effects of mRNA stability. Furthermore, the results suggest that the extent of tolerance varies from gene to gene in a lipid A dose-dependent manner, and with the extent of tolerance diminishing to a variable extent with the length of the rest period following tolerance induction.

Global Analysis of Gene Expression Patterns in Tolerized Cells

To investigate endotoxin tolerance at a global level, we performed RNA-seq with non-tolerized and tolerized BMDMs. Based on the pilot studies in Figure 3-1, cells were either not tolerized (T0) or given a tolerizing lipid A dose (T24) and rested for either 0.25 or 8 hours (R0.25 or "short rest", and R8 or "long rest", respectively). After the rest period, cells were given the low dose of stimulating lipid A (10 ng/ml) for 0, 0.5, or 2 hours (S0, S0.5, or S2). Additionally, naïve (non-tolerized) cells were given the low dose of stimulating lipid A for 0, 0.5, or 2 hours (S0, S0.5, or S2). The experimental design with sample name designations are indicated in Table 3-1.

Because our initial studies indicated that primary transcripts more accurately measured the effect of endotoxin tolerance at the transcriptional level, RNA-seq was performed on chromatin-associated RNA isolated from biochemically fractionated cell lysates. Nascent transcript RNA-seq has previously been shown to provide accurate kinetics and quantitation of transcriptional activation and inactivation of genes[24].

Preliminary examination of well-studied genes affected in tolerized cells verified that this approach would be useful to better understand endotoxin tolerance at a global level. "Tolerizable" genes, such as *Il12b, Il6, Tnf,* and *Il1b,* were induced by lipid A in non-tolerized (T0) cells, but these genes were activated to less than half of that stimulated level in tolerized cells (T24) treated with lipid A (Figure 3-2A). Of these genes, *Il12b* was the most potently suppressed, as its nascent transcript level was less than 1% of the T0-R0-S+ level (S+ is

henceforth used to designate cells given a stimulating dose of LPS for either 0.5 or 2 hours) in the T24-R0.25-S+ samples, and less than 3% of the T0-R0-S+ level in the T24-R8-S+ samples. Notably, the *Tnf* nascent transcript level in the T24-R.5-S+ samples were 37.7% of that observed in the T0-R0-S+ sample, whereas, after the long rest period (T24-R8-S+), it was 59.4% of the level observed in the T0-R0-S+ sample. Thus, the strength of tolerance observed at *Tnf* appeared to diminish with the length of the rest period.

The *Il1b* gene provides an example of another type of profile (Fig. 3-2A). In tolerized cells rested for 0.25 hrs, the nascent transcript level for *Il1b* remained high prior to re-stimulation, presumably because active transcription was taking place at the end of the 24-hr initial stimulation period and had not yet subsided by the end of the 0.25 hr rest period. However, slightly lower transcript levels were observed 0.5 hrs after re-stimulation. After the 8-hr rest period, *Il1b* nascent transcript levels returned to background levels and the gene was susceptible to induction, but only to 39.2% of the induced level observed without tolerance induction.

We next assessed the expression patterns of the “not tolerizable” genes. The *Ccl5* and *Oasl1* nascent transcripts were induced to 62-63% of T0-R0-S+ samples after the long rest. Strikingly, we observed that some “not tolerizable” genes such as *Oasl1, Saa3,* and *Fpr1* were expressed higher in T24-R0.25-S0 and T24-R8-S0 compared to T0-R0-S+ samples. For example, *Oasl1* and *Saa3* were expressed between 30-50% of T0-R0-S+ samples after the long rest before given the stimulating LPS dose, and over 200% of T0-R0-S+ samples for *Fpr1.* Since the data represent actively synthesized transcripts, this suggests that in addition to mRNA stability precluding the analysis for genes such as *Ccl5* and *Saa3* (Figure 3-1B), LPS tolerance may be challenging to interpret for the genes that remain actively transcribed after removal of the tolerizing LPS. It is important to note that the initial examination of *Ccl5* primary transcript levels through qRT-PCR analysis suggested that T24-R0.25-S0 levels were close to background levels. In contrast, the nascent RNA-seq data revealed that *Ccl5* transcript levels

were expressed at 42% in T24-R0.25-S0 samples relative to T0-R0-S0 (Figure 3-2A). The inconsistency observed could be explained by experimental differences: In pilot studies using qRT-PCR to measure tolerance, T24-R0.25-S0 cells were rested for two hours in addition to the 0.25-hour rest while the T24-R0.25-S+ samples were given the stimulating dose of LPS. When preparing samples for RNA-seq, the T24-R0.25-S0 cells were rested for 0.25 hours only. The two additional hours of rest in the pilot studies therefore likely allowed *Ccl5* primary transcripts to diminish. Furthermore, RNA-sequencing is a much more quantitative and accurate method to measure gene expression, and therefore is likely to be capable of detecting low levels of gene expression more accurately than qRT-PCR methods. Together, this provides higher confidence in the information obtained from RNA-seq than those gathered through qRT-PCR.

To focus on the effects LPS tolerance has on LPS inducible genes, we limited the analysis to examine genes that were induced 5-fold or greater and expressed at least 1 RPKM in T0-R0-S+ relative to T0-R0-S0. The 554 5-fold induced genes were then hierarchically clustered to identify gene expression patterns in the T24 samples (Figure 3-2B). This approach revealed subsets of genes that remained inducible and other subsets of genes that were suppressed in both T24-R0.25-S+ and T24-R8-S+ samples. Interestingly, we observed a large proportion of genes whose expression remained high in T24-R0.25-S0 samples, and a smaller subset of genes with high expression in T24-R8-S0 samples (Figure 3-2B, right columns). Specifically, 26.7% and 41.7% of the 5-fold induced genes were expressed greater than 50% and between 10 and 50% in T24-R0.25-S0 samples relative to the maximum expression in T0-R0-S+ samples, respectively (Figure 3-2C, blue). Furthermore, 6.5% and 33.2% of the 5-fold induced genes were expressed greater than 50% and between 10 and 50% in T24-R8-S0 samples relative to the maximum expression in T0-R0-S+ samples, respectively (Figure 3-2C, red). Similar results were observed when we examined more potently induced genes (10-fold induction, 3 RPKM threshold; Figure 3-2D), indicating that this was not an artifact due to the induction levels of genes. Since the global analysis focused on newly synthesized rather than

mature transcripts, the data suggest that a large subset of LPS-inducible genes remain highly transcribed after the short rest, with a subset of these genes maintaining active transcription even after the long rest period prior to the stimulating LPS dose.

Inducible Genes With Sustained Transcription Tend to be Secondary Response Inferferon-Stimulated Genes

To better understand the properties of genes that sustained high expression levels after removal of the tolerizing LPS and before the stimulating LPS dose (henceforth called “high basal” genes), we focused on the 226 genes induced 10-fold or greater and expressed at least 3 RPKM in response to lipid A that were characterized in Chapter 2. Importantly, the analysis in Chapter 2 classified the 226 genes into those that do not (primary response) or do (secondary response) require new protein synthesis for their activation. We therefore examined the expression levels of T24-R0.25-S0 and T24-R8-S0 samples relative to naïve T0-R0-S0 samples for the 226 genes classified as either primary or secondary response to determine if the “high basal” genes tended to be in one class or the other. We found that 8.6% and 34.1% of primary response genes had expression levels greater than 50% and 10-50% of T0-R0-S0 after the short rest, respectively (Figure 3-3A, left, blue). After the long rest, the distribution decreased to 1.5% and 17.4% of primary response genes with expression greater than 50% and 10-50%, respectively (Figure 3-3A, left, red). In contrast, 14.9% and 61.7% of the secondary response genes exhibited expression levels greater than 50% and 10-50% of T0-R0-S0 after the short rest (Figure 3-3A, right, blue), and 10.6% and 52.2% of the secondary response genes remained expressed greater than 50% and 10-50% after the long rest, respectively (Figure 3-3A, right, red).

To gain further insight as to what signaling pathways or transcription factors may be contributing to the high basal expression of the secondary response genes, we examined the expression of the 226 genes in T24-R0.25-S0 and T24-R8-S0 samples in the context of the

classification described in Chapter 2. In general, we observed broad and robust levels of transcription after the short rest, including moderately high expression levels of IRF3 and TRIF gene classes from the primary response and high expression of the IFN-dependent secondary response gene class (Figure 3-3B). After the long rest, expression of most genes across all of the gene classes diminished to levels near 10% of T0-R0-S0 samples with the exception of the IFN-dependent secondary response genes, whose median expression was maintained above 10% of T0-R0-S0 samples. Together, these results suggest that there is broad sustained expression of many genes after the short rest, but the IFN-dependent secondary response genes remain highly expressed even after the long rest.

Due to the finding that genes with “high basal” expression after the long rest tended to be IFN-dependent secondary response genes, we further extended the analysis to explore the relationship between IFN-dependence and “high basal” expression of the 554 5-fold induced genes. The 554 genes were first classified into “high basal” or “low basal” based on the lowest expression they reached in either T24-R0.25-S0 or T24-R8-S0 samples. If a gene reached less than 10% of T0-R0-S0 in either short or long rested samples that were not given the stimulating LPS dose, it was classified as a “low basal” gene (further discussed in Figure 3-5 and 3-6). On the other hand, if a gene was expressed greater than 10% of T0-R0-S0 in both short and long rested samples, it was classified as a “high basal” gene. The “low basal” and “high basal” genes were then examined for their sensitivity to cycloheximide (CHX) as well as their dependence on IFN-signaling.

We observed that 63.4% of the “high basal” genes had low expression in CHX-treated cells (< 33% expression in CHX relative to wildtype), compared to 33.8% of the “low basal” genes that exhibited CHX sensitivity (Figure 3-3C, left). Similarly, 48.1% of the “high basal” genes were IFN-dependent (< 33% expression in *Ifnar-/-* relative to wildtype), compared to 16.3% of “low basal” genes (Figure 3-3C, right). Furthermore, an analysis of all cycloheximide-sensitive genes in the group of 554 5-fold induced genes revealed a relationship between IFN-

dependence and T24-R0.25-S0 or T24-R8-S0 expression levels: genes that were more strongly dependent on IFN signaling exhibiting higher expression in T24-R0.25-S0 or T24-R8-S0 samples (Figure 3-3D). Taken together, the findings from the analysis of T24-R0.25-S0 and T24-R8-S0 samples in the context of the 226 well-characterized genes were consistent with the observations made when analyzing in the context of the 554 5-fold induced genes, and that the "high basal" genes tended to be secondary response and IFN-dependent.

Persistent IFN-β Contributes to Prolonged Transcription of ISGs

To investigate why many interferon stimulated genes (ISGs) maintained high transcription levels after the cells were removed of the tolerizing LPS and rested for both short and long periods, we assessed the expression of the *Ifnb1* gene, which encodes for the type I interferon IFN-β. Since many ISGs remained highly transcribed, we hypothesized that this could be due to high expression of *Ifnb1* transcripts. qRT-PCR analysis revealed that *Ifnb1* transcripts were induced in T0-R0-S+ samples, with maximal induction in T0-R0-S2 (Figure 3-4A). However, after the 24-hour tolerizing LPS, *Ifnb1* expression returned to background levels and was not induced when given the stimulating LPS dose in both short and long rested cells. ELISA analysis confirmed that IFN-β secretion was suppressed in both T24-R0.25-S2 and T24-R8-S2 samples relative to T0-R0-S2 (Figure 3-4B). We next measured IFN-β levels after the 24-hour tolerizing LPS but just prior to stimulus removal, as well as after the short and long rest period but prior to receiving the stimulating LPS dose to assess the presence of the cytokine in media that may be contributing to IFN-dependent gene expression. Interestingly, we observed low levels of IFN-β present in the media after the 24-hour tolerizing LPS treatment (Figure 3-4C, blue columns). Although these levels were low (~4 pg/ml) relative to the amount of IFN-β secreted after 2 hours of lipid A stimulation in non-tolerized macrophages (~100 pg/ml, Figure 3-4B), it is well documented that low levels of type I interferons is sufficient to stimulate expression of ISGs[25].

Furthermore, we observed trace levels of IFN-β present in the media after removal of the tolerizing LPS and the long rest (T24-R8-S0), indicating that low amounts of IFN-β continued to be secreted out of the cell even after removal of the tolerizing LPS. Together, these results suggest that the "high basal" expression of IFN-dependent genes in T24-R0.25-S0 and T24-R8-S0 samples could be due to the presence of IFN-β in the media 24 hours after the tolerizing LPS was given, even after transcription of the *Ifnb1* gene has terminated.

LPS Tolerance Broadly Impacts Inflammatory Gene Induction

We next sought to better understand the "low basal" genes, which we defined as those that returned to within 10% expression in T24-R0.25-S0 or T24-R8-S0 samples relative to T0-R0-S0. The 337 "low basal" genes within the 554 5-fold induced genes from Figure 3-2A were first separated into 222 primary and 115 secondary response genes based on their sensitivity to cycloheximide (33% threshold). The secondary response "low basal" genes were further classified as IFN-independent (n=70) and IFN-dependent (n=45) based on their expression in *Ifnar*$^{-/-}$ macrophages stimulated with lipid A (33% threshold). Each of these classes of genes was then hierarchically clustered, taking into consideration all stimulation conditions (Figure 3-5A). This revealed that a large proportion of genes in the three classes exhibited tolerance in T24-R0.25-S+ and T24-R8-S+ relative to T0-R0-S0.

Importantly, we also observed that nearly half of the 337 "low basal" genes were expressed greater than 10% in T24-R0.25-S0 samples relative to T0-R0-S0 (Figure 3-5A, fourth column; Figure 3-5B, top panel). Specifically, 103 of 222 primary response, 31 of 70 IFN-independent secondary response, and 28 of 45 IFN-dependent secondary response genes exhibited expression levels greater than 10% in T24-R0.25-S0 samples relative to T0-R0-S0. This indicated that the "low basal" expression and inclusion in the analysis of these genes was due to T24-R8-S0 samples reaching within 10% of the T0-R0-S0 samples. Examination of the

maximum expression in T24-R0.25-S+ samples (Figure 3-5B, top panel) revealed that 80%, 87%, and 82% of the primary response, IFN-independent, and IFN-dependent secondary responses (excluding those indicated as “high basal”) were expressed at less than 25% relative to T0-R0-S+, respectively. Furthermore, the T24-R08-S+ samples (Figure 3-5B, bottom panel) exhibited a similar broad level of tolerance, with just 10%, 17%, and 2% of the primary response, IFN-independent, and IFN-dependent secondary responses reaching at least 75% of the T0-R0-S+ stimulation level, respectively. Thus, the data indicate that after the short rest, the majority of genes exhibited “high basal” expression levels prior to treatment with the stimulating dose of LPS. Furthermore, nearly all of the genes that exhibited “low basal” expression were robustly tolerized. After the long rest however, many of these genes become inducible to 50% of the initial induction magnitude. These observations are further addressed below in Figure 3-6.

To more concisely measure the level of tolerance, we used the maximum expression level of each “low basal” gene from either T24-R0.25-S+ or T24-R8-S+ samples to represent the overall degree of tolerance. This revealed that both primary response and IFN-independent secondary responses contained large subsets of genes that were “tolerized” and fewer genes that were “not tolerized”. For example, 80 of 222 (36%) and 35 of 70 (50%) of the genes were induced to less than 25% in the primary response and IFN-independent secondary response, respectively (Figure 3-5C). In contrast, just 24 of 222 (11%) and 12 of 70 (17%) of the genes were inducible to at least 75% of T0-R0-S+ in the primary response and IFN-independent secondary response, respectively. Furthermore, of the 45 IFN-dependent secondary response genes, 33 were expressed at less than 25% relative to T0-R0-S+ samples, while just one was expressed at greater than 75% in either T24-R0.25-S+ or T24-R8-S+ samples relative to T0-R0-S+. Among the “not tolerized” subset were the primary response genes *Ptgs2* and *Tlr2,* two genes that are involved in mediating inflammatory responses. Interestingly, both of these genes were previously classified as NF-κB target genes (Chapter 2). Further examination of the status of NF-κB signaling components in cells given the tolerizing dose of LPS and either long or short

rested may provide insight into the mechanism of tolerance for this subset of genes. Additionally, the IFN-independent secondary response genes *Itgb8* and *Flrt3* were inducible when given the stimulating dose of LPS after treated with the tolerizing LPS and rested. While these genes play roles in mediating cell migration and adhesion, how their expression is regulated is not known. It is important to note that these "not tolerized" genes exhibited high expression in T24-R0.25-S0 samples, and therefore their inducibility in T24-R0.25-S+ samples could not be examined.

Because of the possibility that the classification of "tolerized" and "not tolerized" genes could be an artifact of the initial T0-R0-S+ induction levels, we next examined the fold inductions of genes in T0-R0-S+ samples in various degrees of tolerance bins. This revealed a broad fold induction distribution for genes in T0-R0-S+ samples, and across all of the defined tolerance bins for both the primary and secondary response classes (Figure 3-5D), suggesting that the broad downregulation of gene induction in tolerized cells was not due to an artifact of the dynamic range of the T0-R0-S+ fold induction levels.

To gain insight as to what regulatory pathway the "tolerized" and "not tolerized" genes belonged to, we next assessed the maximum expression levels in either the T24-R0.25-S+ or T24-R8-S+ samples for the 149 of 226 "low basal" genes from Chapter 2 (Figure 3-4E). Interestingly, each co-regulated gene class demonstrated a broad range of expression levels relative to T0-R0-S+. For example, some genes were expressed at less than 10% in T24-R0.25-S+ and T24-R8-S+ samples relative to T0-R0-S+ while others were expressed at greater than 90% relative to T0-R0-S+ samples. Examination of median expression values within each co-regulated gene cluster revealed a similar downregulation of expression, with median values ranging between 30-45% relative to T0-R0-S+ samples across all of the co-regulated gene classes. Notably, the three gene classes with the lowest median expression levels were the NF-κB/IRF3 primary response, the IFN-dependent secondary response, and the IFN-independent secondary response. However, because the NF-κB/IRF3 cluster contained just three genes (the

other two NF-κB/IRF3 genes were “high basal” and not included in this analysis), it is unclear how significant this observation is. Nevertheless, both the IFN-dependent and IFN-independent secondary response genes exhibited lower expression levels in T24-R0.25-S+ and T24-R8-S+ compared to expression levels in the other co-regulated gene classes. This suggests a possibility that the low T24-R0.25-S+ and T24-R8-S+ expression levels observed for the secondary response genes could be due to the overall dampened response of the primary response in T24-R0.25-S+ and T24-R8-S+ samples.

The Effect of LPS Tolerance Diminishes Over Time

One observation made of the “low basal” genes was the moderately transient nature of gene suppression in tolerized cells, as discussed above (Figure 3-5B). Although we noted a broad downregulation of gene activation in T24-R0.25-S+ and T24-R8-S+ samples (Figure 3-5C,E), many genes appeared to be induced more highly after the long rest and treatment with the stimulating LPS compared to the short rest (Figure 3-5B). To more quantitatively investigate the differences, we subcategorized the “low basal” genes into 4 primary response clusters, 3 IFN-independent secondary response clusters, and 1 IFN-dependent secondary response cluster based on the hierarchical clustering of the “low basal” genes (Figure 3-5A). We then performed box plot analysis of the nine stimulation conditions for each cluster of genes (Figure 3-6A). This revealed that some clusters such as Primary 3, Primary 4, IFN-i 1, and IFN-i 3 exhibited a higher dynamic range of induction in T24-R8-S+ samples compared to T24-R0.25-S+ samples. Furthermore, correlation dendrograms of the primary response and IFN-independent secondary response genes revealed that the long rest stimulation conditions (T24-R8) grouped with the non-tolerized samples (T0-R0), while short rested stimulation conditions (T24-R0.25) grouped with each other (Figure 3-6B). In contrast, the IFN-dependent secondary response genes generally clustered into the rest periods: the naïve non-tolerized samples grouped with each other while the T24-R0.25 samples clustered together. The T24-R8 samples were scattered

throughout the dendrogram, with the T24-R8-S+ samples more correlated with the T24-R0.25 samples and the T24-R8-S0 sample grouped with the T0-R0 samples. This suggests that the tolerance of the long rested samples stimulated with LPS may be wearing off, and therefore exhibit similarities to T24-R0.25 samples that exhibit high transcription levels of the majority of IFN-dependent genes. Furthermore, the finding that the T24-R0.25 samples clustered with each other supports the hypothesis that many of the "high basal" IFN-dependent genes continued to be highly expressed in all T24-R0.25 samples due to the presence of IFN-β in the media (Figure 3-4C). Taken together, the data suggests that although there is a broad downregulation of inflammatory gene expression during LPS tolerance, the effect of tolerance may be transient.

Expression of Regulators Downstream of TLR4 Activation

Numerous regulators of the TLR4 signaling pathway and response have been implicated in establishing the LPS tolerant state. We therefore examined expression patterns of the genes encoding these regulators using the RNA-seq datasets to gain insight as to what role they might play in LPS tolerance. Specifically, we compared gene expression in non-tolerized resting cells (T0-R0-S0) to cells given the tolerizing LPS but not the stimulating LPS (T24-R0.25-S0 and T24-R8-S0). This allows us to examine any changes in expression before and after establishment of tolerance. Examination of TLR4 signaling molecules and downstream transcription factors did not reveal substantial downregulation of expression for these genes in the tolerized state (T24, both short and long resting periods; Table 3-2, column 12 and 13). Nearly all genes examined after removal of the tolerizing LPS (T24-R0.25-S0 and T24-R8-S0) were expressed within 2-fold relative to T0-R0-S0. Furthermore, we also did not observe a substantial upregulation of anti-inflammatory mediators of the TLR4 response such as *Il10* (Table 3-3). However, it is important to note that unaltered gene expression does not necessarily translate to unaltered function of the protein encoded by the gene. In contrast, we observed upregulated expression of genes encoding for negative regulators of TLR4 signaling

in T24-R0.25-S0 and T24-R8-S0 samples relative to T0-R0-S0. For example, expression of the *Irak3* gene encoding for IRAK-M was nearly 4-fold and 2-fold increased in T24-R0.25-S0 and T24-R8-S0 samples, respectively (Table 3-4, column 12 and 13). This agrees with previous studies demonstrating that IRAK-M inhibits TLR4 signaling, and that IRAK-M deficient mice fail to develop LPS tolerance[7]. We also observed a 8-fold and 4-fold higher induction of *Socs3* in tolerized T24-R0.25-S0 and T24-R8-S0 samples relative to T0-R0-S0, respectively. Although SOCS3 has not been demonstrated to have a role in mediating LPS tolerance, it is a well-studied molecule that negatively regulates of cytokine signaling through inhibition of STAT proteins[26,27]. Other genes encoding for negative regulators of the TLR4 response such as *Dusp1* and *Trafd1* also exhibited elevated levels in cells given the 24-hour tolerizing LPS dose compared to naïve, unstimulated cells. Taken together, the findings suggest that the global downregulation of gene expression observed in tolerized cells may be due to upregulation of various known negative regulators of TLR4 proximal signal transduction.

Discussion

The results of this study provide evidence that LPS tolerance broadly impacts inflammatory gene induction. Performing RNA-seq on chromatin-associated transcripts enabled us to examine LPS tolerance at the transcriptional activation level. By examining gene expression in tolerized cells compared to non-tolerized cells given a stimulating LPS dose in the context of the TLR4 network analysis (Chapter 2), we found only a small subset of not tolerized genes (inducible in T24-R0.25-S+ and T24-R8-S+ samples), but a much larger subset of genes that were tolerized in cells given the tolerizing LPS dose (not inducible in T24-R0.25-S+ and T24-R8-S+ samples). Furthermore, the tolerized genes were broadly distributed across all of the co-regulated gene classes. The finding that many well-known negative regulators of TLR4 signal transduction such as IRAK-M were upregulated after the 24-hr tolerizing LPS treatment suggest

that the broad inhibition of gene activation may be due to a proximal block in TLR4 signal transduction events.

The broad suppression of gene induction also suggests that multiple mechanisms likely contribute to LPS tolerance. In contrast to a previous study[17], we did not find substantial upregulation of *Il10*. However, this may be due to differences in cell type or tolerance protocol, or due to regulation of Il-10 through post-translational mechanisms. Further investigation of Il-10 levels in tolerized and non-tolerized cells should clarify this. Furthermore, although we observed broad gene suppression in tolerized cells across nearly all of the co-regulated gene clusters, a closer examination of how components of signaling pathways such as NF-κB and MAP kinase are affected will help clarify both the degree and duration of LPS tolerance for each pathway.

It is worthy to note that although we observed a broad suppression of TLR4-induced genes in tolerized macrophages, the suppression of genes was less substantial in the long rest compared to short rest cells given the stimulating LPS. This suggests that the hyporesponsive state of tolerized cells could be transient, and the resting time between tolerizing and stimulating LPS treatments allows cells to return to baseline. Although our analysis in context of the TLR4 transcriptional network study did not reveal co-regulated gene clusters that were “not tolerized”, it will be interesting to further investigate what signaling molecules have returned to basal levels after the long rest period but not the short rest period that may account for the moderate de-suppression in gene expression observed in T24-R8-S+ samples compared to T24-R0.25-S+. It will also be informative to better understand the regulatory mechanisms underlying activation of the genes that are “not tolerized” such as *Tlr2 and Ptgs2*, as it may reveal specialized mechanisms that are unaltered in tolerized cells.

The global analysis of LPS tolerance uncovered a large subset of genes whose expression remained highly transcribed after stimulus removal (“high basal”). The high levels of expression during the rest period for these genes (After the tolerizing LPS was removed but before the stimulating LPS was given) proved to be difficult to classify as “tolerized” or “not

tolerized" because it appeared that the genes were not shut off after removal of the tolerizing LPS. Difficulty arose in determining what sample condition to use as a baseline for determine the level of tolerance, and it therefore seemed reasonable to perform a separate assessment of the "high basal" genes. Analysis in the context of the TLR4 transcriptional network revealed that nearly two-thirds of the "high basal" genes were secondary response genes. The secondary response genes require new protein synthesis, and their activation kinetics is delayed relative to the primary response, indicating that the secondary response genes likely have complex regulatory mechanisms responsible for their activation by the 24-hour tolerizing LPS period due to multiple layers of feedback. For example, transcription of these genes may be prolonged and can result in "high basal" expression levels if cytokines synthesized during the primary response have the capability to activate them and are synthesized at different times during the tolerizing LPS dose. Additionally, expression of these genes may be prolonged if the cytokine involved in its activation is synthesized during the primary response and has a long half-life. The finding that IFN-β was present in culture media after the 24-hour tolerizing LPS period, and the observation that nearly half of the "high basal" genes were IFN-dependent, supports this idea.

For this study, we found it necessary to examine chromatin-associated transcripts due to differences in mRNA stability that complicated the analysis. It will be interesting to compare these datasets to a global analysis of LPS tolerance using mRNA to highlight any differences in LPS tolerance for different subsets of genes, which may indicate mechanisms of tolerance at post-transcriptional levels such as mRNA processing, export, and stability.

Methods and Materials

Cell Culture and Activation

Bone marrow-derived macrophages were prepared from C57Bl/6 mice. Bone marrow cells were harvested from the femurs and tibiae and cultured for 6 days in M-CSF conditioned media containing 20% serum. Macrophages were activated on day 6 with a tolerizing dose of lipid A (Sigma, 100 ng/ml). After 24 hours, the lipid A-containing media was removed and macrophages were washed with warm PBS before replacing with fresh lipid A-free media. Cells were rested for varying amounts of time before challenging with a second dose of lipid A (10 ng/ml or 100 ng/ml).

qRT-PCR

RNA was extracted using TRI-Reagent (Molecular Research Center). RNA was DNase I treated (Qiagen) and purified using an RNeasy kit (Qiagen). RNA was reverse transcribed into cDNA using the Omniscript RT kit (Qiagen) and primed with random hexamers. cDNA fragments were analyzed by qRT-PCR using SensiMix *Plus* (Quantace) using primer pairs to amplify both mRNA and primary transcript products.

RNA-seq Library Preparation

Chromatin-associated RNA was purified as described. Chromatin purity was confirmed by immunoblot analysis of SNRP70, β-Tubulin (Sigma), and Histone H3 (Abcam). Chromatin RNA was depleted of ribosomal RNA using the Ribominus Eukaryote kit (Life Technologies). Strand-specific libraries were generated by using 60 ng of chromatin RNA according to manufacturer's instruction from the Truseq RNA Sample Preparation Kit v2 (Illumina), with the following modifications: second strand cDNA was synthesized in the presence of deoxyuridine triphosphate (dUTP) according to the dUTP method. cDNA libraries were single-end sequenced (50 bp) on an Illumina HiSeq 2000.

RNA-seq Read Mapping and Processing

Reads were aligned to the mouse genome (NCBI37/mm9 build) with TopHat v1.3.3 and allowing reads to be aligned once with up to two mismatches per read permitted. RPKM values were calculated as described. Due to the possibility that chromatin transcripts may be spliced but remain associated with the chromatin/RNA polymerase complex, RPKM values were calculated by counting reads mapped to introns and divided by the sum of the length of all introns within the transcription unit. All RPKMs represent an average from two biological replicates.

To determine the expression relative to non-tolerized samples, the RPKM value in basal, unstimulated samples was set to 0% and the maximum RPKM value from the non-tolerized samples given a stimulating LPS dose was scaled to 100% for each gene. The RPKM expression values for each sample were converted to percent expression using this scale.

Figure Legends

Figure 3-1: Primary and Mature Transcripts Exhibit Distinct Gene Expression Patterns

(A) Basic experimental design to investigate LPS tolerance. Cells were first given a tolerizing (T) LPS dose. After 24 hours, the stimulus was removed and rested (R) for 0.25, 4, 8, or 24 hours prior receiving a stimulating LPS (S) dose. (B) The fold induction of mRNA (top panel) and primary transcripts (bottom panel) for representative genes in tolerized cells stimulated with a low (10 ng/ml) or high (100 ng/ml) dose of LPS relative to naïve, unstimulated samples, quantified by qRT-PCR is shown. Cells that did not receive the second, stimulating LPS dose are indicated by the white bars. The rest periods between the tolerizing LPS and stimulating LPS are indicated in hours.

Figure 3-2: Global Patterns of Gene Expression and Suppression Analyzed by RNA-seq of Chromatin-Associated RNA

(A) Expression of genes previously described as either “tolerized” or “not tolerized” are shown and represented as a percent expression relative to the maximum expression in T0-R0-S+ (naïve cells given a stimulating LPS dose). (B) The 554 5-fold, 1 RPKM LPS-induced genes were hierarchically clustered taking all stimulation conditions (described in Table 3-1) into consideration. The heatmap is colored based on percentile of RPKM values. The expression levels after removal of the tolerizing LPS and before given the stimulating LPS are indicated to the right of the heatmap after the short rest (R0.25, left column) or long rest (R8, right column) but prior to the stimulating LPS dose. The distribution of T24-R0.25-S0 (blue) and T24-R8-S0 (red) percent expression relative to T0-R0-S0 samples for the (C) 5-fold and (D) 10-fold induced genes is shown. The horizontal dashed grey lines indicate the 10% expression threshold. The

tables below each graph indicate the number of genes in T24-R0.25-S0 and T24-R8-S0 samples with expression levels of varying ranges.

Figure 3-3: LPS-Inducible Genes with Sustained Expression Tend to be Interferon-Dependent Secondary Response Genes

(A) The 226 inducible genes from TLR4 network studies (Chapter 2) were subdivided into primary response (left) and secondary response (right). The distribution of the percent expressions in T24-R0.25-S0 (blue) and T24-R8-S0 (red) samples relative to T0-R0-S0 for the primary and secondary responses is shown. (B) The RPKM values in T24-R0.25-S0 (blue) and T24-R8-S0 (red) samples are plotted for each gene in each co-regulated gene class. The median RPKM values within each class for each resting period are indicated as horizontal black bars. The dashed grey line indicates the 10% expression threshold. (C) The 554 5-fold induced genes were separated based on the minimum percent expression reach in either T24-R0.25-S0 or T24-R8-S0 samples. If a gene expressed within 10% of T0-R0-S0 samples in either T24-R0.25-S0 or T24-R8-S0 samples, it was classified as “low basal”. If a gene was expressed greater than 10% in both T24-R0.25-S0 and T24-R8-S0 relative to T0-R0-S0, it was classified as “high basal”. The distribution of expression in cycloheximide-treated samples (CHX; left) and *Ifnar-/-* cells (right) are shown for the “high basal” (blue) and “low basal” (grey) genes. (D) The 554 5-fold induced genes were first separated into those that were or were not sensitive to CHX treatment (33% expression threshold). The expression in *Ifnar*$^{-/-}$ macrophages (x-axis) and the minimum percent expression from either T24-R0.25-S0 or T24-R8-S0 (y-axis) is shown for the 253 CHX-sensitive genes. The horizontal dashed grey line represents the 10% expression threshold. The table below the scatterplot indicates the distribution of the 253 CHX-sensitive genes based on their IFN-dependence and T24-R0.25-S0 or T24-R8-S0 expression.

Figure 3-4: Persistent IFN-β Contributes to Prolonged Transcription of ISGs

(A) The expression of *Ifnb1* primary transcript in non-tolerized (T0) or tolerized (T24) macrophages given a stimulating dose of LPS for 0, 0.5, 1, 2, or 3 hours was quantified by qRT-PCR. Tolerized macrophages were rested for 0.25 or 8 hours prior to treatment with the stimulating LPS. The values represent the relative amounts of transcript (ng) relative to genomic DNA standard curves. (B) IFN-β secretion was measured in non-tolerized (T0) and tolerized (T24) macrophages that were activated with LPS for 2 hours by ELISA. The tolerized macrophages were rested for either 0.25 (R0.25) or 8 (R8) hours after removal of the tolerizing LPS before given the stimulating dose of LPS. (C) IFN-β secretion was measured in macrophages treated with the 24-hour tolerizing LPS, both before the tolerizing LPS was removed, and after the indicated rest period (R0.25 or R8) by ELISA.

Figure 3-5: LPS Tolerance Broadly Impacts Inflammatory Gene Induction

(A) The 337 of 554 5-fold induced genes that had expression levels within 10% of T0-R0-S0 in either T24-R0.25-S0 or T24-R8-S0 samples ("low basal") were separated into primary and secondary response based on their sensitivity to cycloheximide (33% expression threshold). The secondary response genes were further separated into IFN-dependent and IFN-independent based on their expression in *Ifnar*$^{-/-}$ macrophages (33% expression threshold). The primary response and IFN-independent secondary response classes were then each hierarchically clustered, taking all stimulation conditions into account. The IFN-dependent genes were not clustered any further. Heatmap colors are based on percentile of expression. (B) The 337 "low basal" genes were classified into primary response, IFN-independent secondary response, and IFN-dependent secondary response as described in (A). The maximum expression in T24-R0.25-S+ (top panel) and T24-R8-S+ (bottom panel) samples were binned

and is shown as a percent of total within each category. If expression in either T24-R0.25-S0 or T24-R8-S0 did not reach within 10% of T0-R0-S0 but did reach that threshold in the other, it was categorized as “high basal” for the sample that did not fall within the 10% threshold. (C) The distribution of maximum percent expression values in either T24-R0.25-S+ or T24-R8-S+ samples is shown for the “low basal” primary response, IFN-independent secondary response, and IFN-dependent secondary response classes. Values represent the maximum expression relative to T0-R0-S0 samples. (D) The primary response (top) and secondary response (bottom) genes were binned based on their maximum expression in tolerized macrophages treated with a stimulating LPS dose (T24-R0.25-S+ or T24-R8-S+). Each of these bins was further grouped based on the maximum fold induction in naïve, non-tolerized macrophages given the stimulating LPS dose (T0-R0-S+). The values indicate the relative distribution of these fold induction values within each expression bin. (E) The 226 inducible genes from Chapter 2 were analyzed for their expression in T24-R0.25-S0 and T24-R8-S0 samples. The 147 of 226 genes with “low basal” expression were grouped into co-regulated gene classes. The values represent the maximum expression level in either T24-R0.25-S+ or T24-R8-S+ samples. The median expression for each co-regulated group is indicated as horizontal red dashes.

Figure 3-6: The Effect of LPS Tolerance Diminishes Over Time

(A) The 337 5-fold induced genes with “low basal” expression were classified into primary response, IFN-independent secondary response, and IFN-dependent secondary response genes as described in Figure 3-5A. The primary response was separated into 4 sub-classes and the IFN-independent secondary response was separated into 3 sub-classes based on the hierarchical clustering in Figure 3-5A. We then performed box plot analysis on each of the 8 groups of genes to identify the distribution of RPKM values within each group. (D) The 337 5-

fold induced, “low basal” genes were separated into primary response, IFN-independent secondary response, and IFN-dependent secondary response. The experimental datasets (described in Table 3-1) were hierarchically clustered to identify experimental conditions that exhibited the highest degree of similarity to each other.

Table 3-1: Experimental Design and Sample Name Designations of Global Analysis of LPS Tolerance

The experimental design and sample naming system is shown in the table. The tolerizing LPS dose is designated as T, followed by either 0 or 24 to denote if the cells were non-tolerized (T0) or tolerized (T24) for 24 hours. The resting period is denoted by R, followed by the number of hours the cells were rested (0, 0.25, or 8). The stimulating LPS dose is designated as S, followed by a number to indicate the number of hours the cells received the stimulating dose of LPS (0, 0.5, or 2). In the text, “S+” is used to indicate any stimulated timepoint (0.5 or 2 hours) for simplicity.

Table 3-2: Expression of TLR4 Signaling Molecules and Related Factors During LPS Tolerance

The RPKM values of genes encoding for components of the TLR4 signal transduction pathway and related factors that promote inflammatory gene activation are shown. RPKM values are given for both tolerized (T24) and non-tolerized (T0) cells. The last two columns indicate the fold induction of each gene in T24-R0.25-S0 and T24-R8-S0 relative to expression in T0-R0-S0 cells.

Table 3-3: Expression of Anti-Inflammatory Mediators During LPS Tolerance

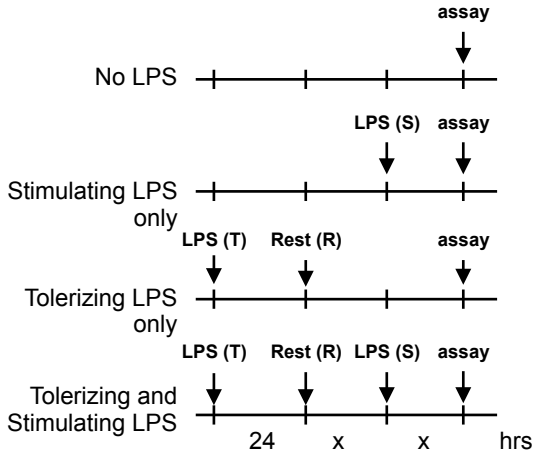
The RPKM values of genes encoding for anti-inflammatory mediators previously demonstrated to be critical to establish LPS tolerance are shown. RPKM values are indicated for both tolerized (T24) and non-tolerized (T0) cells. The last two columns indicate the fold induction of each gene in T24-R0.25-S0 and T24-R8-S0 relative to expression in T0-R0-S0 cells.

Table 3-4: Expression of Negative Regulators of TLR4 Signal Transduction During LPS Tolerance

The RPKM values of genes encoding for negative regulators of TLR4 signal transduction are shown. RPKM values are indicated for both tolerized (T24) and non-tolerized (T0) cells. The last two columns indicate the fold induction of each gene in T24-R0.25-S0 and T24-R8-S0 relative to expression in T0-R0-S0 cells.

Figure 3-1: Primary and Mature Transcripts Exhibit Distinct Gene Expression Patterns

A



B

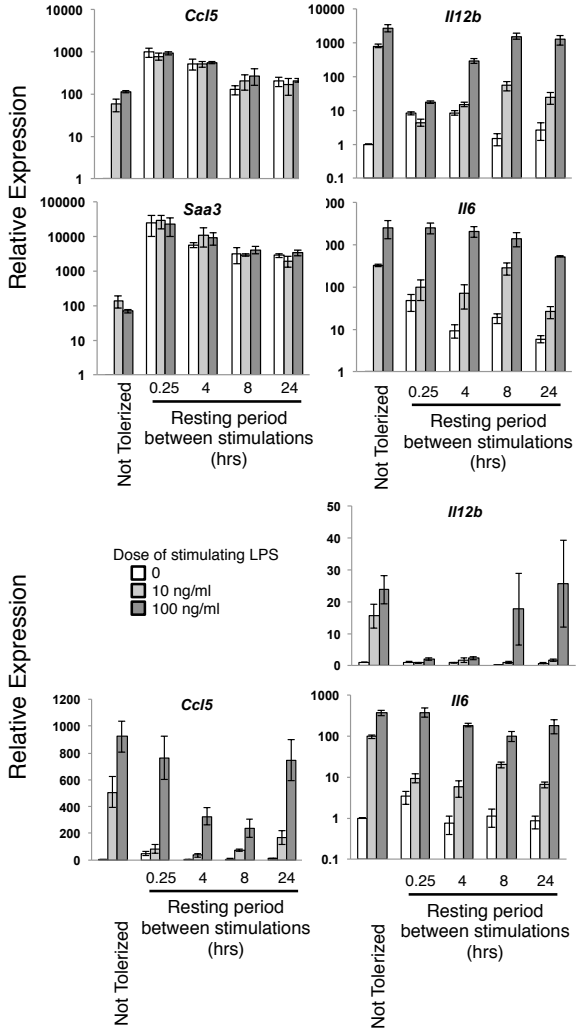


Figure 3-2: Global Patterns of Gene Expression and Suppression Analyzed by RNA-seq of Chromatin-Associated RNA

A

	Percent Expression Relative to Not Tolerized Resting Cells (T0-R0-S0)								
	T0-R0			T24-R0.25			T24-R8		
Gene	0	0.5	2	0	0.5	2	0	0.5	2
Il12b	0.0	16.1	100.0	0.4	0.7	0.8	0.2	1.3	3.0
Il6	0.0	0.7	100.0	20.1	23.6	26.3	0.8	7.3	29.2
Tnf	0.0	100.0	51.8	9.6	37.7	21.3	7.8	59.4	22.3
Il1b	0.0	100.0	29.7	62.6	38.9	28.5	1.4	39.2	15.3
Serpine1	0.0	95.0	100.0	0.0	1.1	1.9	-0.4	17.6	7.0
Mx1	0.0	-1.7	100.0	54.8	44.8	48.6	17.2	8.0	24.5
Oasl1	0.0	-1.2	100.0	147.0	123.8	102.1	32.6	24.2	61.9
Saa3	0.0	4.2	100.0	499.9	621.8	485.8	49.9	410.9	613.8
Ccl5	0.0	0.9	100.0	42.6	41.6	52.5	3.4	14.2	62.9
Fpr1	0.0	3.5	100.0	712.4	816.2	1235.6	245.9	783.0	1147.3

B

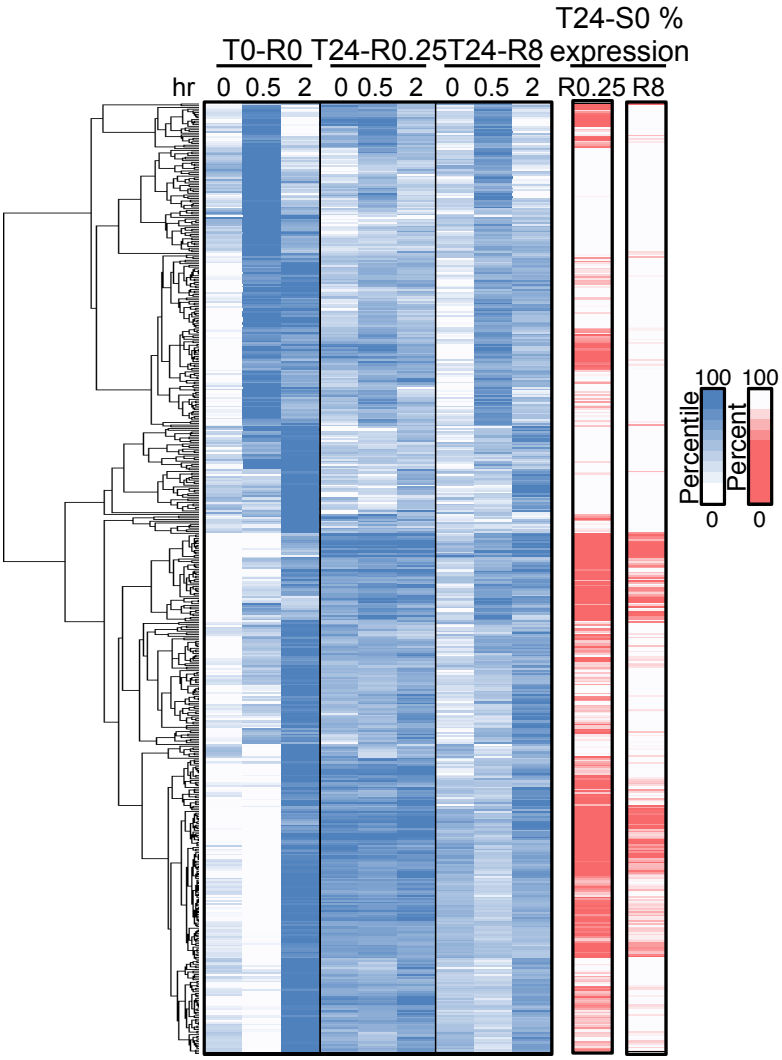
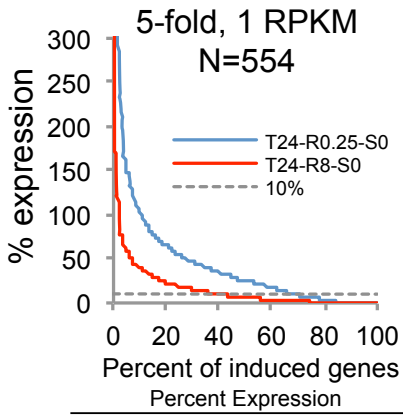


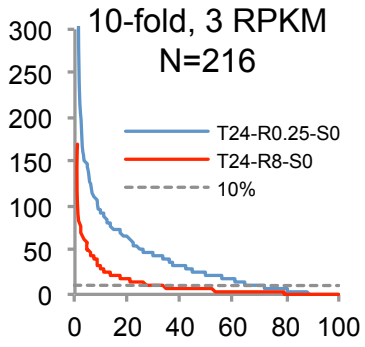
Figure 3-2 (continued)

C



	Percent Expression				
	< 10	10-20	20-50	50-90	> 90
T24-R0.25-R0	175	62	169	80	68
T24-R8-S0	334	93	91	24	12

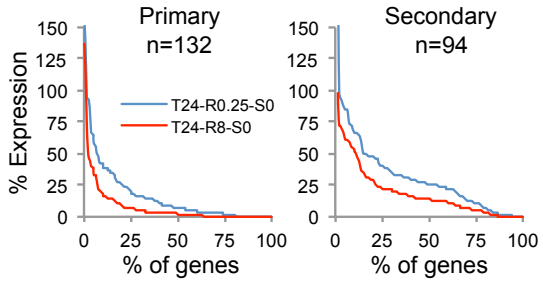
D



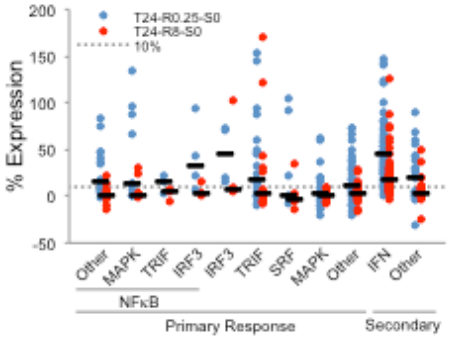
	Percent expression				
	< 10	10-20	20-50	50-90	> 90
T24-R0.25-R0	70	27	65	31	23
T24-R8-S0	151	30	26	6	3

Figure 3-3: LPS-inducible Genes with Sustained Expression Tend To Be Interferon-Dependent Secondary Response Genes

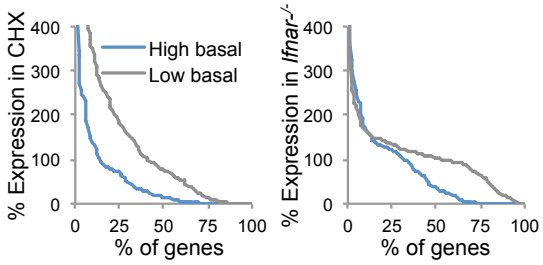
A



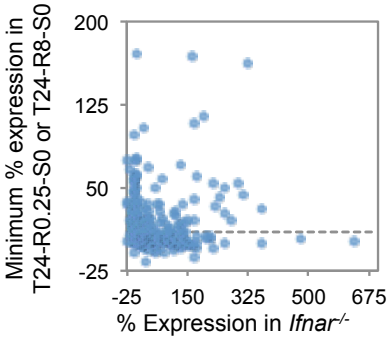
B



C



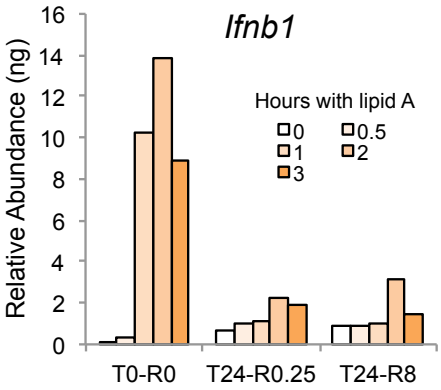
D



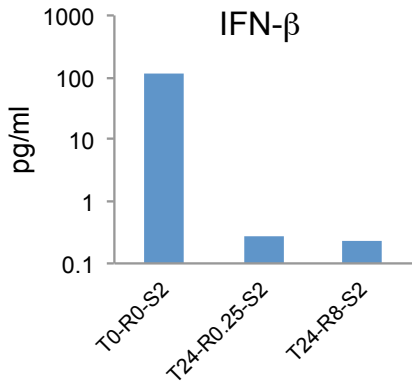
Min. % expression in T24-R0.25-S0 or T24-R8-S0	IFNAR % Expression			
	< 10	10-33	33-50	> 50
< 50	15	1	1	9
25-50	23	3	2	15
10-25	32	12	7	18
< 10	21	24	11	59

Figure 3-4: Persistent IFN-β Contributes to Prolonged Transcription of ISGs

A



B



C

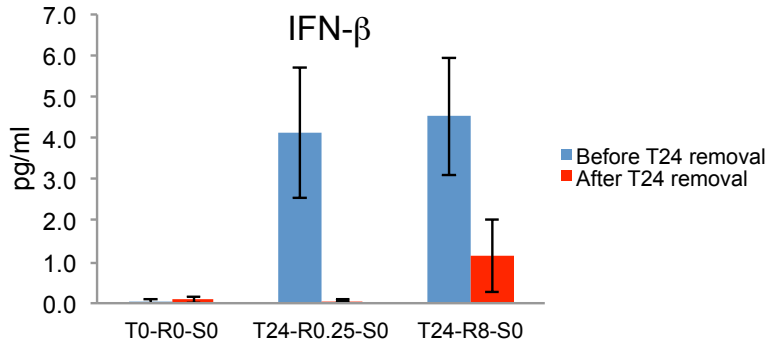
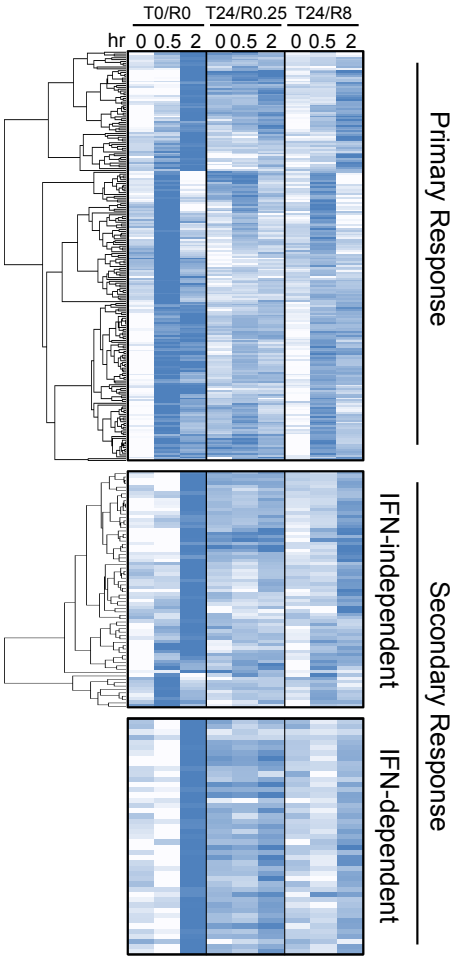


Figure 3-5: LPS Tolerance Broadly Impacts Inflammatory Gene Induction

A



B

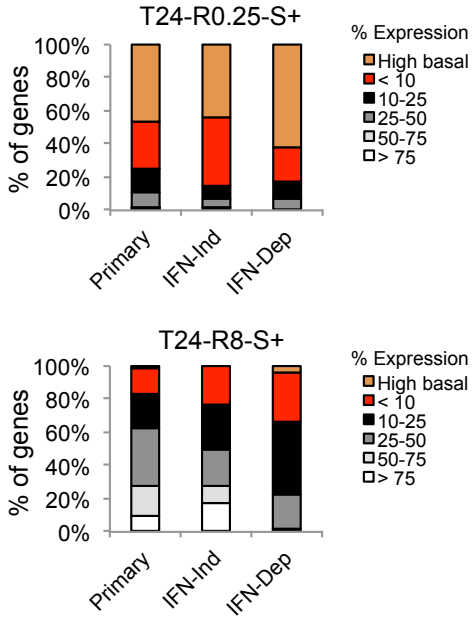
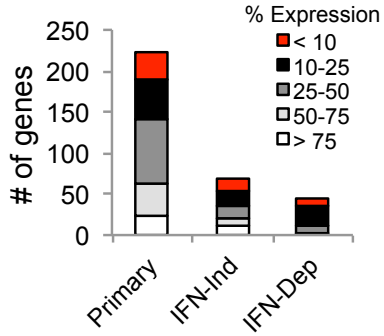
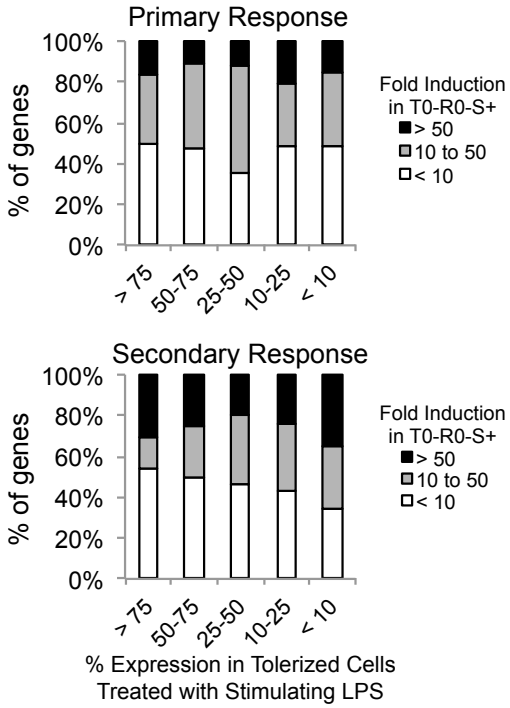


Figure 3-5 (continued)

C



D



E

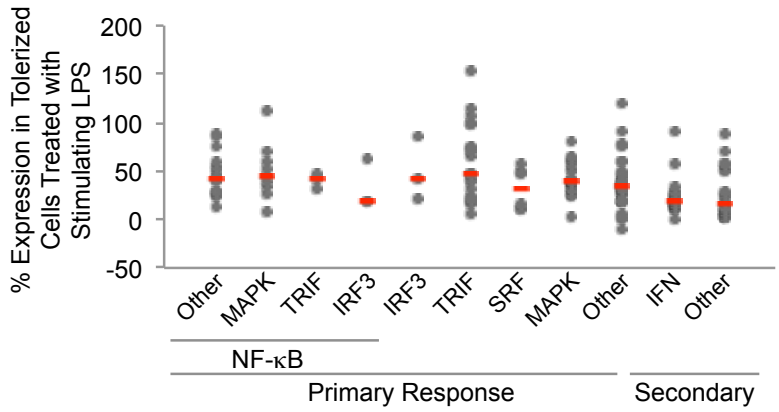
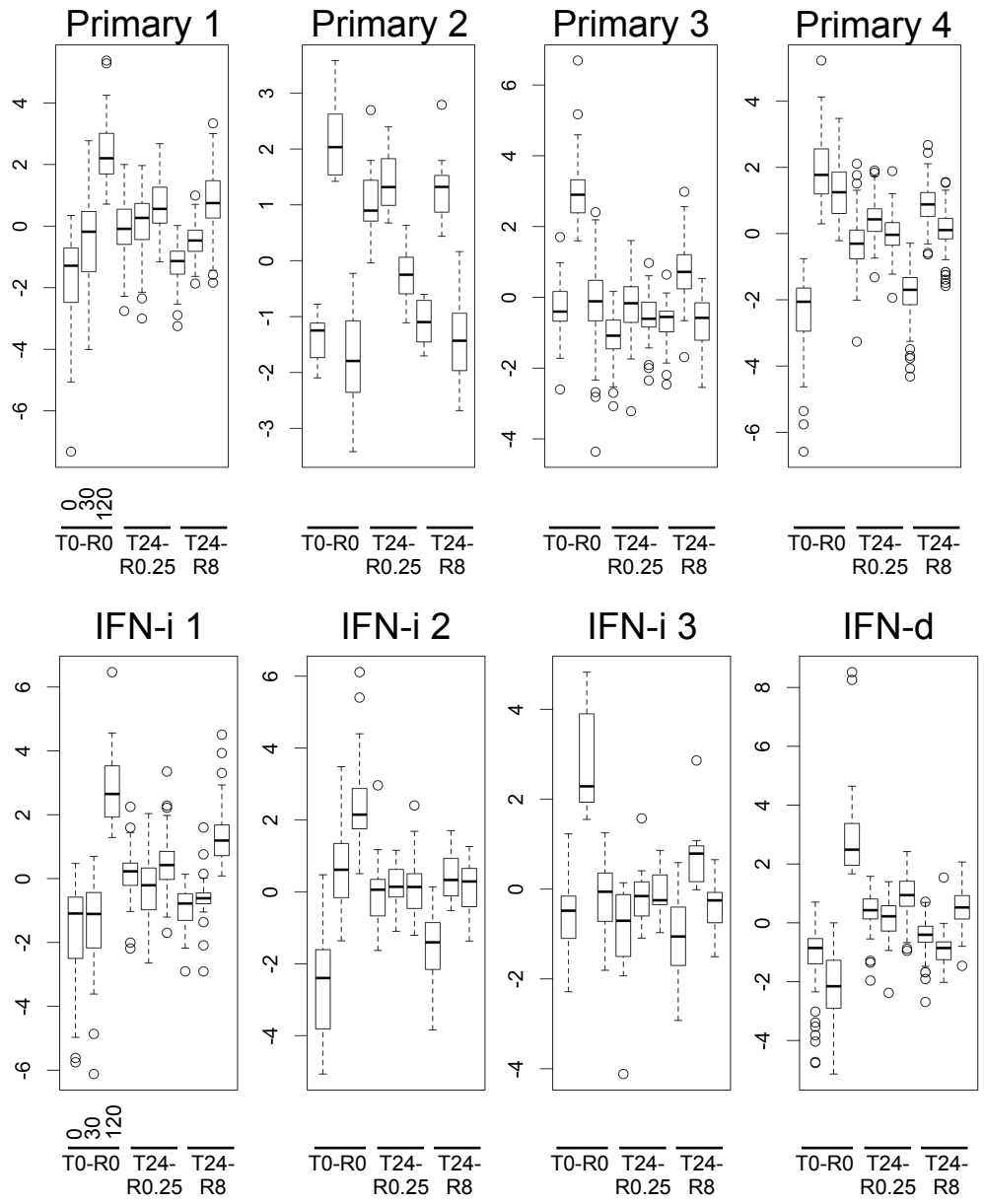


Figure 3-6: The Effect of Tolerance Wears Off Over Time

A



B

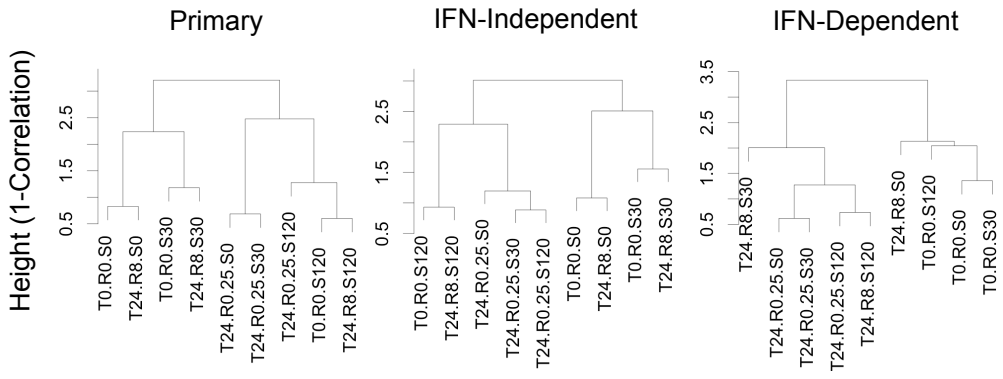


Table 3-1: Experimental Design and Sample Name Designations of Global Analysis of LPS Tolerance

	Name	Tolerizing LPS (T)	Rest period (R)	Stimulating LPS (S)
1	T0-R0-S0	0	0	0
2	T0-R0-S0.5	0	0	0.5
3	T0-R0-S2	0	0	2
4	T24-R0.25-S0	24	0.25	0
5	T24-R0.25-S0.5	24	0.25	0.5
6	T24-R0.25-S2	24	0.25	2
7	T24-R8-S0	24	8	0
8	T24-R8-S0.5	24	8	0.5
9	T24-R8-S2	24	8	2

Table 3-2: Expression of TLR4 Signaling Molecules and Related Factors During LPS Tolerance

		Reads Per Kilobase x Total Mapped Reads (RPKM)									Fold Induction	
Product	**Gene**	**NT 0**	**NT 30**	**NT 120**	**T15 0**	**T15 30**	**T15 120**	**T8 0**	**T8 30**	**T8 120**	**T15 0 / NT 0**	**T8 0 / NT 0**
Irak4	*Irak4*	1.3	0.5	0.6	1.4	1.1	1.1	1.7	1.0	1.4	1.1	1.3
Irak1	*Irak1*	4.3	3.3	2.6	3.0	3.4	2.4	3.8	3.6	3.0	0.7	0.9
p105 (p50)	*Nfkb1*	3.6	14.5	24.4	5.3	12.1	17.0	3.9	11.9	16.1	1.5	1.1
p100 (p52)	*Nfkb2*	3.6	30.1	36.9	7.7	17.4	17.4	4.1	17.7	17.3	2.1	1.1
RelA	*RelA*	3.8	14.1	9.3	4.6	8.6	6.3	3.2	11.4	5.5	1.2	0.8
cRel	*Rel*	2.4	12.9	19.6	4.2	7.9	6.9	3.0	11.1	8.5	1.8	1.3
Myd88	*Myd88*	2.3	4.8	12.1	4.4	7.0	7.1	4.1	6.7	5.4	1.9	1.8
Traf6	*Traf6*	1.7	2.2	1.8	2.0	2.0	2.0	1.8	2.1	1.9	1.2	1.1
Mal	*Mal*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.9
Tirap	*Tirap*	1.0	0.2	1.5	1.0	0.6	1.2	1.0	0.4	1.5	0.9	0.9
Tbk1	*Tbk1*	5.2	9.0	8.6	9.8	11.9	9.1	7.0	8.8	7.3	1.9	1.3
IKKβ	*Ikbkb*	3.1	6.1	8.5	3.4	4.4	5.5	3.1	4.8	6.6	1.1	1.0
IKKγ	*IKBKg*	2.2	1.4	3.1	3.0	2.8	3.4	1.8	1.4	2.1	1.4	0.8
Tram1	*Tram1*	3.7	0.5	0.8	3.1	2.1	2.3	3.1	1.3	2.2	0.9	0.9
Tlr4	*Tlr4*	3.3	0.4	1.2	5.6	2.6	3.4	4.7	0.9	1.7	1.7	1.4
Lbp	*Lbp*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	1.2
Cd14	*Cd14*	11.9	98.7	18.8	42.6	68.1	37.8	13.4	88.6	22.1	3.6	1.1
Md2	*Ly96*	1.5	0.9	2.2	2.6	2.7	2.8	2.1	1.8	2.6	1.8	1.5
Trem1	*Trem1*	0.1	0.7	0.0	0.8	3.1	0.7	0.3	2.4	0.2	6.2	2.3
Itgb2	*Itgb2*	6.8	1.9	0.5	10.7	11.2	6.5	8.6	7.3	4.6	1.6	1.3
Usp18	*Usp18*	2.4	0.5	28.6	15.0	9.7	22.8	10.6	3.6	10.7	6.2	4.4
Tram2	*Tram2*	1.0	0.8	1.3	0.6	0.6	1.0	0.6	0.5	0.8	0.6	0.6
Irak2	*Irak2*	4.4	14.9	18.8	9.1	14.5	11.4	5.1	17.6	13.6	2.1	1.2
Tak1	*Map3k7*	2.3	1.9	1.6	2.8	2.8	2.7	2.9	2.5	2.8	1.2	1.2
Tab2	*Tab2*	6.4	9.2	16.8	7.7	7.1	8.6	7.3	7.1	10.1	1.2	1.1
Tab1	*Tab1*	3.6	1.1	1.0	1.7	1.4	1.3	2.4	1.5	1.4	0.5	0.7
Mekk3	*Map3k3*	3.2	1.6	0.8	2.5	1.6	1.9	3.6	2.4	2.9	0.8	1.1
Mkk3	*Map2k3*	4.8	11.2	0.9	4.9	7.7	3.2	3.3	6.3	1.4	1.0	0.7
Mkk6	*Map2k6*	0.1	0.1	0.0	0.1	0.1	0.2	0.1	0.1	0.1	0.5	0.6
Mkk7	*Map2k7*	5.4	4.2	3.2	5.1	5.4	4.8	4.7	4.7	4.3	0.9	0.9
p38	*Mapk14*	5.6	2.2	3.6	4.5	4.2	5.4	6.5	4.1	7.2	0.8	1.2
Jnk	*Mapk8*	1.2	0.8	0.5	1.2	1.3	1.4	1.1	1.1	1.3	1.0	0.9
Erk	*Mapk14*	5.6	2.2	3.6	4.5	4.2	5.4	6.5	4.1	7.2	0.8	1.2

Table 3-3: Expression of Anti-Inflammatory Mediators During LPS Tolerance

		Reads Per Kilobase x Total Mapped Reads (RPKM)									Fold Induction	
Product	**Gene**	**NT 0**	**NT 30**	**NT 120**	**T15 0**	**T15 30**	**T15 120**	**T8 0**	**T8 30**	**T8 120**	**T15 0 / NT 0**	**T8 0 / NT 0**
Il10	*Il10*	0.5	2.2	2.2	0.8	0.9	0.7	1.2	1.1	0.7	1.6	2.4
Tgfb1	*Tgfb1*	6.8	7.2	2.1	4.3	5.5	4.6	5.1	8.3	4.1	0.6	0.8
PI3K p85	*pik3r1*	2.8	3.9	2.2	2.2	2.4	1.8	2.7	2.6	1.8	0.8	1.0
PI3K p110	*pik3ca*	1.7	2.3	1.3	2.8	2.2	3.1	2.6	1.8	3.4	1.6	1.5

Table 3-4: Expression of Negative Regulators of TLR4 Signal Transduction During LPS Tolerance

		Reads Per Kilobase x Total Mapped Reads (RPKM)									Fold Induction	
Product	**Gene**	**NT 0**	**NT 30**	**NT 120**	**T15 0**	**T15 30**	**T15 120**	**T8 0**	**T8 30**	**T8 120**	**T15 0 / NT 0**	**T8 0 / NT 0**
Tyro3	*Tyro3*	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.3
ST2L	*Il1rl1*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3
TRAILR	*Tnfrsf10b*	0.1	0.2	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.6
Sigirr	*Sigirr*	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.6
Mer	*Mertk*	2.6	1.2	2.8	3.4	1.9	2.3	3.3	2.2	1.8	1.3	1.3
Axl	*Axl*	2.0	1.9	9.7	1.8	1.8	2.3	1.6	2.2	2.1	0.9	0.8
Tollip	*Tollip*	2.4	2.5	1.0	2.3	2.8	2.0	2.1	2.4	2.1	0.9	0.9
IRAK-M	*Irak3*	1.1	1.6	3.2	4.4	7.9	12.3	2.7	4.4	11.6	3.9	2.3
SHIP1	*Inpp5d*	11.6	6.7	0.3	11.9	10.6	4.4	11.1	11.7	2.9	1.0	1.0
Socs1	*Socs1*	0.4	0.1	9.3	0.9	0.2	1.9	0.6	0.1	0.9	2.5	1.7
Mkp1	*Dusp1*	10.7	103.4	7.4	31.5	32.3	13.8	13.9	50.4	9.6	2.9	1.3
Atf3	*Atf3*	13.8	49.5	42.0	20.6	14.0	21.2	15.0	26.2	12.9	1.5	1.1
Trafd1	*Trafd1*	6.0	3.0	34.7	18.8	17.6	22.2	13.6	8.4	15.2	3.1	2.3
Socs3	*Socs3*	0.4	38.7	44.4	3.3	14.3	18.1	1.8	20.5	21.0	7.9	4.3
IκBα	*Nfkbia*	11.5	154.1	106.3	18.8	64.9	46.5	16.6	76.2	69.4	1.6	1.4
IκBβ	*Nfkbib*	1.3	26.0	4.4	2.5	8.2	3.4	1.3	12.7	3.8	1.9	1.0
IκBε	*Nfkbie*	2.0	30.3	21.0	9.5	18.1	11.6	2.8	27.8	12.4	4.8	1.4
IκB–NS	*Nfkbid*	11.6	38.8	6.1	9.1	13.3	9.0	8.0	25.4	7.4	0.8	0.7
IκBζ	*Nfkbiz*	6.9	161.6	58.5	18.0	57.7	22.5	12.2	90.0	31.6	2.6	1.8
A20	*Tnfaip3*	3.8	168.0	86.7	10.3	57.5	23.2	5.5	65.4	22.9	2.7	1.4

Works Cited

1. Beeson, P. B. Tolerance to Bacterial Pyrogens: I. Factors Influencing Its Development. *J. Exp. Med.* **86,** 29–38 (1947).

2. Beeson, P. B. Tolerance to Bacterial Pyrogens: II. Role of the Reticulo-Endothelial System. *J. Exp. Med.* **86,** 39–44 (1947).

3. Nomura, F. *et al.* Cutting Edge: Endotoxin Tolerance in Mouse Peritoneal Macrophages Correlates with Down-Regulation of Surface Toll-Like Receptor 4 Expression. *J. Immunol.* **164,** 3476–3479 (2000).

4. Biswas, S. K. & Tergaonkar, V. Myeloid differentiation factor 88-independent Toll-like receptor pathway: Sustaining inflammation or promoting tolerance? *Int. J. Biochem. Cell Biol.* **39,** 1582–92 (2007).

5. Sato, S. A variety of microbial components induce tolerance to lipopolysaccharide by differentially affecting MyD88-dependent and -independent pathways. *Int. Immunol.* **14,** 783–791 (2002).

6. Escoll, P. *et al.* Rapid up-regulation of IRAK-M expression following a second endotoxin challenge in human monocytes and in monocytes isolated from septic patients. *Biochem. Biophys. Res. Commun.* **311,** 465–472 (2003).

7. Kobayashi, K. *et al.* IRAK-M Is a Negative Regulator of Toll-like Receptor Signaling. *Cell* **110,** 191–202 (2002).

8. Van ’t Veer, C. *et al.* Induction of IRAK-M Is Associated with Lipopolysaccharide Tolerance in a Human Endotoxemia Model. *J. Immunol.* **179,** 7110–7120 (2007).

9. Kravchenko, V. V, Steinemann, S., Kline, L., Feng, L. & Ulevitch, R. J. Endotoxin tolerance is induced in Chinese hamster ovary cell lines expressing human CD14. *Shock* **5,** 194–201 (1996).

10. Wahlstrom, K., Bellingham, J., Rodriguez, J. L. & West, M. A. Inhibitory kappaBalpha control of nuclear factor-kappaB is dysregulated in endotoxin tolerant macrophages. *Shock* **11,** 242–247 (1999).

11. Medvedev, A. E., Kopydlowski, K. M. & Vogel, S. N. Inhibition of Lipopolysaccharide-Induced Signal Transduction in Endotoxin-Tolerized Mouse Macrophages: Dysregulation of Cytokine, Chemokine, and Toll-Like Receptor 2 and 4 Gene Expression. *J. Immunol.* **164,** 5564–5574 (2000).

12. Ziegler-Heitbrock, L. The p50-homodimer mechanism in tolerance to LPS. *J. Endotoxin Res.* **7,** 219–222 (2001).

13. Kastenbauer, S. & Ziegler-Heitbrock, H. W. L. NF-kappa B1 (p50) Is Upregulated in Lipopolysaccharide Tolerance and Can Block Tumor Necrosis Factor Gene Expression. *Infect. Immun.* **67,** 1553–1559 (1999).

14. Ziegler-Heitbrock, H. W. *et al.* Tolerance to lipopolysaccharide involves mobilization of nuclear factor kappa B with predominance of p50 homodimers. *J. Biol. Chem.* **269,** 17001–4 (1994).

15. Kraatz, J., Clair, L., Rodriguez, J. L. & West, M. A. In vitro macrophage endotoxin tolerance: defective in vitro macrophage map kinase signal transduction after LPS pretreatment is not present in macrophages from C3H/HeJ endotoxin resistant mice. *Shock* **11,** 58–63 (1999).

16. Sfeir, T., Saha, D. C., Astiz, M. & Rackow, E. C. *Role of interleukin-10 in monocyte hyporesponsiveness associated with septic shock. Critical care medicine* **29,** (2001).

17. Frankenberger, M., Pechumer, H. & Ziegler-Heitbrock, H. W. Interleukin-10 is upregulated in LPS tolerance. *J. Inflamm.* **45,** 56–63 (1995).

18. Del Fresno, C. *et al.* Potent phagocytic activity with impaired antigen presentation identifying lipopolysaccharide-tolerant human monocytes: demonstration in isolated monocytes from cystic fibrosis patients. *J. Immunol.* **182,** 6494–507 (2009).

19. Rosenzweig, H. L. *et al.* Endotoxin preconditioning protects against the cytotoxic effects of TNFalpha after stroke: a novel role for TNFalpha in LPS-ischemic tolerance. *J. Cereb. Blood Flow Metab.* **27,** 1663–74 (2007).

20. Hafenrichter, D. G., Roland, C. R., Mangino, M. J. & Flye, M. W. The Kupffer cell in endotoxin tolerance: mechanisms of protection against lethal endotoxemia. *Shock* **2,** 251–256 (1994).

21. Foster, S. L., Hargreaves, D. C. & Medzhitov, R. Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* **447,** 972–8 (2007).

22. Biswas, S. K. & Lopez-Collazo, E. Endotoxin tolerance: new mechanisms, molecules and clinical significance. *Trends Immunol.* **30,** 475–87 (2009).

23. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–8 (2008).

24. Bhatt, D. M. *et al.* Transcript Dynamics of Pro-Inflammatory Genes Uncovered by RNA-Seq Analysis of Subcellular RNA Fractions. *Cell* **In Press,** 1–3 (2012).

25. Gough, D. J., Messina, N. L., Clarke, C. J. P., Johnstone, R. W. & Levy, D. E. Constitutive type I interferon modulates homeostatic balance through tonic signaling. *Immunity* **36,** 166–74 (2012).

26. Carow, B. & Rottenberg, M. E. SOCS3, a Major Regulator of Infection and Inflammation. *Front. Immunol.* **5,** 58 (2014).

27. Yoshimura, A., Naka, T. & Kubo, M. SOCS proteins, cytokine signalling and immune regulation. *Nat. Rev. Immunol.* **7,** 454–465 (2007).

CHAPTER 4

Concluding Remarks:

Future Perspectives and Directions

Although Toll-like receptors (TLRs) were discovered less than 20 years ago[1], an extensive number of studies related to TLRs now exist and continues to rapidly grow. The studies now provide us with the understanding that each member of the TLR family recognizes a unique and conserved pathogen-associated molecular pattern (PAMP) derived from foreign material. This recognition initiates numerous complex signaling events, resulting in the activation of many transcription factors that coordinate a stimulus-specific transcriptional output. Furthermore, mechanisms at the post-transcriptional, translational, and post-translational levels also exist to regulate the innate immune response. Importantly, Charles Janeway postulated over 20 years ago that the innate immune system is the critical link to adaptive immunity, as antigen alone is not sufficient for activation[2]. Indeed, it is now appreciated that the innate immune response activates adaptive immunity through cytokine release, antigen processing, and upregulation of co-stimulatory molecules.

An appropriately balanced response is required for effective immunity. For example, a rapid and efficient response is necessary to counteract foreign pathogens and dangers, but excessive or prolonged responses can have pronounced damage to host tissues that can lead to chronic inflammatory disorders, autoimmune disorders, and tumor development. Current therapies are in use that target cytokines and signaling cascades to treat inflammatory diseases such as Crohn's disease, colitis, and arthritis[3]. However, because these therapies target broad cytokine signaling pathways that can affect a wide array of targets, patients undergoing these treatment plans are often susceptible to opportunistic infections. Therefore, a more refined understanding of the precise transcriptional events that occur in response to pathogen recognition and how specificity of the response is achieved in innate immune cells is necessary to be able to develop more specific therapies for the future.

The innate immune response is initiated by the transcriptional upregulation of hundreds of genes that is stimulus and cell-type specific. However, the mechanisms underlying selectivity

of the response remains unresolved. Additionally, the field of innate immunity continues to grow as more stimuli, signaling molecules, and transcription factors are identified and implicated to play indispensible roles in mediating inflammatory responses. This has resulted in an extremely complex picture of the events that occur, beginning from pathogen recognition to activation of numerous signaling molecules and transcription factors that mediate the upregulation of hundreds of genes to resolve infection. Therefore, in order to make progress towards understanding how selectivity of the innate immune response is achieved, a more careful examination of the contributions from various signaling and transcriptional cascades is necessary.

The studies described in the dissertation attempt to uncover the mechanisms and regulatory logic underlying transcriptional activation of genes during the TLR4-mediated innate immune response. Previous work from the lab began to assemble a framework describing the mechanisms of gene activation in response to LPS in macrophages, and demonstrated that LPS-inducible genes could be classified based on promoter properties, chromatin conformation, and transcription factor dependence[4]. However, because the studies were performed on a limited number of well-known proinflammatory genes using quantitative real-time PCR (qRT-PCR), whether or not the findings were relevant on a global scale at all LPS-inducible genes were unclear.

In Chapter 2, we expanded as well as refined these studies. Recent advancements in technology have transformed the fields of genetics and molecular biology. In particular, high-throughput sequencing methods such as RNA-seq and ChIP-seq have allowed for a deeper understanding of the principles governing innate immune responses at a genome-wide level. Although microarray-based studies have greatly contributed to our knowledge of molecular immunology, next-generation sequencing methods are beneficial because they produce highly specific and quantitative information, and are not limited to probes of known genomic regions.

Moreover, studies from our lab demonstrated that RNA populations from subcellular fractions could be isolated biochemically and subjected to RNA-seq[5,6]. This revealed a number of insights into the dynamic regulation of gene expression cascades at the levels of transcription, RNA processing, and RNA transport. One key finding was that RNA-seq analysis chromatin-associated transcripts isolated from stimulated cells provided a highly quantitative and accurate view of the activation kinetics of inducible genes. Using the approach of analyzing chromatin-associated transcripts in stimulated cells by RNA-seq, LPS-inducible genes were classified into primary and secondary response based on their requirement for new protein synthesis. Additionally, using BMDMs derived from gene knockout mice provided highly quantitative information as to what subsets of genes require signaling cascades such as MyD88, TRIF, IRF3, or MAPK. The inducible genes were further characterized by taking a unique approach to analyze the genes in the context of ChIP-seq and transcription factor binding motif datasets. This revealed critical insights into the quantitative relationship between ChIP-seq peaks and transcription factor binding motifs. Furthermore, using these methods, putative NF-κB, IRF3, and SRF target genes were identified. These findings in conjunction with their expression dynamics and requirement for the signaling pathways and transcription factors known to activate them provided strong evidence for their regulation. This approach provided the ability to connect a transcription factor to a set of target genes with higher confidence than previously appreciated, and should advance our understanding of how selective immune responses to pathogens are achieved.

The studies presented in Chapter 3 demonstrated how the framework established in Chapter 2 could be applied to a physiological setting by examining the mechanisms of LPS tolerance. Notably, a global downregulation of inflammatory gene activation in tolerized macrophages given a second LPS treatment was observed. The finding that genes encoding for numerous negative regulators of the TLR4 response were upregulated, such as those blocking

proximal signal transduction events at the TLR4 receptor, in tolerized cells suggest that multiple mechanisms targeting a broad range of signaling cascades likely coordinate the tolerant state. In addition, the interferon receptor-dependent genes identified in Chapter 2 remained highly transcribed after the tolerizing LPS was removed. This was partially explained by the finding that IFN-β was present at low levels in the tolerant state. Thus, although LPS tolerance was exerting its effects downstream of TLR4, many secondary response genes could still be transcribed due to prolonged half-lives of their regulators that were induced during the primary response. Further studies focusing on candidate regulators should expand our understanding of the degree to which these factors contribute to the tolerant state.

The studies presented in the dissertation describe how signaling pathways and transcription factors contribute to the TLR4 transcriptional activation network. Importantly, the framework is not comprehensive but serves as a foundation to begin understanding how selectivity of innate immune responses is achieved. In this chapter, I briefly discuss future directions that should advance our understanding of selective transcription during immune responses.

Importantly, the mechanisms regulating inflammatory gene activation are not limited to the signaling cascades discussed in the dissertation. Although numerous transcription factors and signaling pathways have been implicated to be critical for the inflammatory response, how these various cascades act together to coordinate gene activation is unclear. Therefore, it will be critical to continue dissecting the TLR4 response through whole-genome profiling of activated macrophages deficient for various signaling pathways. This would most certainly refine the framework into additional co-regulated classes and would provide a deeper understanding of the inflammatory gene program.

A critical aspect of the TLR4 transcriptional network that needs to be addressed is the role that chromatin plays in regulating selective transcription in the context of the described framework. First, previous studies from the lab highlighted the important role that chromatin plays in mediating selective gene activation. A subset of TLR4-inducible genes including *Ccl5* requires chromatin remodeling by the SWI/SNF remodeling complex to induce promoter accessibility that is necessary for its activation. The findings also coincided with the delayed activation kinetics of these genes, supporting the hypothesis that nucleosome remodeling at selective promoters allows access for transcription factors to bind and recruit the transcription machinery. Indeed, the ChIP-seq studies described in Chapter 2 demonstrated that RelA bound to the promoter of *Ccl5* with delayed kinetics relative to other RelA target genes, likely because RelA did not have access to DNA until SWI/SNF complexes reconfigured the promoter to an open and accessible conformation. Along these lines, it will be informative to understand in the context of the framework which subset of TLR4 responsive genes remodel their promoters in response to lipid A through genome-wide DNase hypersensitivity approaches, as well as identifying the subset of these genes that require the SWI/SNF chromatin remodeling complex or other remodeling factors.

Secondly, there is still a limited understanding of the role that histone modifications play in regulating inducible gene activation. Numerous studies of the epigenetic landscape in macrophages have demonstrated extensive diversity of histone modifications present in resting cells, as well as modifications that are inducibly gained or lost after stimulation[7]. Importantly, specific histone modifications are more often found at a subset of rather than at all inflammatory genes. Both the diversity and specificity of these epigenetic markers indicates that distinct mechanisms likely exist to regulate select subsets of genes to achieve a transcriptional output that is tailored towards the stimulus detected. Therefore, further investigation of the role of

histone modifications in inducible transcription will reveal properties of select subsets of genes and should provide insight into their unique regulatory mechanisms.

Selectivity of inducible gene activation occurs not only in response to different stimuli, but in different cell types responding to the same stimulus. For example, preliminary studies from our lab have demonstrated that *Il6* transcription is a secondary response in LPS-stimulated macrophages, but a primary response in endothelial cells stimulated with LPS[4]. Furthermore, the findings described in Chapter 2 suggest that a subset of putative NF-κB target genes may be activated in a cell type-specific manner. A large proportion of the promoter-regulated NF-κB genes encode for NF-κB family members and regulators itself, suggesting a possibility that they may be activated by diverse stimuli and in various cell types. Notably, a subset of primary response genes exhibited similar activation kinetics and RelA dependence, yet did not have promoter-bound RelA in response to stimulus. This implies that RelA may be functioning at distal enhancer elements that are established through binding of lineage-determining factors such as PU.1 in macrophages to induce a cell type-specific response. Genome-wide profiling of diverse cell types stimulated with LPS will reveal if promoter- or putative enhancer-regulated NF-κB targets are differentially expressed between the cell types. Furthermore, because enhancers are important in mediating tissue-specific responses, detailed enhancer analysis by examining the overlap between distal RelA binding peaks, PU.1, and enhancer markers such as H3K4me1 should further clarify the prevalence of RelA binding at enhancers. Mapping enhancers to their target genes is a major challenge in the field that has been difficult to overcome because enhancers can regulate its targets as much as hundreds of kilobases upstream or downstream from the transcriptional start site and do not necessarily regulate the closest gene[8]. Chromatin-conformation studies such as Hi-C will be useful to decipher the three-dimensional organization of chromosomes in the LPS-inducible system. While these studies will certainly advance our understanding of the interactions made between distal functional

elements such as enhancers and their putative target genes, additional studies will be necessary to determine the in vivo functional relevance by connecting these interactions with changes in gene expression. The recent advancement of engineered nucleases such as the CRISPR/Cas9 system that enables rapid genome editing[9] has already begun to transform biological research. Using this strategy to disrupt sequence-specific binding sites at enhancers should reveal the functional significance they have in regulating expression of their target genes, and will undoubtedly play a critical role in advancing our understanding of mechanisms that establish cell type-specific responses.

It is clear that our understanding of selective transcription during the innate immune response has only just begun. However, technological advances have come to a point where the *cis*-elements and *trans*-factors for every gene can be investigated to better understand how they coordinate gene expression in response to an environmental change. Furthermore, investigating how these components interact in three-dimensional space, the functional consequences of these interactions, as well as the heterogeneity of inflammatory responses through single-cell studies will be essential to understand the selectivity of transcriptional activation in the innate immune system.

Works Cited

1. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. . A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. **388,** 394–397 (1997).

2. Janeway, C. A. Approaching the Asymptote? Evolution and Revolution in Immunology. *Cold Spring Harb. Symp. Quant. Biol.* 1–3 (1989). at <http://symposium.cshlp.org/content/54/1.full.pdf>

3. McInnes, I. B. & Schett, G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat. Rev. Immunol.* **7,** 429–42 (2007).

4. Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138,** 114–28 (2009).

5. Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150,** 279–290 (2012).

6. Pandya-Jones, A. *et al.* Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* **19,** 811–27 (2013).

7. Smale, S. T. Selective transcription in response to an inflammatory stimulus. *Cell* **140,** 833–44 (2010).

8. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488,** 116–20 (2012).

9. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–21 (2012).

APPENDIX A

Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions

Cell

Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions

Dev M. Bhatt,[1,5] Amy Pandya-Jones,[1,2,5] Ann-Jay Tong,[1,3] Iros Barozzi,[4] Michelle M. Lissner,[1] Gioacchino Natoli,[4] Douglas L. Black,[1,2,3,6,*] and Stephen T. Smale[1,3,6,*]
[1]Department of Microbiology, Immunology, and Molecular Genetics
[2]Howard Hughes Medical Institute
[3]Molecular Biology Institute
University of California, Los Angeles, Los Angeles, CA 90095, USA
[4]Department of Experimental Oncology, European Institute of Oncology, Milan I-20139, Italy
[5]These authors contributed equally to this work
[6]These authors contributed equally to this work
*Correspondence: dougb@microbio.ucla.edu (D.L.B.), smale@mednet.ucla.edu (S.T.S.)
http://dx.doi.org/10.1016/j.cell.2012.05.043

SUMMARY

Macrophages respond to inflammatory stimuli by modulating the expression of hundreds of genes in a defined temporal cascade, with diverse transcriptional and posttranscriptional mechanisms contributing to the regulatory network. We examined proinflammatory gene regulation in activated macrophages by performing RNA-seq with fractionated chromatin-associated, nucleoplasmic, and cytoplasmic transcripts. This methodological approach allowed us to separate the synthesis of nascent transcripts from transcript processing and the accumulation of mature mRNAs. In addition to documenting the subcellular locations of coding and noncoding transcripts, the results provide a high-resolution view of the relationship between defined promoter and chromatin properties and the temporal regulation of diverse classes of coexpressed genes. The data also reveal a striking accumulation of full-length yet incompletely spliced transcripts in the chromatin fraction, suggesting that splicing often occurs after transcription has been completed, with transcripts retained on the chromatin until fully spliced.

INTRODUCTION

Macrophages play an important role in regulating immune responses following exposure to a microbial insult or danger signal. Their response involves widespread changes in gene expression that are often tailored to the stimulus. Some inducible genes encode proteins that help polarize the subsequent innate and adaptive immune responses, whereas others encode effector molecules that protect the host from the insult. Microbial pathogens have acquired mechanisms to subvert immune responses, and deregulated inflammation has been associated with numerous diseases ranging from inflammatory autoimmune disorders to atherosclerosis and cancer. A major challenge has been to elucidate the molecular circuitry that regulates an inflammatory gene program, with the ultimate aim of selectively altering the response to diminish disease pathology while maintaining antimicrobial immunity (Medzhitov and Horng, 2009; Smale, 2010).

To understand the transcriptional response to an inflammatory stimulus, several studies have focused on the regulatory properties of primary response genes, which are rapidly induced in the absence of new protein synthesis (reviewed in Fowler et al., 2011). Primary response genes induced by Toll-like receptor (TLR) ligands and tumor necrosis factor (TNF) usually contain CpG island promoters. Such genes are associated with RNA polymerase II in unstimulated cells and possess histone tail modifications commonly found at the promoters of actively transcribed genes (Hargreaves et al., 2009; Ramirez-Carrozzi et al., 2009). These genes can be activated by inducible transcription factors in the absence of SWI/SNF-dependent nucleosome remodeling (Ramirez-Carrozzi et al., 2006, 2009).

Rapid transcriptional induction may also be facilitated by the active release of polymerase molecules that have initiated transcription in unstimulated cells but are paused near the transcription start site (TSS) (Nechaev and Adelman, 2011). Paused polymerase molecules can be released by signal-dependent recruitment of P-TEFb, phosphorylation of RNA polymerase II on serine 2 of the C-terminal domain (CTD), and release of the NELF repressor. Although some genes are regulated by pausing in close proximity to the TSS, others are reported to be efficiently transcribed into precursor transcripts, with inducible P-TEFb recruitment regulating the efficiency of transcript processing (Adelman et al., 2009; Hargreaves et al., 2009).

In contrast to the features described above, many secondary response genes and some primary response genes require SWI/SNF-dependent nucleosome remodeling for their induction (Ramirez-Carrozzi et al., 2006, 2009; Fowler et al., 2011). These

genes often lack CpG islands and are inaccessible to nucleases in unstimulated cells. Prior to stimulation, they also lack histone modifications characteristic of active genes. The remodeling requirement may facilitate highly selective regulation by imposing an additional barrier to activation.

A major long-term goal is to elucidate the precise network of events that coordinates the complex, temporally ordered gene expression cascade induced by an inflammatory stimulus. Progress toward this goal has been made primarily through the use of microarrays to examine temporal changes in messenger RNA (mRNA) levels, followed by efforts to connect transcription factors and signaling pathways to clusters of coregulated genes (Ramsey et al., 2008; Amit et al., 2009; Litvak et al., 2009). However, temporal changes in mRNA levels can diverge considerably from temporal changes in transcription. For example, a lag period can follow the maximal activation of transcription before the peak mRNA level is reached (Hao and Baltimore, 2009; Rabani et al., 2011). Microarrays also tend to underestimate the magnitudes of changes in mRNA levels, further limiting efforts to examine transcriptional cascades (Marioni et al., 2008).

RNA sequencing (RNA-seq) has emerged as an improved method for examining transcriptomes. Although this method is usually used to study mRNAs, three strategies have been described for the analysis of nascent transcripts, thereby allowing dynamic changes in transcription to be distinguished from changes in mRNA levels. First, the genome-wide nuclear run-on (GRO-seq) involves the incubation of nuclei with labeled nucleoside triphosphates, followed by high-throughput sequencing of complementary DNA (cDNA) prepared from the labeled nascent transcripts (Core et al., 2008; Hah et al., 2011). In a second method, native elongating transcript sequencing (NET-seq), ternary complexes containing RNA polymerase, the DNA template, and the nascent transcript are isolated by cross-linking, chromatin fragmentation, and polymerase immunoprecipitation, followed by high-throughput sequencing of cDNAs (Churchman and Weissman, 2011). Finally, metabolic labeling has been used to monitor the fate of nascent transcripts (Friedel and Dölken, 2009, and references therein). This method was recently employed to characterize clusters of coexpressed genes in mouse dendritic cells stimulated with lipopolysaccharide (LPS) (Rabani et al., 2011).

Here, we describe RNA-seq analysis of biochemically fractionated chromatin-associated, nucleoplasmic, and cytoplasmic transcripts from unstimulated and stimulated macrophages. Like the three methods described above, this method provides basic information about nascent transcription. However, the results have provided several additional insights of broad relevance concerning the dynamic regulation of RNA processing and transport and have allowed us to define, at high resolution, the properties of specific classes of coexpressed genes activated by an inflammatory stimulus.

RESULTS

RNA-Seq Analysis of Fractionated Transcripts

Inflammatory gene transcription was studied in mouse bone-marrow-derived macrophages stimulated with lipid A (the active component of LPS) for 0, 15, 30, 60, and 120 min. We fractionated cells (Wuarin and Schibler, 1994) to isolate chromatin-associated, nucleoplasmic, and cytoplasmic transcripts as done previously to analyze nascent transcripts from individual genes (Dye et al., 2006; Pawlicki and Steitz, 2008; Pandya-Jones and Black, 2009). In pilot studies, the purity of the chromatin, nucleoplasm, and cytoplasm was assessed by western blot analysis of β-tubulin, SNRP70, and histone H3, respectively (Figure S1A available online). To facilitate the analysis of nascent transcripts, an oligo-deoxythymine (dT) purification step was omitted during the isolation of RNA from the three fractions. However, in separate experiments, an oligo dT resin was used to isolate polyadenylated RNA from whole-cell macrophage lysates.

Random primed, strand-specific cDNA libraries prepared from the RNA populations were analyzed by high-throughput sequencing. Inspection of the aligned reads for individual genes provided strong support for the notion that the transcripts in the chromatin samples were newly transcribed. First, as illustrated for the inducible *Nfkb1* gene (Figure 1A), the reads from the chromatin samples were broadly distributed across both the exons and introns of many genes, whereas in nucleoplasmic and cytoplasmic samples, exon reads were strongly enriched. Reads were not observed at genomic regions flanking transcription units, demonstrating that the reads were derived from RNA rather than from DNA. Reads from oligo-dT-selected RNA showed a similar pattern of exon enrichment (Figure S1B). Also, transcript levels for inducible genes increased at earlier time points in the chromatin samples than in the nucleoplasmic and cytoplasmic samples (see Figure 1B for *Nfkb1*) or the mRNA (Figure S1C). This finding suggests that an analysis of chromatin transcripts provides a more accurate view of the kinetics of transcriptional induction than is provided by mRNA analyses.

Distinct Subcellular Transcript Profiles

We divided 10,020 expressed RefSeq genes (reads per kilobase per millions of mapped reads (RPKM) >1 in at least one of the three fractions) into 12 clusters based on their transcript profiles in the three subcellular fractions and five time points (Figure 2A). Most lipid-A-induced genes were found in clusters 1–3 (Figures 2A and 2B). Transcriptional downregulation was detected in clusters 4–6 (Figures 2A and 2B). Surprisingly, far more genes were downregulated than upregulated (Figure 2C). However, the average magnitude of downregulation in two clusters (clusters 4 and 6) was only slightly >2-fold in the chromatin sample (Figure 2B). It is noteworthy that reduced transcript levels were observed in the chromatin samples at earlier time points than in the nucleoplasmic and cytoplasmic samples (Figures 2A and 2C). This difference is presumably due to the time required for pre-existing mRNAs to decay. The remaining clusters showed little change in transcript levels during the time course (Figures 2A and 2B, clusters 7–12).

The cluster analysis also reveals extensive variability in the distribution of transcripts between the chromatin, nucleoplasm, and cytoplasm (Figure 2A). Although we cannot accurately calculate the percentage of transcripts for a gene in each of the three fractions, relative differences are apparent. Cluster 7, for example, contains abundant transcripts in the chromatin, with much lower transcript levels in the nucleoplasm and cytoplasm relative to the other clusters. This cluster is dominated

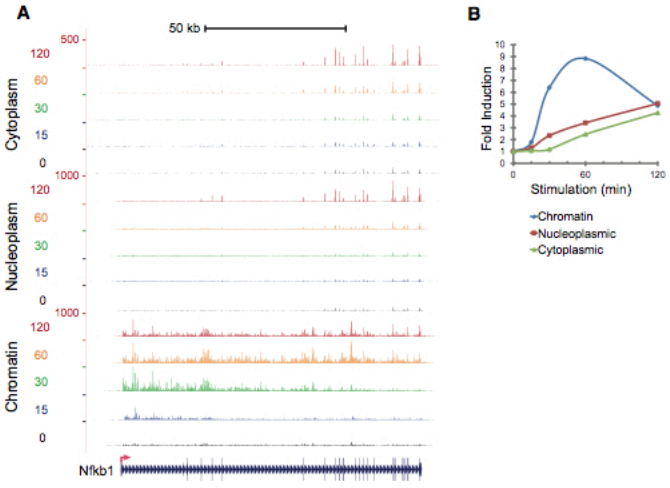


Figure 1. RNA-Seq Read Distributions and Induction Kinetics for *Nfkb1* Transcripts

(A) The distribution of RNA-seq reads at the *Nfkb1* locus is shown for libraries prepared from cytoplasmic, nucleoplasmic, and chromatin-associated RNA. Time points and scale are indicated at the left. Exons and introns are shown at the bottom, and the TSS is indicated (bent arrow).

(B) *Nfkb1* transcript levels in each of the three fractions through the time course are shown as fold-induction values relative to the unstimulated transcript values (determined from RPKMs).

See also Figure S1.

by annotated and unannotated noncoding transcripts and micro RNA (miRNA) precursors. (The analysis excluded transcripts shorter than 400 nucleotides and therefore excluded mature miRNAs and many miRNA precursors.) Some noncoding RNAs, such as *Xist*, are highly enriched in the chromatin because they function at this location (Figure S2). Examples of other noncoding RNAs in cluster 7 are *Neat1* and *Malat1/Neat2* (Figure S2). Further mining of the data sets may provide insights into the subcellular locations at which many noncoding RNAs function. Cluster 7 also includes transcripts from constitutive protein-coding genes that may be highly unstable and therefore present at low levels in the cytoplasm (e.g., *Leng8* in Figure S2).

Cluster 11, by contrast, contains abundant cytoplasmic transcripts, with fewer transcripts in the chromatin and nucleoplasm. This cluster includes genes that are likely to encode highly stable mRNAs, such as *Actb* (β-actin) and *Tuba* (α-tubulin).

Clusters 8 and 9 contain genes whose transcripts are more abundant in the nucleoplasmic fraction than in the chromatin or cytoplasm. Many genes in these clusters encode proteins that would be expected to be translated by ribosomes associated with the endoplasmic reticulum (ER; see Chen et al., 2011). Thus, their enrichment in this fraction is likely due to the copurification of ER-associated transcripts with the nuclear envelope.

Further analysis of individual transcripts or clusters of transcripts enriched in different subcellular fractions is likely to provide important insights into the role of RNA stability in gene regulation at a genome-wide scale, the functions of specific noncoding RNAs, and possibly other facets of RNA dynamics.

Fundamental Properties of the Lipid-A-Induced Transcriptome

For this analysis, we focused on lipid-A-induced genes. An examination of reads for ~26,000 RefSeq genes (normalized as RPKM) revealed changes in chromatin-associated transcript abundance through the time course (Figure S3A). By examining genes with a read coverage of at least 1 RPKM at one of the stimulated time points, we found 560 genes that were induced in the chromatin samples at least 5-fold during the 2 hr stimulation period. 246 genes were induced between 5- and 10-fold, 247 between 10- and 100-fold, and 67 more than 100-fold (Figure S3B). Peak chromatin transcript values were reached at the 15 min time point for only 16 genes, with 145, 94, and 304 genes reaching their highest levels at 30, 60, and 120 min, respectively (Figure S3C). We also determined the time point at which each gene exhibited its maximum fold increase, or "ramp," relative to the preceding time point. Although only 16 genes exhibited their peak chromatin transcript level at the 15 min time point, 154 genes showed maximum ramp values within 15 min of stimulation (Figure S3C).

To examine the reproducibility of the data sets, Pearson correlation values (R values) were determined for chromatin-associated transcripts from two independent experiments performed several months apart with macrophages from different mice (Figure S4). R values derived from a comparison of RPKMs for inducible genes equaled or exceeded 0.94 at each time point (Figure S4A). Hierarchical clustering showed that each time point correlated more closely with the corresponding time point from the second experiment than with any of the other time points (Figure S4B). Similar results were obtained when all expressed genes or all RefSeq genes were examined rather than all induced genes (Figures S4C–S4F).

The 560 lipid-A-induced genes were clustered into six classes, A–F, on the basis of the temporal profiles of their chromatin-associated transcripts (Figures 3A and 3B). We further subdivided class A into classes A1 and A2 to highlight, in class A1, the 16 genes that exhibited peak chromatin transcript levels at the 15 min time point in the two experiments. Interestingly, 10 of these 16 genes encode proteins that directly regulate gene expression: transcription factors (*Fos*, *FosB*, *Jun*, *Atf3*, *Egr1*, *Egr3*, and *Nr4a1* [NUR77]), transcriptional coregulators (*Nfkbid* [IκBNS] and *Btg2*), and an RNA-binding protein (*Zfp36* [TTP]). The other six participate in diverse functions (*Ppp1r15a*, *Plk3*, *Pmaip1*, *Ier2*, *Gdf15*, and *Fabp4*). Class A2 is largely composed

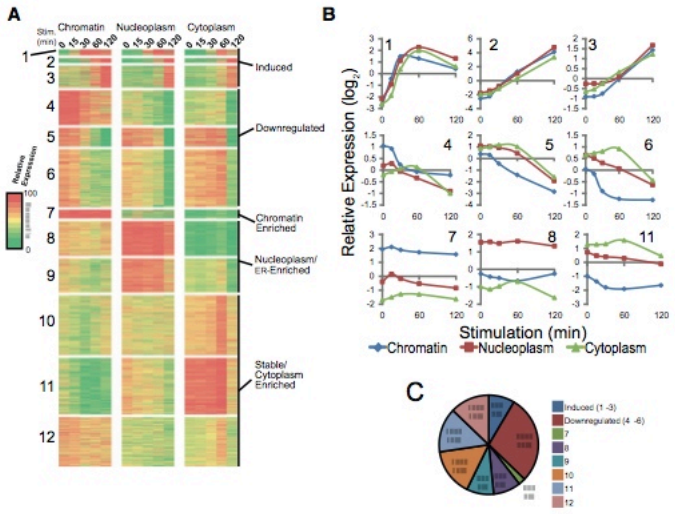


Figure 2. Distributions of Transcripts in the Three Subcellular Fractions

(A) RefSeq genes exceeding 400 bp in length and exhibiting an RPKM of at least one in one of the samples were divided into 12 clusters on the basis of their pattern of transcript levels in the three subcellular fractions and five time points. Colors indicate the percentile of the relative expression level.

(B) The average relative transcript levels within representative clusters are shown. Although relative cluster-to-cluster differences are of interest, absolute transcript abundances in each subcellular compartment cannot be determined because of possible variability in RNA isolation efficiencies from the three fractions.

(C) The number and percentage (parentheses) of genes within each of the 12 clusters in (A) are shown.

See also Figure S2.

of transiently induced genes that exhibit maximal chromatin transcripts 30 min poststimulation. Most class B genes peak at 60 min, with substantial decreases by 120 min. Class C genes, like class A2 genes, are induced early and usually peak at the 30 min time point, but unlike the transient induction of class A2 genes, transcription of class C genes is sustained. Classes D, E, and F exhibit increasingly later induction of chromatin-associated transcripts (Figure 3B).

Progression of Transcripts from Chromatin to Nucleoplasm to Cytoplasm

To monitor the progression of transcripts from the chromatin to the nucleoplasm and cytoplasm, we visualized the profiles of individual genes and analyzed complete coexpression classes. Nucleoplasmic and cytoplasmic transcripts (and mRNAs) largely followed the same temporal profile as the chromatin transcripts but with a clearly detectable delay (Figures 3B and S5). These data are consistent with those obtained by metabolic labeling (Rabani et al., 2011). Thus, transcription appears to be a dominant regulator of mRNA expression kinetics. We have been unable to identify genes that exhibit large changes in mRNA levels without corresponding changes in nascent transcripts, which would be suggestive of regulation primarily at the level of mRNA stability.

Promoter and Chromatin Properties within Coexpression Classes

Our ability to define gene classes from the temporal profiles of nascent transcripts allowed us to examine promoter and chromatin properties within each class. Previous studies suggested that rapidly induced genes frequently contain CpG island promoters, whereas secondary response genes and primary response genes induced with delayed kinetics usually contain low CpG (LCG) promoters (Hargreaves et al., 2009; Ramirez-Carrozzi et al., 2009). Consistent with these studies, 88%, 84%, and 71% of genes in classes A1, A2, and B contain CpG island promoters (Figures 4A and 4B). Interestingly, however, 36%–58% of genes in the remaining classes also contain CpG island promoters, which is higher than predicted from the previous studies of limited numbers of genes. It is noteworthy that a high prevalence of CpG island promoters correlated more closely with transient induction than with rapid induction. That is, classes A1, A2, and B contain a much higher percentage of CpG island promoters than class C (Figure 4B); class C is induced as rapidly as class A2 and more rapidly than class B, but class C transcription is sustained, which is in contrast to the transient induction of classes A1, A2, and B.

In unstimulated cells, inducible CpG island promoters often assemble into chromatin with features commonly associated with transcriptionally active genes, and it has been suggested that this property facilitates rapid transcriptional induction (Hargreaves et al., 2009; Ramirez-Carrozzi et al., 2009). Consistent with this hypothesis, histone H3 lysine 4 trimethylation (H3K4me3), a mark found at the promoters of active genes, was generally higher at CpG island promoters than at LCG promoters in unstimulated cells (Figures 4A and 4C; chromatin immunoprecipitation sequencing (ChIP-seq) data from De Santa et al., 2009). It is noteworthy, however, that this active chromatin mark was enriched at CpG island promoters in all classes. Similarly, RNA polymerase II association was more prevalent in unstimulated cells at CpG island promoters than at LCG promoters in all of the classes (Figures 4A and 4C).

An examination of the repressive histone H3K27me3 modification revealed association with a small fraction of promoters within all classes (Figure 4A). Interestingly, this repressive mark appears to be most prevalent in CpG island promoters that exhibit relatively low levels of the H3K4me3 mark.

Together, these results confirm that rapidly induced genes frequently contain CpG island promoters, high levels of

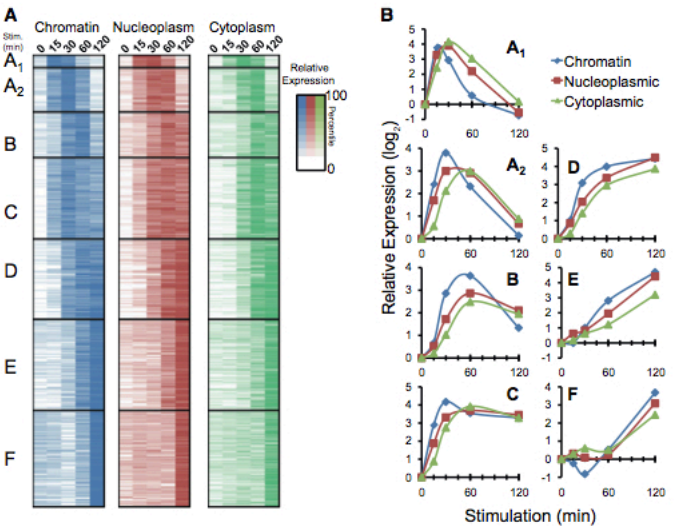


Figure 3. Kinetic Analysis of Lipid-A-Induced Transcripts

(A) Lipid-A-induced genes were divided into six classes (A–F) on the basis of their chromatin-associated transcript profiles. Class A was divided into classes A1 and A2 to highlight 16 genes that exhibit peak or near-peak transcript levels at the 15 min time point. Nucleoplasmic and cytoplasmic transcript levels were aligned on the basis of the clustering of chromatin transcripts. Transcript values are normalized to the average RPKM for each fraction. Colors indicate percentile values.

(B) The average fold induction within each class (y axis) is shown for each time point (x axis). Chromatin (blue), nucleoplasmic (red), and cytoplasmic (green) transcripts were analyzed separately.

See also Figures S3, S4, and S5.

H3K4me3, and RNA polymerase II association in unstimulated cells. However, these features were also found in many genes with delayed induction kinetics. A more detailed examination of the distinct regulatory capabilities conferred by CpG island and LCG promoters is presented below.

Dynamic Range of Inducible Transcription

It has been proposed that rapidly induced genes with active chromatin features in unstimulated cells are transcribed at relatively high levels prior to cell stimulation but that the elevated transcription does not lead to elevated mRNA levels because transcription elongation and RNA processing are inefficient prior to stimulation (Hargreaves et al., 2009). The RNA-seq analysis of chromatin-associated transcripts allowed us to test this hypothesis.

An examination of RPKMs revealed that basal nascent transcript levels span more than three orders of magnitude within most classes (Figure 5A). The median basal RPKMs were similar for classes A2–F, whereas the median RPKM for the 16 genes in class A1 was moderately higher. The significance of this higher median value is difficult to determine due to the small size of class A1. Little difference was observed between genes containing CpG island promoters and LCG promoters (Figure 5A).

An examination of RPKMs at the nascent transcript peak for each gene (i.e., the RPKM at the time point with the highest nascent transcript level) also revealed similarities between all classes (Figure 5B). However, the median peak RPKMs for LCG genes in classes D, E, and F were ~3-fold higher than the RPKMs for CpG island genes in the same classes.

Most interestingly, the median fold-induction values for LCG genes in classes D and E were substantially higher than for CpG island genes (Figure 5C). Of particular relevance, LCG promoters are highly prevalent among genes induced by more than 100-fold, whereas CpG island promoters are more prevalent among genes that are weakly induced (5- to 10-fold) (Figures 5C and 5D). Similar profiles were observed when cytoplasmic RNA or mRNA were examined (data not shown), demonstrating that these differences in nascent transcripts lead to similar distributions in mature mRNA.

These findings explain why LCG promoters were found to be more prevalent during the secondary response in previous studies (Ramirez-Carrozzi et al., 2009; Hargreaves et al., 2009). The previous studies focused on limited numbers of genes that encode critical regulators of immunity, such as *Il6*, *Nos2*, and *Il12b*. Nascent transcripts for these genes were induced in this study by 700-, 1,300-, and 1,600-fold, respectively (see Figure 5C), placing them in a fold-induction range dominated by genes with LCG promoters. The unusually potent induction of these and other LCG genes may be related to their key roles in regulating innate and adaptive immune responses.

Together, these results reveal a broad range of basal and peak nascent transcript levels, with LCG promoters strongly enriched among genes that are most potently induced. We propose that the absence of H3K4me3 and the absence of associated RNA polymerase II at LCG promoters in unstimulated cells contribute to potent induction by limiting basal transcription from promoters that have evolved to support particularly high levels of transcription following stimulation. CpG island promoters can support either high or low levels of transcription, but they rarely need to support the potent induction observed at many LCG genes.

Importantly, the quantitative RNA-seq results fail to support models that proposed that rapidly induced genes with CpG island promoters are transcribed at high levels in unstimulated cells, with mRNA induction due primarily to stimulus-dependent enhancement of RNA processing. Additional results presented below (Figure 6) provide further support for the view that inducible genes are transcribed at variable basal levels, with increased

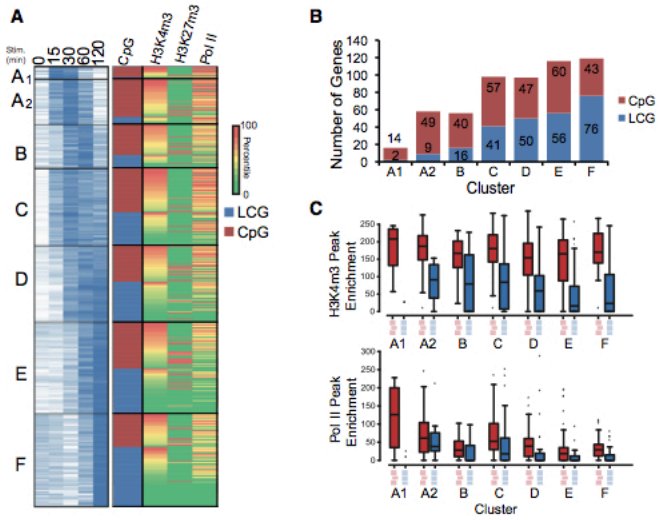


Figure 4. Promoter and Chromatin Properties of Coexpressed Genes

(A) The distribution of CpG island promoters was determined for each class from Figure 3 on the basis of expression kinetics of chromatin-associated transcripts. Levels of histone H3K4me3, H3K27me3, and RNA polymerase II at the promoters in unstimulated cells are shown. ChIP-seq values represent the signal within a 1 kb window centered on the TSS. Each column is normalized to the average enrichment at all inducible promoters and is color-coded according to percentile.

(B) The number of genes that contain CpG island (red) or LCG (blue) promoters within each coexpressed class is shown.

(C) The quantitative distributions of promoter-associated histone H3K4me3 (top) and RNA polymerase II (bottom) in unstimulated cells is shown for each class subdivided by promoter CpG content. The scale of the box plots represents the intensity of ChIP-seq peaks as determined in De Santa et al. (2009).

transcription initiation/promoter release playing a prominent role in the induced expression of all or almost all genes. Our findings are not incompatible with models in which cell stimulation enhances transcription elongation and/or the rate of RNA processing. However, we propose that any enhancement of elongation and processing that occurs following stimulation primarily serves the purpose of promoting the efficient expression of genes that are also strongly upregulated at the level of transcription initiation and/or promoter release.

Class-Specific Differences in Transcriptional Regulation

The high-resolution temporal profiles of nascent transcripts provide a step toward identifying signaling pathways and transcription factors that regulate genes within each coexpression class. To gain insight into factors that may regulate specific classes, a motif enrichment analysis was carried out with promoter sequences (Figures 5E and S6A). NF-κB motifs were overrepresented in most classes, which is consistent with the expectation that this factor contributes to the regulation of many lipid-A-induced genes. In addition, binding sites for factors within the ATF, JUN, and CREB families were strongly enriched in class A1 and A2 promoters. These motifs were more strongly enriched than NF-κB motifs in these classes, whereas NF-κB motifs were strongly overrepresented in class C. Genes in all three classes were potently induced at the 15 min time point, but classes A1 and A2 were transiently induced, whereas class C transcription was sustained (see Figure 3). These results suggest that different sets of factors may coordinate the transient and sustained responses to TLR4 signaling.

The motif analysis also uncovered a high prevalence of STAT and IRF binding sites within the promoters of genes in classes E and F (Figure 5E). Genes containing these motifs may be direct targets of interferon signaling. IFN-β is potently induced by lipid A and can induce interferon (IFN) response genes through the type 1 interferon receptor, IFNAR (Vaidya and Cheng, 2003). To identify IFN-dependent genes, we performed RNA-seq analysis with mRNA from wild-type and *Ifnar*$^{-/-}$ macrophages stimulated with lipid A. Figure S6C highlights genes whose mRNA levels were reduced by 3- to 10-fold (orange) or >10-fold (red) in the *Ifnar*$^{-/-}$ macrophages compared to wild-type. Consistent with the results of the motif analysis, 94% of the genes that exhibit at least 3-fold dependence on IFNAR were located in classes E and F. Also consistent with the motif analysis, IFNAR dependence was biased toward genes with LCG promoters (Figures S6B and S6C). These results are consistent with previous suggestions that, unlike the primary responses to LPS and TNF, the primary response to IFN favors LCG promoters (Ramirez-Carrozzi et al., 2009).

Accumulation of Full-Length, Incompletely Spliced Transcripts in the Chromatin Fraction

As discussed above, the contributions of promoter-proximal pausing, transcription elongation, and transcript processing to the regulation of inducible transcription are of considerable interest. The RNA-seq data from chromatin-associated transcripts allowed an examination of transcript dynamics by analyzing the read density across inducible transcription units over time. An initial expectation was that read density would exhibit a decreasing 5′ to 3′ gradient across actively transcribed genes. This gradient was expected because 5′ end reads will be generated by both partial and complete transcripts, whereas 3′ reads will be generated only by complete or nearly complete transcripts. We also expected to observe read peaks at the 5′ end due to promoter-proximal pausing.

As shown in Figure 1 for the *Nfkb1* gene, a clear 5′ to 3′ gradient was detected at the 15 and 30 min time points in the

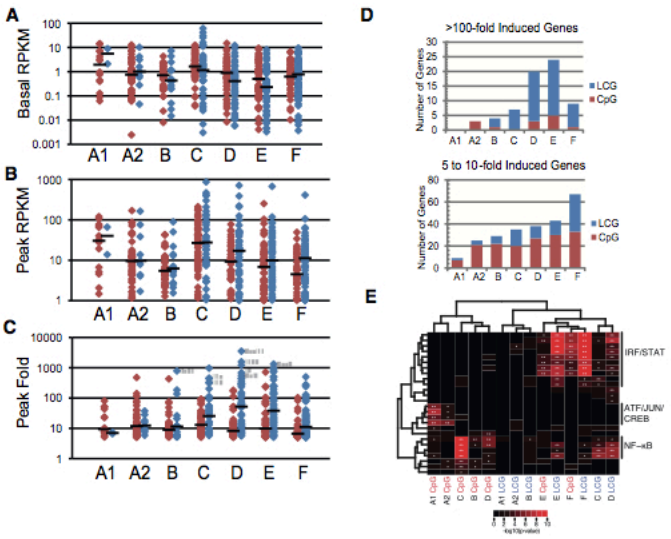


Figure 5. Dynamic Ranges of Expression within Coexpression Classes

(A) Chromatin-associated transcript levels (RPKMs) in unstimulated macrophages are shown for all genes in each class. Genes containing CpG island promoters (red) and LCG promoters (blue) are shown separately. The median transcript level for each group is indicated by a black bar.

(B) Chromatin-associated transcript levels (RPKMs) at their peak time point are shown for all genes in each class, with CpG island and LCG promoters shown separately.

(C) The maximum fold induction of chromatin transcripts for each gene is shown by dividing the peak RPKM by the RPKM in unstimulated cells. Representative genes with LCG promoters that exhibit strong activation are indicated.

(D) The numbers of genes in each class activated by more than 100-fold and by only 5- to 10-fold are shown.

(E) Overrepresented transcription factor binding sites are shown for each class. Data are presented after hierarchical clustering. CpG island and LCG promoters were analyzed separately. Transcription factor families are shown at the right (see Figure S6A for details). Color intensity is proportional to the negative log(p value). Single and double stars indicate $p < 0.01$ and $p < 0.001$, respectively.

See also Figure S6.

chromatin fraction. However, a gradient was not observed in the unstimulated cells, despite substantial basal transcription prior to stimulation (see Figure S1B). The gradient was also absent at the 60 and 120 min time points, after transcription had reached its peak. These results reveal that, during steady-state transcription (prior to induction or after transcription had reached its peak), full-length transcripts accumulated on the chromatin and that release from the chromatin was slow relative to the completion of transcription. Moreover, many of these transcripts remained incompletely spliced (see below).

The features described above for *Nfkb1* were apparent when genes of each class were analyzed together, after normalizing the lengths of all genes (Figure 6A). For example, in class A2, a 5′ to 3′ gradient was observed at the 15 min time point, which is the point at which transcripts were beginning to accumulate. By the 30 min time point, the gradient had disappeared. The nascent transcript levels declined at the 60 and 120 min time points, but without a 5′ to 3′ gradient. The only class that failed to show evidence of a gradient is class A1. This class is notable for its extremely rapid induction and the short length of its genes (1–4 kb; data not shown). Genes of this length would be predicted to be fully transcribed well before the 15 min time point.

These results suggest that, following the completion of transcription, a lag period exists prior to transcript release. Splicing may be completed during this lag period, as introns are abundant on the chromatin but scarce in the nucleoplasm and cytoplasm. Consistent with the hypothesis that completed transcripts accumulate on the chromatin, we found that sequence reads for most genes in the chromatin fraction declined dramatically at the polyadenylation site. This is apparent in Figure 6B, which shows the read distributions across the RefSeq polyadenylation sites for all inducible genes. Similar profiles were observed when polyadenylated mRNA was analyzed, providing additional evidence that the decline observed in the chromatin fraction corresponds to cleavage of completed transcripts at the polyadenylation site (data not shown). RNA-seq analysis of oligo-dT-primed cDNA libraries revealed that polyadenylated transcripts are abundant on the chromatin (data not shown). However, only a fraction of the unspliced transcripts cleaved at the polyadenylation site appeared to be polyadenylated (data not shown), with the precise percentage of polyadenylated transcripts difficult to determine.

The data described above suggest that completed transcripts accumulated at many genes in the chromatin fraction and that these transcripts were usually cleaved at the polyadenylation site but incompletely spliced. Additional insight into transcript dynamics was obtained from an examination of transiently induced genes after transcription had subsided. As shown for the *Il12b* gene in Figure 6C, intronic reads were lost more rapidly than exonic reads from the chromatin fraction after transcription subsided. At the 60 min time point, abundant *Il12b* transcripts were present on the chromatin (Figure 6C) and also in the nucleoplasm and cytoplasm (data not shown); in the chromatin fraction, reads were abundant in both the exons and introns. *Il12b* transcription peaked between 60 and 120 min and then subsided by the 120 min time point. At this latter time point, exon reads were far more abundant than intron reads (Figure 6C), suggesting that excised introns were released from the chromatin and/or degraded more rapidly than spliced transcripts were released. These results demonstrate that excised introns are not stably maintained on the chromatin and suggest that the incompletely spliced transcripts observed at *Il12b* and many other genes are precursors of fully spliced transcripts (see Discussion).

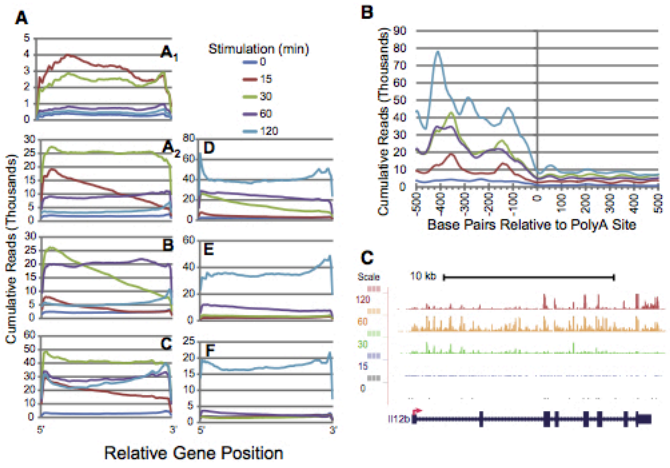


Figure 6. High Prevalence of Completed Transcripts Cleaved at the Polyadenylation Site in the Chromatin Samples

(A) Normalized read distributions from the 5′ to 3′ end of all genes within each class are shown for each time point. Data from chromatin-associated transcripts were used, and genes were normalized for length. Relative gene position is shown on the x axis, and cumulative reads at each relative position are shown on the y axis.

(B) Cumulative read distributions spanning the polyadenylation site are shown for lipid-A-induced genes. The x axis shows the gene position relative to the polyadenylation cleavage site, and the y axis shows cumulative reads at each position.

(C) Read profiles are shown for chromatin-associated transcripts at the *Il12b* locus. Time points and scale are shown at the left, and exon-intron structure of the locus is at the bottom. In the nucleoplasmic and cytoplasmic fractions, abundant exonic reads accumulated by the 60 min time point (not shown).

To examine splicing efficiency of chromatin-associated transcripts in greater depth, we determined the ratio of exon reads to total reads for genes within each constitutive and inducible class from Figures 2A and 3A. Using constitutively expressed cluster 11 as an example (Figure 7A, blue), we found that, in the chromatin fraction, most genes were incompletely spliced. In Figure 7A, the genes are ordered on the basis of splicing level. Genes at the far left correspond to those whose chromatin-associated transcripts are efficiently spliced, as they primarily have exon reads in the chromatin fraction. Genes at the far right correspond to those with greater numbers of reads in introns than exons; examining several of these genes revealed that they contain aberrant intron read spikes that are likely to be derived from additional transcripts from the locus (data not shown). Between these extremes, the large majority of genes were comparable to *Nfkb1* in that only a moderate fraction of introns had been excised from the pool of transcripts. In contrast to the incomplete splicing observed in the chromatin fraction, efficient splicing was observed at almost all genes in the nucleoplasm and cytoplasm (Figure 7A). Similar results were obtained with each constitutive and inducible class (data not shown).

As mentioned above (Figure 6B), the decline in reads at the polyadenylation site suggests that most transcripts in the chromatin fraction are full-length and cleaved at the polyadenylation site despite the incomplete splicing. Additional evidence that incompletely spliced transcripts are frequently cleaved at the polyadenylation site is provided by published RNA-seq data obtained by using the metabolic labeling (4sU-Seq) and GRO-seq methods (Rabani et al., 2011; Escoubet-Lozach et al., 2011). Figure 7B shows the read profiles obtained with these methods at *Nfkb1* in dendritic cells (4sU-Seq) or macrophages (GRO-seq) stimulated with LPS for 1 hr. The 4sU profile was similar to the profile observed with chromatin-associated transcripts; a modest 5′ to 3′ gradient was observed, which was similar to the gradient observed in Figure 1A at the 15 and 30 min time points with the chromatin transcripts. The 4sU profile also shows a dramatic decline in reads at the polyadenylation site, despite abundant intronic reads throughout the gene. This decrease is most apparent in the fifth track of Figure 7B, in which the 15 read scale revealed greatly reduced read numbers after the polyadenylation site in comparison to the final intron. Thus, both the 4sU method and the analysis of chromatin-associated transcripts revealed a high abundance of incompletely spliced transcripts that had been cleaved at the polyadenylation site.

In contrast to the read profiles obtained with the 4sU-Seq and chromatin-associated transcript methods, abundant reads extending past the polyadenylation site were observed by GRO-seq (Figure 7B). This finding is consistent with the knowledge that polymerase molecules continue transcription past the polyadenylation site and that these transcripts can be captured in the short labeling time of the GRO-seq experiment.

Finally, although reads in our chromatin fraction declined at the polyadenylation site on most genes, exceptions were apparent. One example is the *Mapkapk2* gene, for which reads can be detected more than 20 kb downstream of the gene 30 min after induction (Figure S7). Reads from apparent *Mapkapk2* transcripts decline gradually with distance from the gene and extend into the neighboring *Il10* gene; *Il10* is transcribed on the opposite strand and is also induced by many stimuli (Figure S7). It is difficult to predict at this time whether the extensive read-through transcription observed at this gene has functional relevance or simply reflects the fortuitous absence of transcription termination sequences.

DISCUSSION

We describe the use of RNA-seq to examine chromatin-associated, nucleoplasmic, and cytoplasmic transcripts in macrophages exposed to an inflammatory stimulus. This approach allowed us to obtain insights into transcription and RNA dynamics and to examine the inflammatory gene transcription program at high resolution. Although other methods have been developed

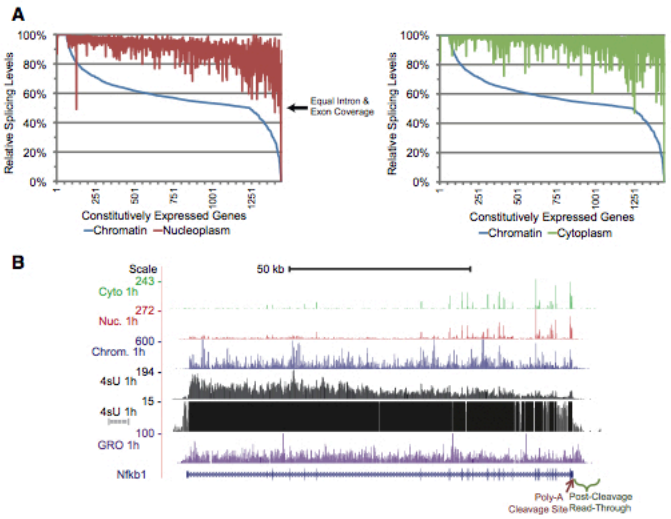


Figure 7. Analysis of Exon:Intron Ratios and a Comparison of Nascent Transcript Methods

(A) Reads mapping to ~1,450 genes from cluster 11 (Figure 2) were analyzed for relative exon and intron coverage. The genes were rank ordered according to splicing levels (length-normalized exon reads over total reads) in the chromatin fraction (blue) and were compared to splicing levels in the nucleoplasm (red) and cytoplasm (green). Similar results were obtained with all gene clusters (data not shown).

(B) RNA-seq read distributions at *Nfkb1* are shown comparing reads from the cytoplasmic (green), nucleoplasmic (red), and chromatin (blue) fractions to data obtained by 4sU-Seq and GRO-seq (Rabani et al., 2011; Escoubet-Lozach et al., 2011). GRO-seq was performed with LPS-stimulated bone-marrow-derived macrophages. 4sU-Seq was performed with LPS-stimulated bone-marrow-derived dendritic cells. The fifth track displays the 4sU-Seq data with a scale of 15 reads to show the reduced number of reads after the polyadenylation site relative to the last intron. The cleavage site (red arrow) and reads representing transcription past this site (green bracket; abundant only in the GRO-seq data set) are indicated. See also Figure S7.

for genome-wide analysis of nascent transcripts, the experimental strategy described here is unique in its ability to provide information about the properties of transcripts that remain associated with chromatin and the changes that occur as the transcripts proceed to the nucleoplasm/ER or cytoplasm.

At most protein-coding genes, full-length transcripts appear to accumulate on the chromatin during steady-state transcription, and these transcripts are often cleaved at the polyadenylation site but are incompletely spliced. The retention of polyadenylated yet incompletely spliced transcripts at a gene locus has been observed previously in an analysis of integrated DNA constructs (Brody et al., 2011). Our results demonstrate that such transcripts accumulate at most endogenous genes in mammalian cells. Fully spliced transcripts were predominant within the nucleoplasm and cytoplasm, suggesting that the temporal delay in transcript release could allow the completion of splicing. Such a delay may be indicative of a quality control step ensuring full processing prior to release (Schmid and Jensen, 2010). It is noteworthy that the splicing delay is not indicative of when spliceosome assembly occurs. Spliceosome assembly may be linked to active transcription, with elongating transcripts assembling factors at exon-intron junctions as the RNA polymerase proceeds along the transcription unit (Perales and Bentley, 2009).

Previous studies have provided strong evidence of cotranscriptional splicing (Singh and Padgett, 2009; Perales and Bentley, 2009). Other studies have indicated that splicing can follow the completion of transcription (Nevins and Darnell, 1978; Vargas et al., 2011). Our results demonstrate that, in mammalian cells, the completion of splicing often follows the completion of transcription, although some genes accumulate chromatin-associated transcripts that have been efficiently spliced, which is consistent with cotranscriptional splicing. One recent study analyzed chromatin-associated transcripts in *Drosophila* at steady state and reported evidence of more extensive cotranscriptional splicing (Khodor et al., 2011). It is possible that cotranscriptional splicing is more prevalent in some species than others. However, differences in the experimental procedures and analysis methods may also have influenced the interpretation of the results. In particular, our finding that full-length transcripts frequently accumulate on the chromatin was strongly dependent on the analysis of inducible transcription; the dynamic transition from a read distribution lacking a 5′ to 3′ gradient to one exhibiting a gradient during transcriptional induction to one in which the gradient is again lost (Figure 6A) provided clear evidence of the accumulation of full-length transcripts. Furthermore, Khodor et al. (2011) depleted polyadenylated transcripts from their chromatin fraction prior to RNA-seq analysis, which may have obscured the abundance of full-length transcripts that are incompletely spliced.

One important question that must be considered when interpreting our RNA-seq data is whether the abundant full-length, incompletely spliced transcripts that accumulate on the chromatin are precursors to productive fully-spliced mRNA or whether they might instead by nonproductive "dead-end" transcripts. The rapid increase and decrease in chromatin transcripts at transiently induced genes (classes A1, A2, and B in Figure 3) argue against the existence of a large pool of nonproductive chromatin-associated transcripts. The hypothetical nonproductive transcripts would need to be eliminated from the chromatin at the same rate as the release of productive transcripts. The fact that the increase in chromatin transcripts is followed by an increase in nucleoplasmic and cytoplasmic transcripts, with a defined temporal delay, further supports the

notion that the chromatin transcripts are precursors to mRNA. In fact, because chromatin transcripts are not maintained on transiently induced genes for a prolonged period, the hypothetical full-length nonproductive transcripts would need to be far more abundant than the pool of nascent productive transcripts to obscure the 5′ to 3′ read gradient generated by these nascent transcripts. Finally, the results in Figure 6C suggest that the full-length incompletely spliced transcripts are competent for the completion of splicing, as the incompletely spliced transcripts resolve into transcripts containing only exons after transcription has subsided.

It was also necessary to consider the possibility that the broad read coverage of both exons and introns on the chromatin may be due to the stable maintenance of introns that were excised immediately after transcription. However, our data suggest that excised introns are short-lived on the chromatin. For example, at the *Fos* gene, intron reads peaked at the 15 min time point and decreased by 8-fold at the 30 min time point (data not shown), suggesting that introns were released from the chromatin and/or degraded with a half-life of ~5 min. *Fos* exon reads decreased by only 3-fold during these 15 min, indicating that introns were lost from the chromatin more rapidly than exons. Introns were also lost more rapidly than the fully spliced RNA at the *Il12b* gene (Figure 6C) and at many other genes (data not shown) after transcription had declined, possibly because the spliced transcripts were not released until all introns had been excised. A detailed analysis of the kinetics of intron and exon accumulation and loss relative to the kinetics of splice exon-exon junction accumulation provides additional evidence of a splicing delay following transcription through an intron (A.P.-J. and D.L.B., unpublished data).

Our results and the results of Rabani et al. (2011) highlight the importance of transcriptional induction kinetics in shaping the temporal mRNA profiles of lipid-A-induced genes. Although genes could not be identified that are induced primarily via enhanced mRNA stability, mRNA stability must act in concert with transcriptional control to shape the temporal expression profile of each gene. Because the transient induction of nascent transcripts (i.e., transcription) for genes in classes A and B is mirrored by the transient induction of the mRNA, these transcripts must either be intrinsically unstable, or their stability must be tightly regulated (Caput et al., 1986; Hao and Baltimore, 2009). Conversely, mRNAs from genes that exhibit sustained induction are likely to be intrinsically stable or may be stabilized in stimulated cells. Thus, our results support the hypothesis that the temporal dynamics of gene expression required the coevolution of mechanisms regulating transcription and mRNA stability.

In previous studies, we defined chromatin and promoter properties of 67 genes by quantitative RT-PCR (qRT-PCR) (Ramirez-Carrozzi et al., 2009). By RNA-seq, we can now discern 560 genes whose nascent transcripts are activated by lipid A at least 5-fold within 2 hr of stimulation. By classifying genes on the basis of nascent transcript kinetics, we found that binding sites for transcription factors in the ATF, JUN, and CREB families are abundant in the most rapidly and transiently induced genes and that IFN-dependent genes are tightly clustered in two gene classes whose nascent transcripts are induced at later times. The framework provided by classifying genes on the basis of their nascent transcript profiles will greatly aid future efforts to understand how a broad range of signaling pathways, transcription factors, and chromatin events shape the inflammatory gene transcription program.

Finally, the current analysis provides genome-scale confirmation of previous models of inflammatory gene activation, while suggesting revisions of other models. The results confirm that a high percentage of genes that are rapidly induced by a TLR stimulus contain CpG island promoters. In unstimulated cells, these promoters exhibit chromatin features of active genes and can be activated in a nucleosome-remodeling-independent manner (Ramirez-Carrozzi et al., 2009). The results also support the previous finding that IFN-induced genes are biased toward LCG promoters, which are not assembled into constitutively active chromatin, consistent with the view that IFN-induced transcription factors promote nucleosome remodeling (Huang et al., 2002; Liu et al., 2002).

However, in contrast to previous suggestions, our results reveal that both CpG island and LCG promoters are abundant among genes induced at late times. One important distinction is that LCG promoters are highly enriched among the most potently induced genes, and CpG island promoters are more prevalent among weakly induced genes. The CpG island promoters associated with late genes possess features of active chromatin in unstimulated cells and are thus similar to CpG island promoters associated with early genes. We propose that late genes containing CpG island promoters are poised for induction just like the early genes but that transcriptional induction of the late genes requires transcription factors or signaling pathways that are expressed following the primary response to the stimulus. These genes may tolerate constitutively active chromatin because the required dynamic range of expression between the unstimulated and stimulated states is relatively small. In contrast, genes that require a large dynamic range of expression often contain LCG promoters. An LCG promoter may help limit basal transcription and, at the same time, may facilitate tight regulation by conferring a requirement for an inducible nucleosome remodeling event. The ability to perform high-resolution genome-scale analysis of nascent transcripts should facilitate future efforts to further dissect and understand inflammatory gene regulation networks.

EXPERIMENTAL PROCEDURES

Cell Culture, Cell Fractionation, and Library Preparation

Macrophages were prepared from the bone marrow of 4-week-old C57BL/6 or *Ifnar*$^{-/-}$ mice (Ramirez-Carrozzi et al., 2009). Adherent macrophages were activated on day 6 with lipid A (100 ng/ml) (Invivogen).

Subcellular fractions were prepared as described (Pandya-Jones and Black, 2009), with minor changes. The cell lysis buffer contained 0.15% NP-40, and the sucrose cushion did not contain detergent. Fraction purity was confirmed by immunoblot analysis of SNRP70, β-tubulin (Sigma), and histone H3 (Abcam). Cytoplasmic and nucleoplasmic RNA was purified by using QIAGEN RNeasy columns. Chromatin RNA was isolated by using TRI-reagent (MRC), followed by further purification with RNeasy columns. All samples were eluted into 100 ul RNase-free water. RNA (10 ug) from each fraction was depleted of ribosomal RNA (rRNA) by using the Mouse/Human Ribominus Kit (Invitrogen). Whole-cell RNA was purified using TRI-reagent and RNeasy columns. Polyadenylated RNA was purified by using the MicroPoly(A) Purist Kit (Ambion).

Strand-specific libraries were generated by using 500 ng RNA input according to the "deoxyuridine triphosphate (dUTP)" method (Levin et al., 2010). Libraries from mRNA were not strand specific. An Illumina HiSeq 2000 was used for sequencing with a single-end-sequencing length of 50 nt.

Sequence Mapping and Analysis

All bioinformatics analyses were conducted by using the Galaxy platform except as noted (Goecks et al., 2010). Reads were aligned to the mouse mm9 reference genome with Tophat (Trapnell et al., 2010) by using most default parameters. Alignments were restricted to uniquely mapping reads, with two possible mismatches permitted. RPKM values were calculated as described (Mortazavi et al., 2008) for mm9 RefSeq genes (Pruitt et al., 2005) by using Seqmonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Because chromatin transcripts were largely unspliced, RPKMs were calculated by counting all locus mapping reads and dividing by the length of the entire locus. RPKMs for nucleoplasmic and cytoplasmic transcripts and mRNAs were calculated by counting exonic reads and dividing by mRNA length. Coexpressed gene classes were generated with Cluster3 by applying k-means clustering to mean-centered $\log_2$(RPKM) expression values. Analysis of 5′ to 3′ read distribution trends was conducted by using the Seqmonk probe trends tool. Splicing levels were analyzed by comparing the base coverage of exons alone divided by the base coverage of an entire locus; the calculations included exonic and intronic reads as well as exon:exon and exon:intron junction-spanning reads. This ratio was subsequently compared to three data points by using the PERCENTRANK function in Excel: 0, X, and 1, where X is the ratio of exon length to locus length. Genome tracks were generated by using the genome coverage utility in the BEDTools suite (Quinlan and Hall, 2010) and visualized in the UCSC genome browser.

ChIP-Seq and Motif Analyses

ChIP-seq data sets were obtained from GEO accession GSE17631 (De Santa et al., 2009). Data sets were converted from mm8 to mm9 genome builds by using the liftover utility in Galaxy. Promoter enrichment was calculated as the height of peaks falling within a 1 kb window centered on the TSS.

Pscan was used to detect DNA motifs overrepresented in each class between nucleotides −500 and +250 relative to the TSS (Zambelli et al., 2009). Significance was tested against CpG-content-matched promoters as background. A binding site was considered significantly overrepresented with a p value <0.01.

ACCESSION NUMBERS

Sequencing data have been submitted to GEO under accessions GSE32916 and GSE38892.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.05.043.

ACKNOWLEDGMENTS

We thank Christopher Glass for sharing results prior to their publication; Alexander Hoffmann, Steve Horvath, Justin Langerman, Steve Ley, Chia-Ho Lin, and Matteo Pellegrini for helpful discussions; and Owen Witte for invaluable support. This work was supported by a UCLA Dissertation-Year Fellowship (to A.P.-J.), by NIH grants R01GM086372 and R01CA127279 (to S.T.S.), and by the Broad Stem Cell Research Center at UCLA. D.L.B. is an Investigator of the Howard Hughes Medical Institute.

Received: November 21, 2011
Revised: March 27, 2012
Accepted: May 6, 2012
Published: July 19, 2012

REFERENCES

Adelman, K., Kennedy, M.A., Nechaev, S., Gilchrist, D.A., Muse, G.W., Chinenov, Y., and Rogatsky, I. (2009). Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. Proc. Natl. Acad. Sci. USA *106*, 18207–18212.

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.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS Biol. *9*, e1000573.

Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S., and Cerami, A. (1986). Identification of a common nucleotide sequence in the 3′-untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl. Acad. Sci. USA *83*, 1670–1674.

Chen, Q., Jagannathan, S., Reid, D.W., Zheng, T., and Nicchitta, C.V. (2011). Hierarchical regulation of mRNA partitioning between the cytoplasm and the endoplasmic reticulum of mammalian cells. Mol. Biol. Cell *22*, 2646–2658.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

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.

De Santa, F., Narang, V., Yap, Z.H., Tusi, B.K., Burgold, T., Austenaa, L., Bucci, G., Caganova, M., Notarbartolo, S., Casola, S., et al. (2009). Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. EMBO J. *28*, 3341–3352.

Dye, M.J., Gromak, N., and Proudfoot, N.J. (2006). Exon tethering in transcription by RNA polymerase II. Mol. Cell *21*, 849–859.

Escoubet-Lozach, L., Benner, C., Kaikkonen, M.U., Lozach, J., Heinz, S., Spann, N.J., Crotti, A., Stender, J., Ghisletti, S., Reichart, D., et al. (2011). Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. PLoS Genet. *7*, e1002401.

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

Friedel, C.C., and Dölken, L. (2009). Metabolic tagging and purification of nascent RNA: implications for transcriptomics. Mol. Biosyst. *5*, 1271–1278.

Goecks, J., Nekrutenko, A., and Taylor, J.; Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Hah, N., Danko, C.G., Core, L.J., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622–634.

Hao, S., and Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. Nat. Immunol. *10*, 281–288.

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

Huang, M., Qian, F., Hu, Y., Ang, C., Li, Z., and Wen, Z. (2002). Chromatin-remodelling factor BRG1 selectively activates a subset of interferon-alpha-inducible genes. Nat. Cell Biol. *4*, 774–781.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., II, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. *25*, 2502–2512.

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.

Litvak, V., Ramsey, S.A., Rust, A.G., Zak, D.E., Kennedy, K.A., Lampano, A.E., Nykter, M., Shmulevich, I., and Aderem, A. (2009). Function of C/EBPdelta in

a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. Nat. Immunol. *10*, 437–443.

Liu, H., Kang, H., Liu, R., Chen, X., and Zhao, K. (2002). Maximal induction of a subset of interferon target genes requires the chromatin-remodeling activity of the BAF complex. Mol. Cell. Biol. *22*, 6471–6479.

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.

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

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: regulating the transition from transcription initiation into productive elongation. Biochim. Biophys. Acta *1809*, 34–45.

Nevins, J.R., and Darnell, J.E., Jr. (1978). Steps in the processing of Ad2 mRNA: poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell *15*, 1477–1493.

Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896–1908.

Pawlicki, J.M., and Steitz, J.A. (2008). Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. J. Cell Biol. *182*, 61–76.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol. Cell *36*, 178–191.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *33*(Database issue), D501–D504.

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

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., 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-2beta nucleosome remodeling complexes during an inflammatory response. Genes Dev. *20*, 282–296.

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.

Schmid, M., and Jensen, T.H. (2010). Nuclear quality control of RNA polymerase II transcripts. Wiley Interdiscip. Rev. RNA *1*, 474–485.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nat. Struct. Mol. Biol. *16*, 1128–1133.

Smale, S.T. (2010). Selective transcription in response to an inflammatory stimulus. Cell *140*, 833–844.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Vaidya, S.A., and Cheng, G. (2003). Toll-like receptors and innate antiviral responses. Curr. Opin. Immunol. *15*, 402–407.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A., Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell *147*, 1054–1065.

Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol. Cell. Biol. *14*, 7219–7225.

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*(Web Server issue), W247–W52.

APPENDIX B

Splicing Kinetics and Transcript Release from the Chromatin Compartment Limit the Rate of Lipid A-Induced Gene Expression

Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression

AMY PANDYA-JONES,[1,2,4] DEV M. BHATT,[1,2,4] CHIA-HO LIN,[1] ANN-JAY TONG,[1,2] STEPHEN T. SMALE,[1,2,4,5] and DOUGLAS L. BLACK[1,2,3,4,5]

[1]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California 90025, USA
[2]Molecular Biology Institute, University of California, Los Angeles, California 90025, USA
[3]Howard Hughes Medical Institute, University of California, Los Angeles, California 90025, USA

ABSTRACT

The expression of eukaryotic mRNAs is achieved though an intricate series of molecular processes that provide many steps for regulating the production of a final gene product. However, the relationships between individual steps in mRNA biosynthesis and the rates at which they occur are poorly understood. By applying RNA-seq to chromatin-associated and soluble nucleoplasmic fractions of RNA from Lipid A-stimulated macrophages, we examined the timing of exon ligation and transcript release from chromatin relative to the induction of transcription. We find that for a subset of genes in the Lipid A response, the ligation of certain exon pairs is delayed relative to the synthesis of the complete transcript. In contrast, 3′ end cleavage and polyadenylation occur rapidly once transcription extends through the cleavage site. Our data indicate that these transcripts with delayed splicing are not released from the chromatin fraction until all the introns have been excised. These unusual kinetics result in a chromatin-associated pool of completely transcribed and 3′-processed transcripts that are not yet fully spliced. We also find that long introns containing repressed exons that will be excluded from the final mRNA are excised particularly slowly relative to other introns in a transcript. These results indicate that the kinetics of splicing and transcript release contribute to the timing of expression for multiple genes of the inflammatory response.

Keywords: gene expression; pre-mRNA splicing; splicing kinetics; mRNA maturation; inflammatory response

INTRODUCTION

Eukaryotic mRNAs mature via a complex series of molecular reactions prior to their export from the nucleus for translation in the cytoplasm. RNA synthesis commences upon the initiation of transcription by RNA polymerase II (Pol II) from a promoter, followed by elongation through the template DNA, and termination often many kilobases downstream (Selth et al. 2010). Early in transcription, a 7-methyl guanosine cap is added to the 5′ end of precursor messenger RNAs (pre-mRNAs) (Cho et al. 1997; McCracken et al. 1997; Shuman 2001). Most nascent Pol II transcripts also contain introns that must be excised during mRNA maturation (Matlin and Moore 2007; Pandya-Jones and Black 2009; Perales and Bentley 2009; Wahl et al. 2009). It is thought that initial spliceosome assembly on the pre-mRNA occurs during synthesis, and some introns on long transcripts are removed prior to complete transcription (Osheim et al. 1985; LeMaire and Thummel 1990; Bauren and Wieslander 1994; Tennyson et al. 1995; Wetterberg et al. 1996; Lacadie et al. 2006; Listerman et al. 2006). Once transcription through the gene is complete, the pre-mRNA must be cleaved and polyadenylated to generate its mature 3′ end (Chan et al. 2011). Full-length transcripts are released from the DNA template to allow export to the cytoplasm (Stewart 2010). These multiple reactions provide many potential steps at which to regulate gene expression, and some of these steps are coupled to each other, either kinetically or mechanistically (Hocine et al. 2010; Hsin and Manley 2012). However, the mechanisms regulating individual steps and how they each contribute to the output of the final product are not well understood.

Our incomplete picture of the gene expression pathway is in part due to limited information on the reaction kinetics of many of the steps. The rate of transcription elongation has been estimated using a variety of approaches (Ardehali and Lis 2009). However, endogenous rates of RNA processing reactions in mammalian systems have proven more difficult to measure, and how they affect the time required for a pre-

[4]These authors contributed equally to this work.
[5]**Corresponding authors**
E-mail dougb@microbio.ucla.edu
E-mail smale@mednet.ucla.edu
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.039081.113.

mRNA to be matured for export is not known (Grunwald and Singer 2010; Schmid and Jensen 2010; Rabani et al. 2011). The kinetics of excision for individual introns have been described, but larger scale analyses are limited (Kessler et al. 1993; Bauren and Wieslander 1994; Tennyson et al. 1995; Singh and Padgett 2009; Hicks et al. 2010). Nevertheless, understanding the sequence of events in mRNA production will be essential before we can accurately describe their mechanisms of regulation.

We recently reported the application of RNA sequencing (RNA-seq) to understanding the Lipid A-induced transcriptional response in mouse bone marrow-derived macrophages (Bhatt et al. 2012). This analysis of RNA from the chromatin, soluble nucleoplasmic, and cytoplasmic compartments over time after Lipid A stimulation enabled us to characterize the pro-inflammatory gene expression program at a new level of detail. Notably, it was possible to follow the transcription kinetics of individual genes from initiation to completion on chromatin, the release of transcripts into the nucleoplasm, followed by appearance of mRNAs in the cytoplasm. The varying kinetics by which transcripts proceeded through these steps permitted a quantitative and comprehensive categorization of different components of the Lipid A response. An interesting feature of these data was the apparent abundance of fully transcribed and only partially spliced RNAs that remained associated with chromatin, beyond the time required for transcription to complete. This implied that transcript release from this compartment might be a slower step than was previously thought and presented the question of what was necessary for release to occur.

Using this Lipid A induction system, we sought to examine the timing of the splicing reaction and transcript release for genes of the inflammatory response. We find that particular introns in this regulatory cascade are often excised with markedly slower kinetics than has been measured in constitutive transcripts and that incompletely processed RNAs are held within the chromatin fraction until fully spliced. These results have important implications for understanding mechanisms of splicing regulation and of quality control during mRNA biogenesis. Most notably, our data indicate that transcript release from chromatin plays a substantial role in the timing of inflammatory gene expression.

RESULTS

Assessing splicing of Lipid A-induced transcripts

In a recent study, we applied RNA-seq to chromatin-associated, nucleoplasmic, and cytoplasmic RNA isolates from macrophages (Bhatt et al. 2012). Using this RNA sampling, we examined gene responses at 15, 30, 60, or 120 min post-Lipid A treatment (Fig. 1A; Materials and Methods). This work defined groups of Lipid A-responsive genes with different induction kinetics that we tracked from transcription initiation and elongation, through the movement of completed messages from the chromatin to the nucleoplasm and into the cytoplasm. We noted that many introns were present within the chromatin compartment at times when transcription through the gene appeared complete. Using these data sets, we wanted to closely examine the timing of intron removal relative to the initiation of transcription, the completion of transcript synthesis, and the release of the message from chromatin. Our goal was to understand how these different steps contribute to the overall timing of inflammatory gene induction.

In previous studies of fractionated cell nuclei, nascent pre-mRNAs co-purified with the salt, detergent, and urea-resistant pellet while the soluble nuclear extract largely contained complete, spliced messages, as well as many components of nuclear speckles (Wuarin and Schibler 1994; Pandya-Jones and Black 2009; Khodor et al. 2011; Bhatt et al. 2012). The nascent RNAs are released from the pellet by DNase treatment, indicating their association requires intact chromatin (A Pandya-Jones and DL Black, unpubl.). In primary mouse macrophages we found a similar distribution of steady-state RNA populations. For constitutive genes, whose expression was unaffected by Lipid A treatment, nucleoplasmic reads were highly enriched for exonic regions, as would be expected of fully spliced RNA (Supplemental Fig. S1A,B, red). For most of these genes, intron reads in the nucleoplasm were at the level of background contamination from the chromatin, which was estimated from the amount of the known chromatin-associated RNA *Xist* found in the nucleoplasm (see Materials and Methods; Supplemental Fig. S1C, red). However, this pattern was variable and certain constitutive genes produced higher numbers of nucleoplasmic reads mapping to intronic sequences (Supplemental Fig. S1D, red). It is not clear whether introns in this group of constitutive transcripts are spliced away from the chromatin or are more stable after their excision.

In the chromatin fraction patterns of read accumulation on constitutive genes were more variable. For some genes at steady state, decreasing gradients of read density were observed across the locus, presumably derived from the 5′-to-3′ polarity of transcription, but this was not a general pattern (Supplemental Fig. S1A, green; Bhatt et al. 2012). We also observed some constitutive genes exhibiting higher levels of reads mapping to exons than to adjacent introns (Supplemental Fig. S1B, green), and some genes exhibited almost entirely exon reads even on chromatin (Supplemental Fig. S1E, green). If introns are excised soon after their 3′ exon is synthesized, then exons should be enriched relative to their neighboring intron sequences, particularly toward the 5′ end of the gene. However, again this was not a consistent pattern and many genes exhibited reads that were evenly distributed across the locus regardless of exon or intron location (Supplemental Fig. S1A,B, green and data not shown; see also Ameur et al. 2011; Khodor et al. 2011, 2012). Given the variability in the steady-state read accumulation patterns of constitutive genes, it is difficult to draw conclusions about their rates of RNA synthesis and processing. Thus, we used

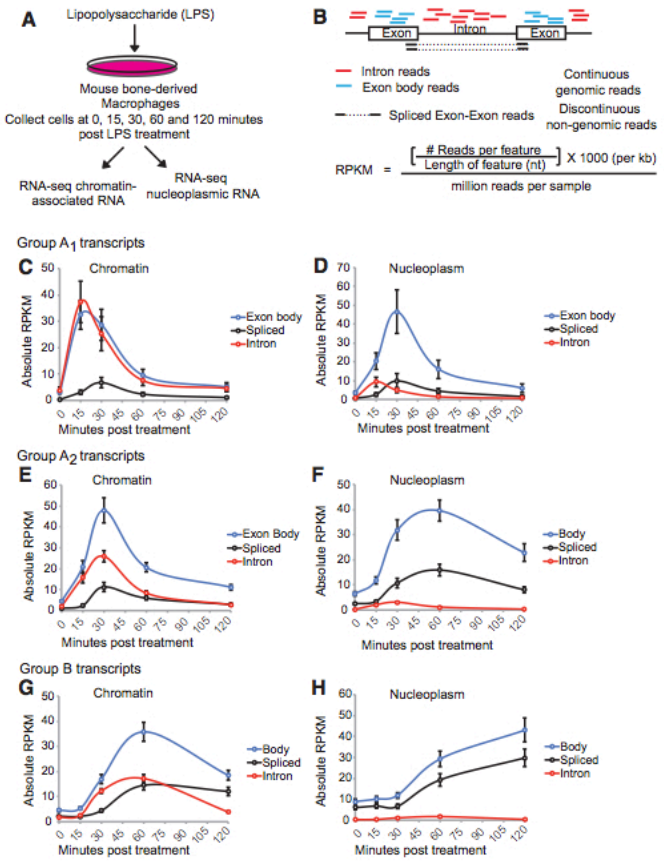


FIGURE 1. RNA synthesis and splicing can be resolved over time on Lipid A-responsive genes. (*A*) In vitro differentiated mouse macrophages were treated with Lipid A and collected at 0, 15, 30, 60, or 120 min post-simulation. Chromatin-associated and nucleoplasmic RNA fractions from the same biological sample, for each time point, were analyzed by RNA-seq. (*B*) Reads mapping to genic regions were classified by their location. Continuously mapping genomic reads include the following: exon body reads (blue) that map entirely within exons; intron reads (red) that map entirely within introns; spliced reads (black) mapping discontinuously to the genome represent reads that cross a spliced exon–exon junction. (*C*) Line graph representing the total RPKM of exon body (blue), intron (red), and spliced reads (black) averaged from all chromatin-associated Group A_1 transcripts at each time point post-Lipid A treatment. (*D*) Same as *C* except that the exon, intron, and spliced RPKMs were calculated for all nucleoplasmic Group A_1 transcripts. (*E*) Same as *C* except that RPKMs were calculated for all chromatin-associated Group A_2 transcripts. (*F*) Same as *D* except that RPKMs were calculated for all nucleoplasmic Group A_2 transcripts. (*G*) Same as *C* except that RPKMs were calculated for all chromatin-associated Group B transcripts. (*H*) Same as *D* except that RPKMs were calculated for all nucleoplasmic Group B transcripts. In all figures, error bars represent the standard error of the mean.

the Lipid A-induced genes to examine the rates of intron excision relative to initiation of transcription and release of the completed transcripts from the template.

To identify genes activated by Lipid A treatment, an averaged Reads Per Kilobase of gene per Million mapped reads (RPKM) value was assigned to each gene, at each time point, for all isoforms within the RefSeq database. This included all exon body, intron body, intron–exon junction, and spliced exon–exon reads (Fig. 1B and below). The fold induction was calculated by comparing the isoform with the highest gene RPKM value over the time course relative to the zero time point. We limited our analysis to those genes up-regulated fivefold or more. This included a majority of the previously identified Lipid A-responsive genes, but excluded genes with substantial constitutive activity that show a small increase in expression upon stimulation (Ramirez-Carrozzi et al. 2009).

Exons constitute only a small portion of the precursor mRNA sequence. To ensure that a majority of exons within the up-regulated transcripts had sufficient coverage for splicing analysis, only genes with an RPKM of ≥60 at one or more time points were selected. At this RPKM threshold, the fraction of exons exhibiting coverage across all nucleotides approached 100% at peak transcription (data not shown). To assess splicing relative to transcription, we determined RPKM values for four features, at each time point in each fraction (Fig. 1B). The exon body RPKM included only those reads that map entirely within an exon. Since a majority of exons within a precursor are included in the mature message, the exon body RPKM was used as a measure of overall RNA volume in subsequent analyses. The intronic RPKM was calculated from reads that mapped entirely within an intron. We measured splicing by quantifying the reads that map discontinuously with the genome and cross an annotated exon–exon junction (Fig. 1). Note that the transcripts studied here are minimally expressed in macrophages prior to Lipid A exposure, allowing the accumulation of spliced exon reads to be measured over time. This is in contrast to the measurement of steady-state RNA, where ratios of spliced to unspliced introns are inferred from the ratio of reads mapping across a 3′ intron–exon junction to reads within the adjacent exon (Khodor et al. 2011, 2012). Since individual sequences along a transcript generate reads with

varying efficiencies, we found that this ratio is highly variable when applied to individual introns. By examining genes induced by Lipid A, we could directly measure the rates of splicing from the accumulation of individual exon–exon junction reads over time, and avoid comparing read levels between different short segments of a transcript. We compared the increase in spliced RNA-seq reads with the gain and loss of intron reads over the same time course. Levels of selected exon, intron and spliced exon–exon sequences were also measured by qPCR and RT/PCR and found to match the temporal profiles determined by RNA-seq (data not shown). We also determined RPKM values for reads spanning 3′ intron–exon boundaries and found that these reads show similar temporal profiles to those mapping within the intron body (Supplemental Fig. S2).

Splicing of some Lipid A-induced transcripts is delayed relative to transcription

Genes that made both cutoffs (equal to or greater than fivefold induction and RPKM ≥60) were grouped by their kinetics of induction, as described previously (Fig. 1C–H; Bhatt et al. 2012). The 11 Group A_1 genes reached peak RNA volume on the chromatin within 15 min (Figs. 1C, blue; 2A, green; Supplemental Fig. S3A,E, green). This group included the immediate early genes, such as *c-Fos* and *Nfkbid*. Group A_2 included 34 genes whose RNA accumulation on chromatin peaked at 30 min (Figs. 1E, blue; 3A, green). An additional 23 genes that reached maximal transcriptional volume at 60 min were classed as Group B genes (Figs. 1G, blue; 4A; Supplemental Fig. S4A, green). We focused our splicing analyses on these three gene sets because each of them exhibits a burst of activity upon Lipid A stimulation and returns to near basal transcription within the 2-h time course, thereby allowing us to track the processing and movement of transcripts within the nuclear fractions. Additional classes of Lipid A-responsive genes (Groups C–F) did not return to basal transcription levels during the time course, precluding further splicing analysis (data not shown). Characterization of the expression kinetics of genes comprising Groups A_1–F is presented elsewhere (Bhatt et al. 2012).

To determine the timing of splicing relative to transcription initiation, the exon body RPKM was compared with the intron and the spliced exon RPKMs over time. We find that Group A_1 genes show similar patterns of exon body and intron body read accumulation on the chromatin between 0 and 15 min (Fig. 1C, blue and red). Transcriptional volume on these genes is maximal at 15 min, when the majority of A_1 transcripts appear to be completely synthesized (Fig. 2A; Supplemental Fig. S3A,E, green; and data not shown). Interestingly, the peak of spliced RNA from these genes is not reached until 30 min (Fig. 1C, cf. the black and blue lines). Thus, chromatin-associated Group A_1 transcripts exhibit an observable delay between the accumulation of nascent transcripts and that of the spliced RNA (Fig. 1C, cf. the red and black lines). The length of this delay cannot be precisely measured given rapid induction of A_1 genes and the time resolution of the experiment. However, the time of maximum increase in spliced reads is offset from the rapid rise in intron and exon reads by 15 min and occurs after transcription through the genes is complete (see below). We find that the nucleoplasmic accumulation of A_1 transcripts also begins between 0 and 15 min, but does not peak until 30 min (Fig. 1D, blue). The early accumulating nucleoplasmic RNA includes greater numbers of intron body reads than are seen in more slowly induced genes (cf. Fig. 1D and Fig. 1F,

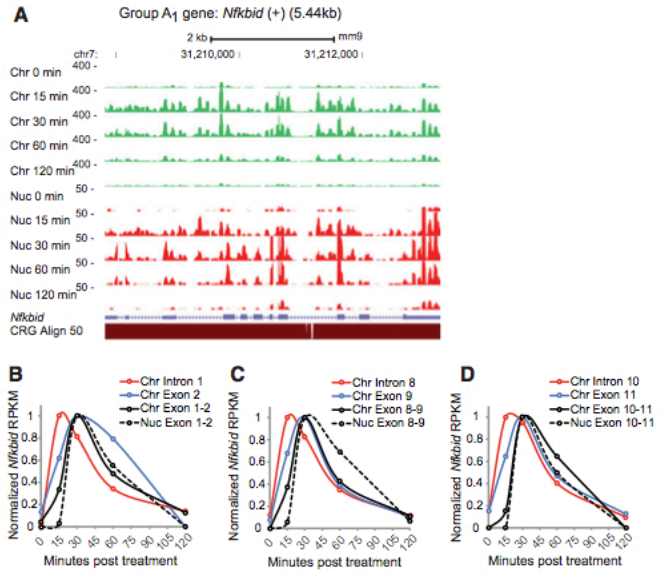


FIGURE 2. Splicing and release of *Nfkbid* transcripts. (*A*) Plot of all reads mapping to the 5.44-kb *Nfkbid* gene from the chromatin (green) and nucleoplasmic (red) fractions at each time point post-Lipid A treatment aligned to the UCSC genome browser track. The gene structure and mappability for the 50-nt reads are depicted *below*. (*B–D*) Line graphs representing the RPKM (normalized to the maximum value) of individual intron (red) or exon (blue) regions within the *Nfkbid* gene in the chromatin fraction at each time point post-Lipid A treatment. The normalized read numbers mapping to the corresponding exon–exon junctions in the chromatin fraction (black, solid) and nucleoplasm (black, dashed) are also plotted.

H, red). These reads may derive from excised introns released from the chromatin and/or a minor subset of incompletely spliced transcripts being released into the nucleoplasm. Taken together, these data show that, for Group A_1 transcripts, the nucleoplasmic accumulation of RNA is delayed relative to its production on the chromatin, but that, since the maximal accumulation of spliced RNA coincides with that of the exon sequences in the nucleoplasm (Fig. 1D, blue and black), the RNA released from the chromatin after 15 min is largely spliced.

A delay in splicing relative to the onset of transcription is also observed for A_2 genes. Production of RNA on the chromatin from A_2 genes begins between 0 and 15 min after Lipid A treatment (Fig. 1E, blue) and is accompanied by a commensurate rise in intron reads (Fig. 1E, red). There is no observable accumulation of spliced reads during this time period (Fig. 1E, black). In the next interval (15–30 min), a rapid increase in spliced exon reads is observed (Fig. 1E, black). Intron reads continue to increase during this interval, but at a slower rate than exon body reads, presumably due to splicing. In contrast to the A_1 transcripts, exon body, intron, and spliced exon reads mapping to A_2 genes all peak simultaneously on the chromatin, in spite of the delayed onset of the spliced reads (Fig. 1E). For this group, the nucleoplasmic intron reads are at the level of background contamination from the chromatin (Fig. 1F, red). Notably, the exon body and spliced reads do not reach maximal values in the nucleoplasm until 60 min. Thus, Group A_2 spliced RNAs exhibit an observable delay in nucleoplasmic accumulation relative to their apparent completion on the chromatin (Fig. 1E,F, black).

Group B transcripts show a pattern of read accumulation on the chromatin similar to Group A_2 transcripts, except that the maximal values are shifted to a later time. Exon body and intron reads accumulate at equal rates between 15 and 30 min (Fig. 1G, blue and red), but substantial spliced read accumulation does not begin until after 30 min (Fig. 1G, black). After this time, accumulation of exon body reads continues, but intron reads begin to plateau, presumably due to splicing. Similar to Group A_2, the peak of spliced reads for Group B genes coincides with the peak of total transcript, even though its initial onset is delayed (Fig. 1E,G, cf. black and blue). In the nucleoplasm, both exon body and spliced reads continue to accumulate at 2 h post-Lipid A stimulation, a time when their abundance on chromatin is decreasing. The intron body reads in the nucleoplasm are at background levels, indicating that these transcripts are nearly completely spliced upon release or are rapidly degraded if released when splicing is incomplete (Fig. 1H, red).

Taken together, these data indicate that transcription, splicing, and chromatin release can be followed over time. Moreover, the three groups of genes differ in their accumulation rates for the different types of reads, but splicing generally precedes nucleoplasmic accumulation. Excision of constitutive introns has been measured as ranging from 2 to 10 min after synthesis of the 3′ exon in mammalian cells, and faster in yeast (Singh and Padgett 2009; Alexander et al. 2010a; Schmidt et al. 2011). The delayed accumulation of spliced exon reads relative to RNA synthesis in the Lipid A-induced genes indicates that some introns in this set might exhibit significantly slower than typical excision rates. To measure these rates, we examined individual genes and introns within each of the three groups.

Individual Group A_1 transcripts exhibit differences in their kinetics of splicing and chromatin release

Almost all Group A_1 genes are <10 kb long and contain three or fewer introns (Supplemental Table S1; Bhatt et al. 2012). As expected, this makes the kinetics of their induction, splicing, and release too fast to completely resolve using 15-min time intervals. For example, at 15 min the *c-Fos* gene has already reached peak values for intron, exon body, and spliced reads on the chromatin, as well as exon and spliced reads in the nucleoplasm (Supplemental Fig. S3A–D). Interestingly, not all Group A_1 genes had the same kinetics for splicing and nucleoplasmic accumulation. The gene *Nfkbid* is only 5.44 kb long and shows complete transcription across the locus by 15 min (Fig. 2A, green). *Nfkbid* differs from other Group A_1 genes in that it contains 10 introns (Fig. 2A, short isoform). We examined the accumulation of reads for several introns and exons along its length. As expected, we found that reads for introns 1, 8, and 10 peak between 15 and 30 min on the chromatin, after which RPKM decreases for all introns (Fig. 2B–D, red). The drop in intronic RPKM after 15 min coincides with a rapid increase in spliced exon reads, which reach their maxima ~30 min (Fig. 2B–D, black). Reads for the body of exons 2, 9, and 11 show similar rapid increases in the first 15 min but peak later as these reads derive from both spliced and unspliced RNA (Fig. 2B–D, blue). Thus, even though this gene undergoes rapid transcription to synthesize all of its introns and exons by 15 min, the accumulation of spliced RNA is delayed (Fig. 2B–D, black; see also Supplemental Fig. S2 for examples of absolute RPKMs per feature). For these introns, the gain in spliced exons coincides with the decrease in intron sequences, indicating that the excised lariats are short lived. This is seen for many but not all introns (see below and data not shown). Differences in the splicing kinetics of different introns cannot be resolved on this short gene (see below). Interestingly, nucleoplasmic RNA shows no accumulation prior to 15 min and then rapidly rises to peak sometime after 30 min (Fig. 2B–D, dashed). These data show that both splicing and transcript release occur more slowly than synthesis of full *Nfkbid* transcripts on the locus. A similar delay in the splicing relative to the accumulation of total reads on chromatin was observed for the longest gene within this group, *Atf3* (Supplemental Fig. S3E–G). This confirms our previous observations of full-length but incompletely processed RNAs associated with chromatin whose transcription is induced by Lipid A (Bhatt et al. 2012). We suggest that the delayed nucleoplasmic

accumulation allows splicing of all introns to be completed prior to release from the chromatin compartment (see below). It is notable that *Nfkbid* and *c-Fos* are of comparable length and appear to be transcribed at similar rates. However, their rates of splicing differ, and by influencing the time that a transcript is retained on the chromatin these splicing kinetics are determining the rate at which the encoded protein can be expressed in the cytoplasm (Fig. 2; Supplemental Fig. S3).

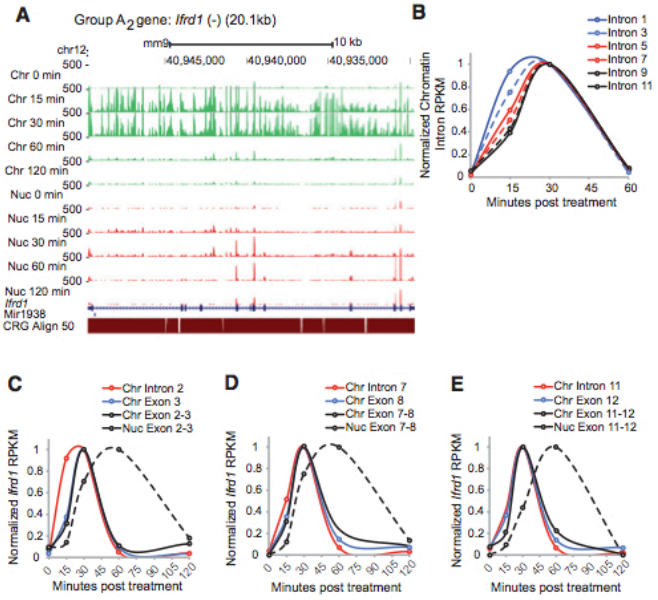


FIGURE 3. mRNA release from chromatin can be significantly delayed relative to synthesis. (*A*) Plot of all reads mapping to the 20.1-kb *Ifrd1* gene from the chromatin (green) and nucleoplasmic (red) fractions at each time point post-Lipid A treatment aligned to the UCSC genome browser track. The gene structure and mappability for the 50-nt reads are depicted *below*. Note that intron 1 contains the gene for mir1938. A microsatellite repeat across the intron 1–exon 2 junction prevents identification of 3′ intron junction and spliced exon 1–2 reads. (*B*) Line graph showing the RPKM (normalized by the maximum value) of the odd-numbered *Ifrd1* introns within chromatin-associated transcripts at each time point post-Lipid A treatment. (*C–E*) Line graphs of the normalized RPKM for individual intron (red) or exon (blue) regions within the *Ifrd1* gene in the chromatin fraction. Reads mapping to the corresponding exon–exon junctions in the chromatin fraction (black, solid) and nucleoplasm (black, dashed) are also shown.

Nucleoplasmic RNA accumulation can be delayed relative to synthesis and splicing on chromatin

Although Group A_1 genes are all short, the other Lipid A-induced gene groups showed broad length distributions (Supplemental Table S1; Bhatt et al. 2012). We found that transcriptional gradients could be detected on Lipid A-responsive genes that were >20 kb. As shown for the Group A_2 gene *Ifrd1*, at 15 min the chromatin-associated reads are more abundant in the 5′ portion of the transcription unit than toward the 3′ end (Fig. 3A, green). By 30 min, reads are more evenly distributed along the length of the gene. Plotting the read accumulation for individual introns, we found that introns 1 and 3 at the 5′ end of the gene (Fig. 3B, blue) accumulate prior to those in the middle (red) or 3′ portions of the transcription unit (black). Intron 2 also showed early accumulation similar to introns 1 and 3 (Fig. 3C, red). The time resolution of the experiment is not sufficient to distinguish more closely spaced introns. Similarly, for the assayable exons the exon body reads and spliced read accumulation could not be resolved from the intron accumulation (Fig. 3B,D,E). These introns are all spliced relatively quickly, although there was a notable delay in the accumulation of spliced exon 2–exon 3 reads after the synthesis of intron 2 (Fig. 3C, black and red). This difference could be due to polymerase pausing, although the profile of RNA accumulation across this region does not show an obvious upstream enrichment that might indicate such a pause.

Interestingly, the *Ifrd1* transcripts display a further delay between the accumulation of spliced reads on the chromatin and their appearance in the nucleoplasm, where they peak at 60 min (Fig. 3C–E, cf. the solid and dashed black lines). This is in contrast to the transcripts analyzed in the Group A_1 that are more rapidly released into the nucleoplasm after being spliced. The delay in nucleoplasmic accumulation might reflect the time required for splicing introns in *Ifrd1* whose excision is slower than those we were able to assay. Whatever mechanisms are mediating this retention, the lag in *Ifrd1* transcript release from the chromatin is determining when *Ifrd1* protein can finally be expressed after stimulation. Even though Ifrd1 transcription is rapidly induced upon stimulation, a substantial delay occurs before its final mRNA is available for translation. Thus, the timing of transcript release will be an important component of the overall regulatory program of inflammation.

Splicing catalysis can occur substantially later than transcript completion

We next examined long genes that provide greater resolution of splicing and transcription over the time course. *Fchsd2* is 175 kb long and contains 21 exons (Fig. 4A). Introns within this gene accumulate in sequence, 5′ to 3′, in the chromatin

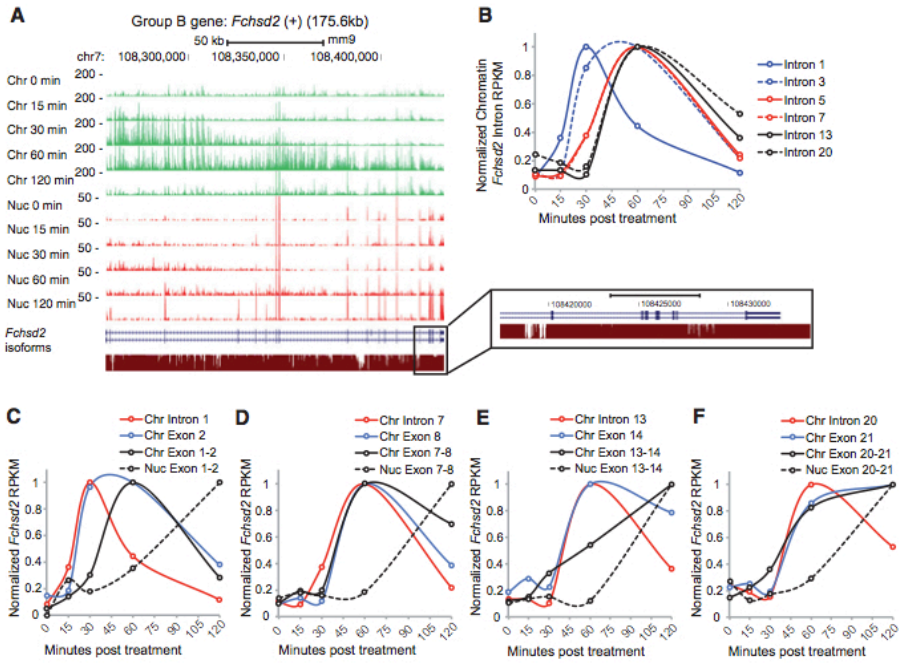


FIGURE 4. Splicing catalysis on long pre-mRNAs can be greatly delayed relative to transcription. (*A*) Plot of all reads mapping to the 175.2-kb *Fchsd2* gene from the chromatin (green) and nucleoplasmic (red) fractions at each time point post-Lipid A treatment aligned to the UCSC genome browser track. The gene structure and mappability for the 50-nt reads are depicted *below*. For clarity, an enlargement of the 3′ region of the gene is included. (*B*) Line graph showing the normalized RPKM of introns within chromatin-associated *Fchsd2* transcripts. (*C–F*) Line graphs representing the RPKM of individual intron (red) or exon (blue) regions within the *Fchsd2* gene in the chromatin fraction. The reads mapping to the corresponding exon–exon junctions in the chromatin fraction (black, solid) and nucleoplasm (black, dashed) are also shown.

fraction (Fig. 4B). We find that intron 1 and its neighboring exon 2 appear almost simultaneously between 15 and 30 min (Fig. 4C, red and blue). However, there is a significant delay in the accumulation of exon 1–2 spliced reads, which begin to accumulate between 30 and 60 min, coincident with the loss of intron 1 reads (Fig. 4C, cf. red and solid black). Similar results are seen for 5′ proximal intron/exon pairs in *Nfkb1* and other long genes (Supplemental Fig. S4 and data not shown). A delay in spliced read accumulation could be due, in part, to pausing of the polymerase within an intron and before its downstream exon. However, we do not observe patterns of read enrichment in introns that might be indicative of transcriptional pauses (Core et al. 2008). In addition, published Chromatin Immunoprecipitation (ChIP-seq) data for RNA Pol II in LPS-stimulated macrophages does not indicate polymerase pausing in these regions (De Santa et al. 2009; Escoubet-Lozach et al. 2011). Most significantly, we find that the spliced exon junction reads from 5′ proximal exons display kinetics of accumulation similar to intron and exon reads mapping to the 3′ terminus of the gene (cf. Fig. 4C, black line, and Fig. 4D–F, blue lines, and Supplemental Fig. S4). These results indicate that splicing catalysis for many of these introns coincides roughly with transcript completion (Fig. 4D,E; Bhatt et al. 2012). Once splicing has begun, spliced reads for exons from the 5′ region generally reach maximal abundance earlier than reads for more 3′ spliced exons (Fig. 4C–F; Supplemental Fig. S4). This may indicate that, although splicing catalysis for some introns is delayed relative to intron synthesis, spliceosome assembly is likely initiating during transcription and affecting the relative timing of some splicing events. However, other factors are also clearly affecting the timing of catalysis. The peak RNA volume on the *Fchsd2* locus occurs at 60 min, but the exons flanking introns 13 and 20 do not reach maximum ligation until at least 120 min. A similar delay is seen in the splicing of the last exon of *Nfkb1* (Supplemental Fig. S4). These data indicate that the timing of inflammatory gene product expression, including *Nfkb1* itself, is not determined solely by the kinetics of transcriptional induction and elongation, but also by the rate of splicing and release of the mRNAs. Importantly, even where splicing

catalysis is completed quite late, the nucleoplasmic accumulation of RNA coincides with the excision of these slowest introns (Fig. 4C–F; Supplemental Fig. S4).

The changes in read accumulation for different portions of a gene over time can be used to estimate the rate of transcription elongation on Lipid A-induced genes. Examining the *Fchsd2* gene, reads at the 5′ end of the gene become detectable at 15 min, whereas those at the 3′ end do not appear until 60 min (Fig. 4A). Thus, RNA Pol II takes at most 45 min to transcribe the 175.2-kb *Fchsd2* gene, yielding a minimum elongation rate of 3.89 kb/min. This value agrees reasonably well with rates of transcription on endogenous mammalian genes determined by a variety of methods (Darzacq et al. 2007; Ardehali and Lis 2009; Singh and Padgett 2009; Brody et al. 2011).

3′ end cleavage and polyadenylation can precede the splicing of nascent transcripts

We previously noted that the reads accumulating on chromatin were largely depleted of sequences downstream from the poly(A) site, and much of the RNA appeared to be polyadenylated (Bhatt et al. 2012). To quantify this 3′ processing and assess its timing, we again stimulated primary macrophages with Lipid A and isolated chromatin-associated RNA at 0, 5, 15, 20, 30, and 60 min. We then primed reverse transcriptase with primers located either downstream from the poly(A) site (Fig. 5A, blue primer) or at this site after cleavage and polyadenylation (Fig. 5A, red primer). The cDNA was subjected to qPCR to determine the molar amount of 3′ terminal exon sequence in each sample. We analyzed short transcripts synthesized from Group A_1 and A_2 genes because their cleavage and polyadenylation status could be assayed early in the time course. Absolute quantification revealed accumulation of both uncleaved and polyadenylated *Nfkbid* RNA on chromatin over the time course (Fig. 5). We find that polyadenylated transcripts account for >75% of the total RNA on chromatin at peak levels of expression, confirming our previous observation of read enrichment upstream of the cleavage and polyadenylation site of most genes (Fig. 5A; Bhatt et al. 2012; and data not shown). This finding is consistent with long-standing observations that polyadenylation occurs rapidly after transcription, and with more recent data that polyadenylated RNAs are retained at a locus after completion of transcription (Derman and Darnell 1974; Nevins and Darnell 1978; Salditt-Georgieff et al. 1980; Brody et al. 2011; Daguenet et al. 2012). Notably, the peak of uncleaved *Nfkbid* RNA on chromatin occurs at 15 min post-Lipid A stimulation, coincident with the first appearance of full-length transcripts (Fig. 5A). Even at this early time, most of the RNA is already cleaved and polyadenylated. 3′ processed RNA continues to accumulate and peaks ~30 min post-stimulation, with similar kinetics to the exon body and spliced exon reads (Figs. 2B–D, 5A). We find that 60% of all reads mapping to spliced exon–exon junctions within *Nfkbid* accumulate on chromatin between 15 and 30 min post-stimulation (Fig. 2B–D, black). During this time period, ~85% of the *Nfkbid* transcripts on chromatin are already polyadenylated (Fig. 5A, red). Thus, a large fraction of intron excision events occur on RNAs that have already been 3′-cleaved and polyadenylated. In the nucleoplasm, uncleaved transcripts are at the level of background, confirming that 3′ end cleavage occurs in association with chromatin (data not shown).

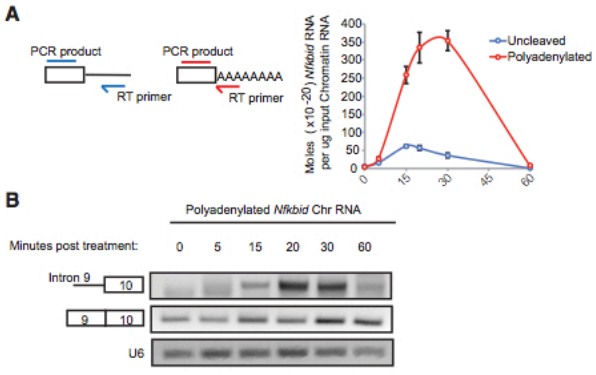


FIGURE 5. Polyadenylation can precede splicing on chromatin. (*A*) Line graph showing the molar amount of *Nfkbid* 3′ terminal exon per microgram of input chromatin-associated RNA at times post-Lipid A simulation. The RNA was primed for reverse transcription using either a primer downstream from the cleavage site (blue) or one complementary to the poly(A) tail (red). Error bars represent the standard error of the mean. (*B*) Ethidium bromide stained agarose gel showing the PCR product of the *Nfkbid* intron 9–exon 10 junction (35 cycles) and the corresponding exon 9–exon 10 spliced product (35 cycles) within the chromatin fraction at times post-Lipid A stimulation. We note that the intron 9–exon 10 product is short, and appears at the 30-min time point. The band visible at the 0- and 15-min time points is a primer dimer. A PCR product (25 cycles) amplified from the U6 snRNA cDNA served as a loading control.

To further confirm that polyadenylated RNAs on chromatin contain introns, we measured the presence of unspliced introns in the polyadenylated *Nfkbid* RNA samples by PCR. The intron 9–exon 10 splice site was readily detected in this RNA and accumulated with the same kinetics as the polyadenylated 3′ exon, peaking on chromatin at 20–30 min post-stimulation (Fig. 5B). Additional introns in *Nfkbid* and other Group A_1 and A_2 transcripts were also seen in the polyadenylated RNA (data not shown). Thus, splicing catalysis for some introns follows 3′ end maturation. This is consistent with long-standing in vivo labeling data, as well as in vitro data showing that cleavage and 3′ end formation can precede splicing, even though

splicing is required for release of transcripts from the polymerase (Derman and Darnell 1974; Nevins and Darnell 1978; Bird et al. 2005; Rigo and Martinson 2009; Darnell 2013).

Evidence for polymerase pausing in the final exon of genes has been observed in yeast (Alexander et al. 2010b; Carrillo Oesterreich et al. 2010). However, transcriptional pauses are likely to occur on a time scale too rapid to be assayed in our data.

Introns containing skipped alternative exons are spliced more slowly than adjacent constitutive introns

We and others previously observed that introns containing skipped exons are present on chromatin at higher levels than other introns, indicating that regions subject to splicing regulation may exhibit slower rates of splicing (Pandya-Jones and Black 2009; Khodor et al. 2011). Our data here suggest that the intron with the slowest excision rate might determine the release kinetics of its parent transcript. To examine an example of exon skipping, we analyzed the Group B *Cd44* gene, which contains a set of well-characterized alternative exons that are organized in tandem (exons 6 through 15) (Fig. 6A; Lynch 2004). These exons are excluded from the mature mRNA in macrophages (Fig. 6A and data not shown). We find that spliced reads across the exon 4–exon 5 junction peak at 60 min, ~30 min after transcription through the intron (Fig. 6B, cf. red and black). In contrast, accumulation of spliced reads across the exon 5–exon 16 junction was delayed relative to all other junctions analyzed (Fig. 6C, black), including the 3′ terminal junction between exons 18 and 19 (Fig. 6D, black). These data confirm the interpretation of steady-state read enrichment within alternatively spliced introns, and indicate that skipping of regulated exons can cause a significant delay in splicing (Pandya-Jones and Black 2009; Khodor et al. 2011, 2012).

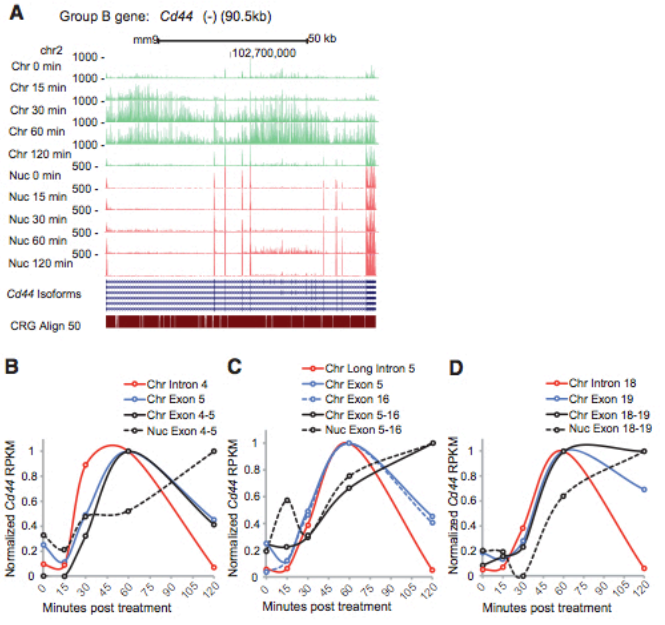


FIGURE 6. The long *Cd44* intron containing skipped exons is spliced more slowly than adjacent constitutive introns. (*A*) Genome browser tracks of reads mapping to the 90.5-kb *Cd44* gene in the chromatin (green) and nucleoplasmic (red) fractions. Alternating SINE and LINE elements interfere in the analysis of reads that map to exon 2. (*B–D*) Line graphs representing the RPKM of individual intron (red) or exon (blue) regions within the *Cd44* gene in the chromatin fraction. The reads mapping to the corresponding exon–exon junctions in the chromatin fraction (black, solid) and nucleoplasm (black, dashed) are also shown.

CD44 is a membrane protein, whose mRNA is associated with the ER. We previously showed that these mRNAs are enriched in the nucleoplasmic fraction, presumably due to the co-fractionation of endoplasmic reticulum (ER) with the contiguous nuclear envelope (Bhatt et al. 2012). This produces a background of *Cd44* mRNA in the nucleoplasmic fraction even prior to induction. Nevertheless, we observe the increase in nucleoplasmic *Cd44* mRNA after gene induction. Notably, this nucleoplasmic accumulation coincides with splicing across exons 5–16 (Fig. 6C), consistent with the idea that the intron with the slowest splicing kinetics within a transcript will determine the timing of release into the soluble nucleoplasm.

To look at other regions encompassing regulated exons, we examined the constitutively expressed gene *Cd45*. The *Cd45* mRNA also contains a group of well-characterized alternative exons that appear to be predominantly skipped in macrophages (Fig. 7A; Tong et al. 2005). As this gene is equally transcribed at all time points, we cannot follow its synthesis and splicing over time. However, as observed on *Cd44*, there is a higher read density in the long *Cd45* intron containing the skipped exons 4, 5, and 6 than elsewhere along the locus (Fig. 7A,B). This intron accumulation might, in part, result from Pol II pausing as described in a T-lymphocyte cell line (Shukla et al. 2011). However, reads for the exons downstream from exon 7 are enriched in the chromatin fraction indicating that they may be spliced prior to exons 3–7 (Fig. 7A). Moreover, the transcriptional pausing in T-cells was associated with exon inclusion and not skipping, and in published data sets for LPS-stimulated macrophages there

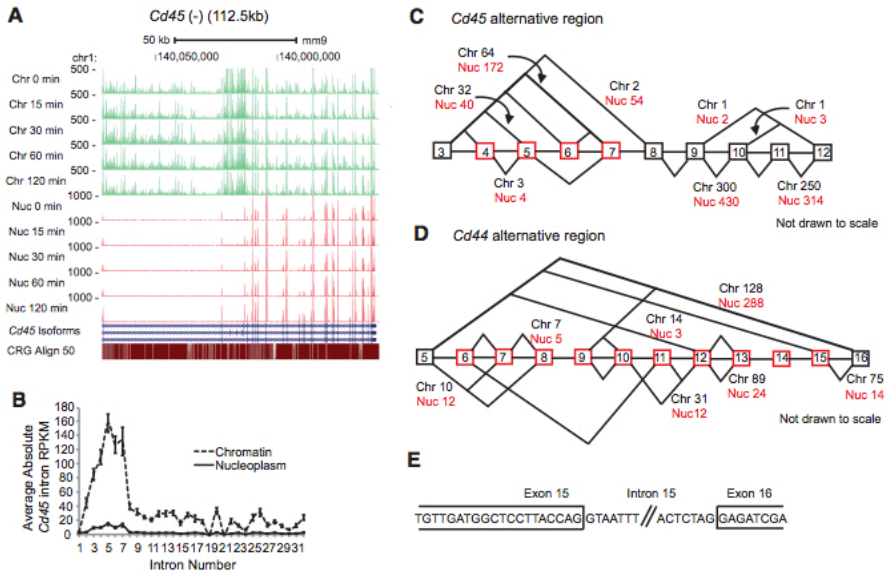


FIGURE 7. Aberrant splicing of skipped alternative exons. (*A*) Genome browser tracks of reads mapping to the 112.5-kb *Cd45* gene in the chromatin (green) and nucleoplasmic (red) fractions. This gene is constitutively expressed, thus each time point can be considered a biological replicate. (*B*) Line graph showing the total RPKM for each intron of the *Cd45* gene in the chromatin and nucleoplasmic fractions, averaged over all five time points. Error bars represent the standard error of the mean. (*C*) Diagram of *Cd45* exons 3–12. Exons 4–7 outlined in red are alternative, although in macrophages the dominant isoform includes exon 7. Diagonal lines join exons where chromatin-associated splicing events were observed. The predominant alternative splicing pattern is shown in bold. Numbers indicate the number of reads mapping to the indicated junction averaged over all five time points in the chromatin fraction. For clarity, the information for all junctions within the depicted region is not included. (*D*) A diagram of the *Cd44* alternative region. Constitutive exons 5 and 16 are outlined in black, and alternative exons 6–15 in red. Diagonal lines join exons where splicing events were observed. Bold lines depict productive splicing between exons 5 and 16. Numbers report the reads per event at 60 min on the chromatin. For clarity, the information for all junctions is not included. The values shown are representative of the range in read number per junction found across the *Cd44* alternative region. (*E*) Sequence of the 3′ end of exon 15 and the 5′ end of exon 16 in the *Cd44* gene.

is no obvious enrichment of Pol II ChIP-seq reads in this region (Escoubet-Lozach et al. 2011; Shukla et al. 2011). In macrophages, the increased chromatin-associated read density extends through the entire long intron encompassing all the skipped exons (Fig. 7A and B). These results are consistent with the observations of *Cd44* that the splicing machinery takes longer to catalyze excision of long introns encompassing exons that will be excluded from the mature RNA.

Splicing to cryptic splice sites and the incorrect joining of exons is rare

Given the observed delays between the synthesis and splicing of some introns, we examined whether aberrant splicing could be observed in these regions. Within *Cd44* intron 1, we found multiple instances of spliced reads that map to pairs of cryptic splice sites internal to the two correct sites (Supplemental Fig. S5). These reads were rare; only one or two reads were found for each aberrant event over all five time points, compared with hundreds of reads mapping to the correct junctions (Supplemental Fig. S5 and data not shown). Almost all the observed aberrant splicing events utilized GU–AG splice sites (Supplemental Fig. S5 and data not shown), indicating their likely origin from cryptic splicing by the spliceosome and not some other process. These cryptic splicing events in intron 1 are notable in that they will be eliminated from the final mRNA by the correct splicing of the intron (Supplemental Fig. S5).

Interestingly, we did not find a single read derived from the incorrect joining of two constitutive exons within *Cd44* (data not shown). We did find some reads of this type in the *Cd45* gene, albeit at extremely low frequency (Fig. 7C). The skipping of constitutive exons can generate mRNAs subject to nonsense-mediated decay (NMD). NMD requires translation and should lead to the loss of cytoplasmic mRNA and not the nuclear species. These observations suggest that the fidelity of splice site recognition by the spliceosome is high, or that processes that remove aberrantly spliced transcripts prior to template release are very efficient.

Nevertheless, there is a measurable error rate, with some of the observed errors possibly eliminated by subsequent correct splicing.

Although the incorrect joining of constitutive exons appears to be rare, we found more significant numbers of reads derived from splicing of the variable exons in *Cd44* and *Cd45* (Fig. 7C,D; variable exons are highlighted in red). Both *Cd44* and *Cd45* encode membrane proteins whose final mRNAs fractionate with the nucleoplasm, preventing measurement of relative exon levels in the cytoplasm. In the nucleoplasm there are very few exon body reads for the variable exons of *Cd44* and *Cd45* (Figs. 6A, 7A). RPKM and total read numbers are not directly comparable between fractions because they are sampling from RNA populations with different overall complexities. As a general comparison, reads for correct exon–exon junctions were generated in higher numbers from nucleoplasmic RNA than from chromatin, because the loss of intron sequence makes the spliced RNA a larger fraction of the nucleoplasmic RNA pool (data not shown). In contrast, reads for exon–exon joints involving the skipped exons decreased in nucleoplasm relative to chromatin, indicating a loss of the incorrectly spliced RNA (data not shown). On the chromatin, we find that spliced reads derived from the joining of skipped alternative exons occurred at higher frequencies than reads mapping to incorrectly paired constitutive exons. In fact, reads for certain alternative exon–exon junctions were nearly as abundant as those mapping to correctly joined constitutive exons (Fig. 7C,D). Thus, there is apparently a higher error rate for the joining of these alternative exons, even though these aberrant products are subsequently eliminated.

In *Cd44*, the variable exons are nested between exons 5 and 16, where their exclusion from the mRNA would be achieved by excision of the long intron, whether they had previously been spliced to each other or not (Fig. 7D). Many of the most abundant splicing events involving the *Cd44* variable exons could be eliminated in this way. In contrast, events joining a variable exon to one of the flanking constitutive exons could be eliminated by a second splicing event only if a new splice site was generated to allow subsequent splicing of exon 5 to exon 16. Interestingly, for exon 16, the only observed alternative splicing event involved splicing to exon 15. This event was quite frequent, occurring on chromatin at ~60% the frequency of exon 5–16 splicing. Notably, the junction between exons 15 and 16 has potential to regenerate a 3′ splice site that could be used in resplicing of exon 16 to exon 5, thus resolving the incorrect splicing event (Fig. 7E). In contrast, splicing events between exon 5 and the downstream alternative exons 7 and 8, and between *Cd45* exons 3 and 5, would not generate new splice sites at the exon–exon junctions. These products would have to be eliminated by other mechanisms, perhaps through decay processes acting on incompletely spliced RNA. These findings indicate that some nonproductive splicing occurs between exons, and that this may be subsequently rescued by resplicing.

DISCUSSION

In this study, we characterize the relative kinetics of transcription, splicing, and transcript release from the chromatin compartment for genes of the inflammatory response. Stimulating primary macrophages with Lipid A, we induced transcription of a large set of endogenous genes. We applied RNA-seq to isolates of chromatin-associated and soluble nuclear RNA collected over time to measure the rates of exon ligation and transcript release from chromatin relative to RNA synthesis. This experimental approach allowed us to follow the RNA products as they were transcriptionally elongated, terminated, spliced, and then released into the nucleoplasm. We find that the rates of later steps in mRNA maturation strongly affect the timing of when a final protein product can be made after stimulation. This timing has important implications for understanding the inflammatory gene cascade. For example, *Nfkbid*, which is rapidly induced by Lipid A, encodes a nuclear inhibitor of NFkB. Unlike another immediate early gene, *c-Fos*, *Nfkbid* mRNA maturation and release are delayed by ~15 min, thereby slowing the negative feedback on NFkB activity that it controls. Thus, these rates of mRNA maturation and release will need to be incorporated into predictive models of the inflammatory gene response (Hetru and Hoffmann 2009).

We find that the appearance of some spliced exon–exon junctions can be substantially delayed relative to synthesis of the exons themselves and that introns in the Lipid A-induced gene set often are excised after the time needed to fully transcribe a locus. These introns can be substantially slower in their excision than other introns whose removal occurs prior to transcript completion. In contrast, 3′ end cleavage of nascent transcripts occurs rapidly after transcription across the poly(A) site. These differing kinetics result in pools of transcripts associated with chromatin that are cleaved and polyadenylated at the 3′ end but are not yet completely spliced (Fig. 8). This is consistent with previous results demonstrating that cleavage and polyadenylation can occur prior to splicing in adenoviral and mammalian transcripts, as well as with observations that pools of polyadenylated RNA can be observed at active gene loci (Nevins and Darnell 1978; Harpold et al. 1979; Weber et al. 1980; Kessler et al. 1993; Rigo and Martinson 2009; Brody et al. 2011).

We observe that a lag occurs after the completion of transcription before transcript release from the chromatin fraction into the nucleoplasm, and that RNA passes through this nucleoplasmic fraction prior to appearance in the cytoplasm (Fig. 8; Bhatt et al. 2012). Introns in sporadic constitutive genes are detectable in the nucleoplasmic fraction, and imaging data indicate that some introns may be excised away from their chromatin template (Supplemental Fig. S1; Vargas et al. 2011; Girard et al. 2012). However, measurements of steady-state RNA levels from most constitutive genes found intronic RNA sequence primarily in the chromatin fraction (Supplemental Fig. S1; Wuarin and Schibler 1994;

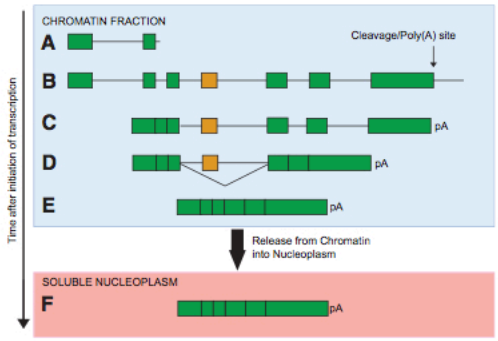


FIGURE 8. The sequence of maturation events observed for Lipid A-induced transcripts. We find that splicing and transcript release from chromatin are slow relative to RNA synthesis and 3′ end cleavage. This gives rise to the following sequence of events for Lipid A-induced transcripts. (*A*) Transcription initiates upon Lipid A stimulation. (*B*) Transcription of the gene completes. Some splicing likely initiates prior to this, but for the genes assayed here, spliced read accumulation largely coincided with the time of transcript completion or later. (*C*) Cleavage and polyadenylation at the 3′ splice site is rapid once transcription proceeds across the site. Introns are removed roughly in a 5′-to-3′ order along the transcript. (*D*) Introns encompassing an alternative exon that is to be excluded from the final message are spliced especially slowly. (*E*) Fully processed, mature mRNAs are released from the DNA template. (*F*) RNA released into the nucleoplasm is largely complete.

Pandya-Jones and Black 2009; Khodor et al. 2011, 2012; Tilgner et al. 2012; and data not shown). If these transcripts become dissociated from the locus after transcription, they must rapidly rebind to components of the chromatin pellet. We find that serine/arginine (SR) proteins and other speckle components do not pellet with the chromatin and that the association of RNA with this fraction is sensitive to DNase treatment, indicating that its interaction requires intact DNA (data not shown). Thus, the RNA in the chromatin fraction is most likely still associated with the locus. However, this RNA is not necessarily still bound to the Pol II on the DNA template. The fraction presumably contains what have been called perispeckles that contain polyadenylated RNA and are adjacent to speckles and gene loci (Brody et al. 2011; Daguenet et al. 2012). Once released into the soluble nucleoplasm, transcripts rapidly appear in the cytoplasmic fraction, in agreement with relatively rapid rates of RNA nuclear export observed by microscopy (Ben-Ari et al. 2010; Grunwald and Singer 2010; Bhatt et al. 2012).

In the Lipid A-induced genes, we find nearly all intron-containing RNA within the chromatin fraction, and that RNA within the soluble nucleoplasm is fully spliced. For these Lipid A-induced genes, the release of RNA into the nucleoplasm coincided with the splicing of the slowest intron. Thus, the lag in transcript release from the chromatin may reflect the time needed to complete splicing. A requirement for complete splicing prior to release is consistent with many studies showing an accumulation of RNA at gene loci in response to perturbations of the splicing reaction or disruptions in mRNA export (Custodio et al. 1999; Jensen et al. 2001; Abruzzi et al. 2006; Brody et al. 2011; Martins et al. 2011). The coupling of final transcript maturation to template release would provide a site of action for quality control mechanisms that act on incorrectly processed RNAs prior to their entry into the nuclear export pathway. It will be very interesting to examine how factors involved in RNA quality control and export, such as the TREX/THO complexes, affect the distribution of mRNAs between the chromatin and nucleoplasmic fractions (Strasser et al. 2002; Rougemaille et al. 2008; Dias et al. 2010; Schmid and Jensen 2010).

Some introns are excised during RNA synthesis and others later (Bauren and Wieslander 1994; Tennyson et al. 1995; Wetterberg et al. 1996). Singh and Padgett (2009) made the most extensive measurements of splicing rates, examining constitutive transcripts restarted after DRB inhibition of transcription. They found that introns are typically excised ~8 min after synthesis of the 3′ exon. Although a large fraction of the introns we examined are excised at comparable rates, some are much slower. One interesting possibility is that the transcriptional induction process by Lipid A contributes to a delay in splicing for certain introns. Factors required for these slow introns may not be immediately recruited to a locus, and this would contribute to the timing of the evolving Lipid A response. In contrast, introns in transcripts from constitutively expressed genes may be tuned for rapid excision. However, the patterns of steady state RNA accumulation on the introns and exons of constitutive genes are highly variable, and this may be an indication of different rates of intron excision on these transcripts as well.

The source of the differing intron excision kinetics is not clear. In vitro, the rate of splicing can be altered by changing the strength of the 5′ splice site or the length of the intron, or by adding enhancer or silencer elements (Hertel and Maniatis 1998; Yu et al. 2008; Hicks et al. 2010). We did not observe correlations of either splice site strength or intron length with the rates of splicing (data not shown). The variability in intron excision rates presumably reflects rate differences in steps of spliceosome assembly and catalysis subsequent to initial splice site recognition and perhaps subsequent to splice site pairing.

Regulated introns can be particularly slow in their excision. In earlier observations, introns containing exons that will be excluded from the final mRNA were found to be present on the chromatin at higher levels than adjacent constitutive introns (Pandya-Jones and Black 2009; Khodor et al. 2011, 2012). In now measuring the excision rate of such an intron in *Cd44*, we find that it is indeed excised much more slowly than the other introns in the transcript. Splicing of complex regions that offer competing choices for spliceosome assembly may involve additional rate limiting regulatory steps prior to proper exon ligation. For example, there is evidence that SR proteins bound within introns are inhibitory for splicing

(Kanopka et al. 1996; Ibrahim et al. 2005). Since exons that are excluded during splicing will likely bind SR proteins at their exonic splicing enhancer elements, these proteins must somehow be removed or ignored during splicing. These results underscore how little we know regarding the kinetics of spliceosome assembly, which steps are rate limiting, and how features of the pre-mRNA might alter those rates.

There is substantial evidence for cross-talk between transcription and splicing. Specifically, transcription elongation rates are thought to affect the choice of some alternative splicing patterns (Perales and Bentley 2009; de la Mata et al. 2010; Pandya-Jones 2011). Although intron excision can be slow relative to RNA synthesis, as we show here, splice site pairing choices may still occur during transcription. In fact, introns near the 5′ end of long genes do appear to be excised prior to introns near the 3′ end, even though some splicing catalysis is occurring long after the intron is synthesized; this may be indicative of the sequence of spliceosome assembly during transcription (de la Mata et al. 2010). On the other hand, the long delay in the excision of regulated introns relative to their synthesis offers additional opportunities for altering the choice of splicing patterns. There is evidence that regulatory proteins can alter splice site choice after the initial pairing of splice sites (Lim and Hertel 2004; Kotlajich et al. 2009). It will be interesting to examine whether pools of completed but unspliced transcripts can be differentially spliced depending on cellular conditions. It has been shown that a pool of partially spliced Clk1/4 mRNA is stored for later completion in response to stress, but this RNA does not appear to be at the Clk1/4 locus (Ninomiya et al. 2011).

In a previous study, we examined the splicing of *c-Src* and *Fibronectin* transcripts in chromatin and nucleoplasmic fractions of tissue culture cells by RT-PCR (Pandya-Jones and Black 2009). As described above, the results of this study were largely consistent with our observations here, except that we found increasing levels of intron sequence at the 3′ ends of genes. We interpreted the 5′-to-3′ gradient in intron abundance as reflecting the excision of introns as polymerase progresses through the locus. This may be correct for *c-Src*, *Fn*, and other constitutive genes. However, these measurements were of steady-state RNA levels where the rates of splicing and transcription could not be directly assessed. We find here that the patterns of intron and exon abundance of constitutive genes measured at steady state are highly variable and that it is difficult to draw general conclusions about the order of intron excision from these data. Moreover, we find that intron excision may still progress in a roughly 5′-to-3′ order along a transcript even if splicing initiates after transcription has proceeded through most of the gene.

Other recent studies also applied RNA-seq analysis to chromatin-associated RNA of constitutive genes (Khodor et al. 2011, 2012; Tilgner et al. 2012). Khodor and colleagues examined *Drosophila melanogaster* and rat liver genes at steady state, with the added step of removing poly(A)$^+$ RNA from the chromatin RNA pool to ensure only nascent chains were present. While the differences in the experimental systems prevent the exact comparison of their results with ours, the authors similarly conclude that splicing largely occurs in the chromatin fraction and that regulated introns are slow to splice. The authors also report higher levels of introns in nascent mammalian transcripts than in *Drosophila*, indicating that splicing may be slower relative to RNA synthesis for mammalian genes (Khodor et al. 2012). Another genome-wide metagene analysis of RNA-seq data from K562 cells also found a modest cross-gene gradient of ligated exons (Tilgner et al. 2012). These authors found that long noncoding RNAs may also have relatively slowly spliced introns, although again the rates could not be determined from steady-state conditions.

In another study, RNA-seq was applied to total human brain RNA. Saw-tooth patterns of decreasing read density observed across introns were interpreted as indicating rapid intron excision as the polymerase passed through a 3′ splice site (Ameur et al. 2011). This analysis was conducted on total, steady-state RNA containing both fully spliced nuclear and cytoplasmic RNA, as well as immature RNA presumably associated with the chromatin. This precluded examination of exon abundance in the nascent RNA. In examining steady-state genes that are not affected by Lipid A stimulation, we also find examples of saw-tooth patterns of read density across introns (Supplemental Fig. S1); however, these patterns are not pervasive across all introns nor all genes (see also Khodor et al. 2012). Moreover, although there is some enrichment of exon reads surrounding these introns, we do not observe decreasing 5′-to-3′ gradients of exon reads along the length of these genes that would indicate very rapid intron excision during transcription. It is important to point out that there is wide variation in the patterns of RNA-seq read accumulation between individual genes, particularly at steady state. Thus, we would not rule out the possible rapid excision of some introns, including in *c-Src* and *Fn* transcripts (although also see Khodor et al. 2012). The pattern of decreasing 5′-to-3′ read densities within introns was not generally seen in the Lipid A-induced transcripts, but again it is possible that there are differences in intron excision rates between some constitutive genes and Lipid A inducible genes.

The error rate of the splicing process is difficult to estimate in vivo. In analyzing total RNA or mRNA, a variety of quality control mechanisms have likely degraded many of the nonfunctional mRNAs produced from aberrant splicing patterns. We examined whether we might observe higher levels of aberrant products in our nascent RNA fraction. In general, the fidelity of splice site selection appears to be high, even when splicing is relatively slow and abundant unspliced RNA accumulates at a gene locus (Fox-Walsh and Hertel 2009). Nevertheless, we do find low levels of aberrant splicing events occurring in the analyzed transcription units. The more commonly observed errors appear to be splicing events that will be resolved by subsequent splicing at the correct junctions. These include splicing at cryptic splice sites internal to the

correct sites and splicing between alternative exons that will not affect the final mRNA after excision of the long intron that contains them. It also appears that *Cd44* exon 16 may join to exon 15 only to subsequently resplice to exon 5. The resplicing of exon 16 is conceptually similar to the recursive splicing that has been described in a long intron of the *Drosophila Ubx* locus (Hatton et al. 1998; Burnette et al. 2005). In that case, a 5′ splice site is repeatedly respliced to a series of 3′ splice sites along the long intron. We did not observe intermediate spliced products for this kind of recursive splicing in the genes analyzed, but it may occur on other mammalian genes (although see Shepard et al. 2009). Aberrant spliced products that can be resolved by subsequent resplicing may not have a strong selection against their occurrence. Further analyses of nascent RNA after induction at greater sequence depth should allow more accurate measurement of splicing error rates.

This experimental approach can be applied to characterizing splicing responses to myriad other stimuli and cellular conditions. The isolation of the chromatin-associated RNA pool offers more rapid measurement of RNA changes than the traditional analysis of cytoplasmic mRNA. The method can also be applied to other RNA processing events, such as the maturation of microRNAs. A majority of miRNA genes are housed within introns and their processing is thought to occur in association with chromatin (Kim and Kim 2007; Pawlicki and Steitz 2008). We identified a miRNA within the Lipid A-induced gene set, miR1938 encoded within intron 1 of the *Ifrd1* gene (Fig. 3). We did not observe obvious differences in the excision kinetics for this intron, nor did the intron show unusually high read numbers after excision, as might be expected if the lariat product were particularly stable (data not shown). Using this approach, it will be interesting to assess where and when miRNA maturation occurs relative to the processing of host mRNAs. In this way, we hope to obtain much needed information on the maturation rates of a variety of transcripts.

MATERIALS AND METHODS

Cell culture

Bone marrow derived macrophages were isolated from mixed-sex C57BL/6 mice and cultured and stimulated with Lipid A as previously described (Bhatt et al. 2012).

RNA isolation

Cells were collected at 0, 15, 30, 60, and 120 min post-stimulation. Chromatin-associated and nucleoplasmic RNA fractions were isolated and characterized as previously described (Pandya-Jones and Black 2009; Bhatt et al. 2012). To estimate the loss of RNA from the chromatin into the nucleoplasm during purification, we examined the nuclear *Xist* transcripts. We found that >95% of the total *Xist* RNA reads at all time points were found in the chromatin fraction (Supplemental Fig. S1A), as would be expected given its function in initiating X chromosome silencing (Wutz 2011). Read numbers between the chromatin and nucleoplasmic fractions cannot be compared directly, because of differences in library construction and the sequence complexity between the two fractions. Nevertheless, the chromatin fraction generated 20–40 times more *Xist* reads than the nucleoplasm, indicating that the *Xist* association with chromatin was stable to the fractionation. These data indicate that there is little loss of RNA from the chromatin into the nucleoplasm during purification and that the two fractions are composed of distinct RNA populations.

Library construction

Chromatin-associated and nucleoplasmic RNA fractions were subjected to library preparation as outlined in the Illumina mRNA library preparation kit. Ribosomal RNA was removed using the Ribominus kit (Invitrogen). A modification to the protocol involved the incorporation of dUTP into the second strand synthesis as described in Levin et al. (2010) to create a strand-specific library.

Mapping statistics

Fifty-nucleotide reads were aligned to the mouse mm9 male canonical genome using TopHat on Galaxy (options: −a = 6; −m = 0; −i = 50; −I = 500,000; −F = 0; −g = 1; segment length = 25, segment mismatch = 2) (Giardine et al. 2005; Trapnell et al. 2009; Blankenberg et al. 2010; Goecks et al. 2010).

Normalization and RPKM calculations

Reads were assigned to genic regions using the RefSeq gene database and the BEDTools suite (Quinlan and Hall 2010). Exon and intron body reads were calculated using the annotated UCSC known gene database and the BEDTools commands coverageBed piped after intersectBed with flag −f 1.0. The same command was used to extract the 3′ intron junction reads, except −f 0.02, since we required that the reads overlap the splice site by 1 nt. The spliced exon–exon reads were identified using the intersectBed command and curated manually. The RPKM values were calculated by normalizing the number of reads mapping to a feature by the total number of reads mapped per sample, as well as the length of the feature in kilobase. For spliced reads crossing an exon–exon junction, the feature size was set at 88 nt because we required a 6-nt overlap across the junction. We limited analysis of ligated exon junctions to those with ≥15 mapped reads at peak expression, with the vast majority of junctions accumulating between 25 and 200 reads at this time point (Pan et al. 2008). For the 3′ intron junction reads, feature length was set at 98 nt to force a 1-nt overlap with the 3′ splice site.

Quantitative RT-PCR

Five-hundred nanograms of input RNA, as quantified on a nanodrop-1000 spectrophotometer, was random-primed according to the manufacturers protocol (Superscript III, Invitrogen). One-tenth of the cDNA and appropriate primers were added to LightCycler qPCR mix (Roche) and used in a qPCR reaction as described by the manufacturer. All qPCR reactions were performed on an MX300P machine (Stratagene). The efficiency of each primer set was deter-

mined using a standard curve and quantification was normalized against absolute values of U6 snRNA. Primer sequences used in the analysis of *Nfkbid* are as follows:

Reverse primer against uncleaved RNA: ccccactccccatttttaat
Reverse primer against poly(A) RNA: ttttttttttttttttttttttttctaa
Exon 11 forward: GGAGTTTCCATGGGATTTGA
Exon 11 reverse: CACAGACCAAGGGGAACAAT
Exon 9 forward: CCATTCTGCACTTGGCTGTA
Intron 9 forward: ggttcaaagacccagagacg
Exon 10 reverse: GCAGTGTCGGGTCTGCTC

DATA DEPOSITION

Sequencing data have been submitted to GEO under accession number GSE32916.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Jim Darnell, Phil Sharp, Tim Nilsen, and members of the Black and Smale labs for helpful discussions and/or for advice on the manuscript. This work was supported by a UCLA Dissertation Year Fellowship (to A.P-J.) and by NIH grants R01GM086372 and R01CA127279 (to S.T.S.), and by the Broad Stem Cell Research Center at UCLA. D.L.B. is an Investigator of the Howard Hughes Medical Institute.

Received February 27, 2013; accepted March 13, 2013.

REFERENCES

Abruzzi KC, Belostotsky DA, Chekanova JA, Dower K, Rosbash M. 2006. 3′-end formation signals modulate the association of genes with the nuclear periphery as well as mRNP dot formation. *EMBO J* **25:** 4253–4262.

Alexander RD, Barrass JD, Dichtl B, Kos M, Obtulowicz T, Robert MC, Koper M, Karkusiewicz I, Mariconti L, Tollervey D, et al. 2010a. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*. *RNA* **16:** 2570–2580.

Alexander RD, Innocente SA, Barrass JD, Beggs JD. 2010b. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* **40:** 582–593.

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18:** 1435–1440.

Ardehali MB, Lis JT. 2009. Tracking rates of transcription and splicing *in vivo*. *Nat Struct Mol Biol* **16:** 1123–1124.

Bauren G, Wieslander L. 1994. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76:** 183–192.

Ben-Ari Y, Brody Y, Kinor N, Mor A, Tsukamoto T, Spector DL, Singer RH, Shav-Tal Y. 2010. The life of an mRNA in space and time. *J Cell Sci* **123:** 1761–1774.

Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150:** 279–290.

Bird G, Fong N, Gatlin JC, Farabaugh S, Bentley DL. 2005. Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. *Mol Cell* **20:** 747–758.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. 2010. Galaxy: A web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* Chapter 19: Unit 19 10 11–21.

Brody Y, Neufeld N, Bieberstein N, Causse SZ, Bohnlein EM, Neugebauer KM, Darzacq X, Shav-Tal Y. 2011. The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol* **9:** e1000573.

Burnette JM, Miyamoto-Sato E, Schaub MA, Conklin J, Lopez AJ. 2005. Subdivision of large introns in *Drosophila* by recursive splicing at nonexonic elements. *Genetics* **170:** 661–674.

Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40:** 571–581.

Chan S, Choi EA, Shi Y. 2011. Pre-mRNA 3′-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* **2:** 321–335.

Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11:** 3319–3326.

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

Custodio N, Carmo-Fonseca M, Geraghty F, Pereira HS, Grosveld F, Antoniou M. 1999. Inefficient processing impairs release of RNA from the site of transcription. *EMBO J* **18:** 2855–2866.

Daguenet E, Baguet A, Degot S, Schmidt U, Alpy F, Wendling C, Spiegelhalter C, Kessler P, Rio MC, Le Hir H, et al. 2012. Perispeckles are major assembly sites for the exon junction core complex. *Mol Biol Cell* **23:** 1765–1782.

Darnell JE Jr. 2013. Reflections on the history of pre-mRNA processing and highlights of current knowledge: A unified picture. *RNA* **19:** 443–460.

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH. 2007. In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14:** 796–806.

de la Mata M, Lafaille C, Kornblihtt AR. 2010. First come, first served revisited: Factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA* **16:** 904–912.

De Santa F, Narang V, Yap ZH, Tusi BK, Burgold T, Austenaa L, Bucci G, Caganova M, Notarbartolo S, Casola S, et al. 2009. Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. *EMBO J* **28:** 3341–3352.

Derman E, Darnell JE. 1974. Relationship of chain transcription to poly (A) addition and processing of hnRNA in HeLa cells. *Cell* **3:** 255–264.

Dias AP, Dufu K, Lei H, Reed R. 2010. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat Commun* **1:** 97.

Escoubet-Lozach L, Benner C, Kaikkonen MU, Lozach J, Heinz S, Spann NJ, Crotti A, Stender J, Ghisletti S, Reichart D, et al. 2011. Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. *PLoS Genet* **7:** e1002401.

Fox-Walsh KL, Hertel KJ. 2009. Splice-site pairing is an intrinsically high fidelity process. *Proc Natl Acad Sci* **106:** 1766–1771.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, et al. 2005. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* **15:** 1451–1455.

Girard C, Will CL, Peng J, Makarov EM, Kastner B, Lemm I, Urlaub H, Hartmuth K, Luhrmann R. 2012. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun* **3:** 994.

Goecks J, Nekrutenko A, Taylor J. 2010. Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11:** R86.

Grunwald D, Singer RH. 2010. *In vivo* imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. *Nature* **467:** 604–607.

Harpold MM, Dobner PR, Evans R, Bancroft FC, Darnell JE Jr. 1979. The synthesis and processing of a nuclear RNA precursor to rat pregrowth hormone messenger RNA. *Nucleic Acids Res* **6:** 3133–3144.

Hatton AR, Subramaniam V, Lopez AJ. 1998. Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol Cell* **2:** 787–796.

Hertel KJ, Maniatis T. 1998. The function of multisite splicing enhancers. *Mol Cell* **1:** 449–455.

Hetru C, Hoffmann JA. 2009. NF-κB in the immune response of *Drosophila*. *Cold Spring Harb Perspect Biol* **1:** a000232.

Hicks MJ, Mueller WF, Shepard PJ, Hertel KJ. 2010. Competing upstream 5′ splice sites enhance the rate of proximal splicing. *Mol Cell Biol* **30:** 1878–1886.

Hocine S, Singer RH, Grunwald D. 2010. RNA processing and export. *Cold Spring Harb Perspect Biol* **2:** a000752.

Hsin JP, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26:** 2119–2137.

Ibrahim EC, Schaal TD, Hertel KJ, Reed R, Maniatis T. 2005. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc Natl Acad Sci* **102:** 5002–5007.

Jensen TH, Patricio K, McCarthy T, Rosbash M. 2001. A block to mRNA nuclear export in *S. cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol Cell* **7:** 887–898.

Kanopka A, Muhlemann O, Akusjarvi G. 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381:** 535–538.

Kessler O, Jiang Y, Chasin LA. 1993. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol Cell Biol* **13:** 6211–6222.

Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT II, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev* **25:** 2502–2512.

Khodor YL, Menet JS, Tolan M, Rosbash M. 2012. Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA* **18:** 2174–2186.

Kim YK, Kim VN. 2007. Processing of intronic microRNAs. *EMBO J* **26:** 775–783.

Kotlajich MV, Crabb TL, Hertel KJ. 2009. Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. *Mol Cell Biol* **29:** 1072–1082.

Lacadie SA, Tardiff DF, Kadener S, Rosbash M. 2006. In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev* **20:** 2055–2066.

LeMaire MF, Thummel CS. 1990. Splicing precedes polyadenylation during *Drosophila E74A* transcription. *Mol Cell Biol* **10:** 6059–6063.

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* **7:** 709–715.

Lim SR, Hertel KJ. 2004. Commitment to splice site pairing coincides with A complex formation. *Mol Cell* **15:** 477–483.

Listerman I, Sapra AK, Neugebauer KM. 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* **13:** 815–822.

Lynch KW. 2004. Consequences of regulated pre-mRNA splicing in the immune system. *Nat Rev Immunol* **4:** 931–940.

Martins SB, Rino J, Carvalho T, Carvalho C, Yoshida M, Klose JM, de Almeida SF, Carmo-Fonseca M. 2011. Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3′ end of human genes. *Nat Struct Mol Biol* **18:** 1115–1123.

Matlin AJ, Moore MJ. 2007. Spliceosome assembly and composition. *Adv Exp Med Biol* **623:** 14–35.

McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL. 1997. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* **11:** 3306–3318.

Nevins JR, Darnell JE Jr. 1978. Steps in the processing of Ad2 mRNA: Poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell* **15:** 1477–1493.

Ninomiya K, Kataoka N, Hagiwara M. 2011. Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. *J Cell Biol* **195:** 27–40.

Osheim YN, Miller OL Jr, Beyer AL. 1985. RNP particles at splice junction sequences on *Drosophila* chorion transcripts. *Cell* **43:** 143–151.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40:** 1413–1415.

Pandya-Jones A. 2011. Pre-mRNA splicing during transcription in the mammalian system. *Wiley Interdiscip Rev RNA* **2:** 700–717.

Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15:** 1896–1908.

Pawlicki JM, Steitz JA. 2008. Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. *J Cell Biol* **182:** 61–76.

Perales R, Bentley D. 2009. "Cotranscriptionality": The transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36:** 178–191.

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

Rabani M, Levin JZ, 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 VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST. 2009. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138:** 114–128.

Rigo F, Martinson HG. 2009. Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. *RNA* **15:** 823–836.

Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F, et al. 2008. THO/Sub2p functions to coordinate 3′-end processing with gene-nuclear pore association. *Cell* **135:** 308–321.

Salditt-Georgieff M, Harpold M, Sawicki S, Nevins J, Darnell JE Jr. 1980. Addition of poly(A) to nuclear RNA occurs soon after RNA synthesis. *J Cell Biol* **86:** 844–848.

Schmid M, Jensen TH. 2010. Nuclear quality control of RNA polymerase II transcripts. *Wiley Interdiscip Rev RNA* **1:** 474–485.

Schmidt U, Basyuk E, Robert MC, Yoshida M, Villemin JP, Auboeuf D, Aitken S, Bertrand E. 2011. Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: Implications for alternative splicing regulation. *J Cell Biol* **193:** 819–829.

Selth LA, Sigurdsson S, Svejstrup JQ. 2010. Transcript elongation by RNA Polymerase II. *Annu Rev Biochem* **79:** 271–293.

Shepard S, McCreary M, Fedorov A. 2009. The peculiarities of large intron splicing in animals. *PLoS One* **4:** e7853.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479:** 74–79.

Shuman S. 2001. Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog Nucleic Acid Res Mol Biol* **66:** 1–40.

Singh J, Padgett RA. 2009. Rates of *in situ* transcription and splicing in large human genes. *Nat Struct Mol Biol* **16:** 1128–1133.

Stewart M. 2010. Nuclear export of mRNA. *Trends Biochem Sci* **35:** 609–617.

Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, et al. 2002.

TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417:** 304–308.

Tennyson CN, Klamut HJ, Worton RG. 1995. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat Genet* **9:** 184–190.

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R. 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22:** 1616–1625.

Tong A, Nguyen J, Lynch KW. 2005. Differential expression of CD45 isoforms is controlled by the combined activity of basal and inducible splicing-regulatory elements in each of the variable exons. *J Biol Chem* **280:** 38297–38304.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.

Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, Schedl P, Tyagi S. 2011. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147:** 1054–1065.

Wahl MC, Will CL, Luhrmann R. 2009. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Weber J, Blanchard JM, Ginsberg H, Darnell JE Jr. 1980. Order of polyadenylic acid addition and splicing events in early adenovirus mRNA formation. *J Virol* **33:** 286–291.

Wetterberg I, Bauren G, Wieslander L. 1996. The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA* **2:** 641–651.

Wuarin J, Schibler U. 1994. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: Evidence for cotranscriptional splicing. *Mol Cell Biol* **14:** 7219–7225.

Wutz A. 2011. Gene silencing in X-chromosome inactivation: Advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12:** 542–553.

Yu Y, Maroney PA, Denker JA, Zhang XH, Dybkov O, Luhrmann R, Jankowsky E, Chasin LA, Nilsen TW. 2008. Dynamic regulation of alternative splicing by silencers that modulate 5′ splice site competition. *Cell* **135:** 1224–1236.

APPENDIX C

Age-Related Gene Expression Differences in Monocytes From Human Neonates, Young Adults, and Older Adults

Abstract

A variety of age-related differences in the innate and adaptive immune systems have been proposed to contribute to the increased susceptibility to infection of human neonates and older adults. The emergence of RNA sequencing (RNA-seq) provides an opportunity to obtain an unbiased, comprehensive, and quantitative view of gene expression differences in defined cell types from different age groups. An examination of *ex vivo* human monocyte responses to lipopolysaccharide stimulation or *Listeria monocytogenes* infection by RNA-seq revealed extensive similarities between neonates, young adults, and older adults, with an unexpectedly small number of genes exhibiting statistically significant age-dependent differences. By examining the differentially induced genes in the context of transcription factor binding motifs and RNA-seq data sets from mutant mouse strains, a previously described deficiency in interferon response factor-3 activity could be implicated in most of the differences between newborns and young adults. Contrary to these observations, older adults exhibited elevated expression of inflammatory genes at baseline, yet the responses following stimulation correlated closely with those observed in younger adults. Notably, major differences in the expression of constitutively expressed genes were not observed, suggesting that the age-related differences are driven by environmental influences rather than cell-autonomous differences in monocyte development.

Introduction

Age-related differences in clinical susceptibility to infection have been extensively documented, with diminished protective responses and enhanced susceptibility observed in pre-term and term infants, as well as in older adults when compared to young adults [1-5]. This clinical observation of an age-dependent risk for infectious morbidity and mortality has led to an interest in identifying the underlying mechanisms and deriving strategies to enhance protective immune responses at the extreme ends of life [1-3].

Differences in innate immune responses are thought to contribute to the overall susceptibility observed in neonates and older adults [2,6]. Neonates have been reported to produce lower levels of effector molecules, such as oxygen radicals [2,7]. A number of other proteins have also been reported at reduced levels in innate immune cells, including reduced expression of IFNα, CD40, CD80, CD83, and CD86 in neonatal plasmacytoid dendritic cells [5]. Furthermore, newborns and older adults produce altered levels of cytokines that regulate the development of adaptive immunity (reviewed in [2]). For example, the heterodimeric, Th1-inducing innate cytokine, interleukin(IL)-12, is expressed at reduced levels in neonates, due to the reduced expression of its p35 subunit [8-10]. In contrast, the anti-inflammatory cytokine, IL-10, and the Th17-inducing cytokines, IL-6 and IL-23, have been observed at elevated levels in neonates [9,11]. In older adults, a variety of innate effector responses appear to be reduced, including superoxide generation and the phagocytosis of microorganisms [12,13]. Systemic low-level inflammation is another common characteristic of older adults that may alter their response to infection (reviewed in [2]).

The approaches used to identify age-dependent differences that lead to an increased risk to suffer from infection at the extreme ends of life have been largely balkanized and focused on a few particular components, the choice of which appears to depend on the expertise of a given group of investigators. What has been missing is an unbiased yet comprehensive interrogation of the events that occur in the very young and the very old following recognition of an infectious threat. In addition to our deficiency in knowledge of age-dependent differences in the immune system, little is known about the molecular mechanisms responsible for these differences. Reduced activation of transcription factors such as interferon response factor 3 (IRF3), defects in nucleosome remodeling, and differences in the expression of pattern recognition receptors and signaling molecules (e.g. MyD88) are among the mechanisms that have been proposed to be responsible for the diminished innate immune responses observed in neonates [2,14-16].

Age-dependent differences in hematopoietic stem cells and in the development of hematopoietic lineages have also been observed, providing one possible explanation for the immune response differences [17-19]. According to this scenario, myeloid cell types may be fundamentally different in neonates, adults, and older adults, resulting in different gene expression responses following stimulation or infection. As an alternative, the myeloid cell populations may be similar, but age-related differences in the blood or tissue microenvironment may lead to different responses [20]. The response differences may be lost when cells from different age groups are cultured under the same conditions, or they may be retained via epigenetic mechanisms or other memory mechanisms [3].

DNA microarrays were previously used to obtain genome-scale insight into age-dependent differences in gene expression following infectious exposure [15]. More recently, RNA sequencing (RNA-seq) has emerged as a more quantitative method for examining transcriptomes [21]. The availability of the RNA-seq method provides an opportunity to unravel, with greater precision, the age-dependent differences in the immune system that increase risk for a serious outcome following infection. As a first step, the identification of age-related differences in gene expression following *ex vivo* infectious exposure of defined cell populations, along with the identification of differences in constitutive gene expression in these populations, would be of considerable value.

In this study, RNA-seq was used to compare the gene expression responses to LPS stimulation or *Listeria monocytogenes* (*Lm*) infection in cord blood monocytes and in peripheral blood monocytes from young and older adults. LPS provides an example of a well-defined innate immune stimulator; *Lm* causes suffering and dying in the very old and the very young, while most young adults rarely even display symptoms if infected [22]. Our data reveal extensive similarities in constitutive gene expression and in the response to stimulation or infection in monocytes from the three age groups. Furthermore, most of the differences identified between neonates and young adults could be connected to the previously reported reduction in IRF3 activity in neonates [15]. In contrast, most differences between young adults

and older adults appeared to result from a low-level inflammatory state ('inflammaging') that characterized monocytes from older adults. Interestingly, large differences in the expression of constitutively expressed genes, which would be expected if blood monocytes from neonates, adults, and older adults were fundamentally different, were not identified. This finding supports a hypothesis in which age-related environmental differences are responsible for the inability of neonatal monocytes to mount a robust IRF3-mediated response.

Materials and Methods

Isolation of cells and stimulation conditions

All studies were approved by the Institutional Ethics Review Board at the University of British Columbia. Samples of cord blood from healthy, full-term elective Caesarean sections without labor and samples of healthy young adult and older adult peripheral blood were collected directly into sodium heparin-containing vacutainers (BD Biosciences). Within two hrs of the blood draw, mononuclear cells were isolated by density gradient centrifugation [11]. Positive selection of monocytes from mononuclear cells was then carried out using Miltenyi microbeads according to the manufacturer's protocol with some revisions. Briefly, mononuclear cells were incubated with 800 uL MACS buffer and 200 uL anti-human CD14 microbeads at 4^{o}C. Cells were then washed with MACS buffer prior to positive selection of monocytes using Miltenyi selection columns. Purified monocytes from each donor were cultured in RPMI 1640 medium supplemented with Glutamax (Gibco, Life Technologies) and 10% human AB serum (Gemini Bio Products). The monocytes were counted and plated onto 96 well plates at a density of 1x10^{6} cells/well. Monocytes were stimulated with LPS (10 ng/ml) (InvivoGen) or infected with *Lm* at MOI=5. Wild-type (WT) *Lm* strain 10403s was provided by Dr. D. Portnoy (University of California, Berkeley, CA) and grown as described [23].

Mouse macrophages were prepared from the bone marrow of 6-week-old C57BL/6, IRF3$^{-/-}$, or IFNAR$^{-/-}$ mice as described [24,25], and were stimulated with lipid A (100 ng/mL) (Sigma) after 6 days of differentiation.

RNA isolation, library preparation, and sequencing

Human monocyte RNA was purified using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Strand-specific libraries were prepared using 120 ng RNA input according to the "deoxyuridine triphosphate (dUTP)" method [26]. Mouse macrophage experiments involved analyses of chromatin-associated RNAs, as previously described [25]. A HiSeq 2000 (Illumina) was used for sequencing, with a single end sequencing length of 50 nucleotides. Sequencing data have been submitted to GEO under accession number GSE60216.

Bioinformatic analyses

All bioinformatic analyses were conducted using the Galaxy platform [27]. Reads were aligned to the human GRCh37 or mouse mm9 reference genomes with Tophat [28] using most default parameters. Alignments were restricted to uniquely mapping reads with two possible mismatches permitted. RPKM (reads per kilobase pair per million mapped reads) were calculated using Seqmonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Coexpressed gene classes were evaluated with Cluster3 by applying k-means clustering to mean-centered log2(RPKM) expression values [29]. Statistically significant gene expression differences were evaluated using DESeq [30]. Mouse orthologs of human genes were identified using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Pscan was used to detect DNA motifs overrepresented in each class between nucleotides -450 and +50 relative to the transcription start site [31].

Results

Gene expression cascades induced in monocytes by LPS and *Lm*

An attractive starting point toward a full understanding of age-related differences in immune responses is to employ RNA-seq to carefully examine mRNA transcript levels following stimulation or infection of defined cell types. Toward this goal, peripheral blood monocytes were

obtained from healthy young adults between the ages of 19 and 45, and healthy older adults aged 65 years and older. In addition, neonatal monocytes were obtained from umbilical cord blood samples. The monocytes were stimulated with LPS or infected with *Lm* immediately after isolation to avoid alterations in cell properties caused by culturing. For both the LPS and *Lm* experiments, three individuals from each age group were analyzed. For the LPS experiments, samples for RNA-seq were secured at 0, 1, and 6 hrs post-stimulation. For the experiments involving live *Lm* infection, samples were collected for RNA-seq 2 and 6 hrs after infection; in this experiment, uninfected cells (referred to as 0-hr time point) were collected after culturing without *Lm* for 2 hrs, whereas the unstimulated cells in the LPS experiment were collected immediately after isolation. After mRNA isolation and cDNA library preparation, RNA-seq was performed. The number of mapped reads ranged from 3.4×10^6 to 1.3×10^7 per sample.

An examination of the data sets from the LPS experiment identified 1147 annotated RefSeq genes that were induced by at least five-fold at the 1- or 6-hr time point (relative to the unstimulated sample) in at least one sample from any age group, and that exhibited a transcript level exceeding four RPKM following induction. To examine the relationship between the different time points and age groups in the response to LPS, hierarchical clustering was performed with these 1147 genes (Figure C-1A). This analysis revealed that each of the nine samples from a given time point was more closely related to the other samples from the same time point than to any sample from the other two time-points. The most significant difference that showed a possible relationship to age was that the three unstimulated samples from older adults (OA1.0, OA2.0, and OA3.0) and one young adult unstimulated sample (A1.0) clustered separately from the remaining unstimulated samples from young adults and neonates.

Small age-related differences were also observed with the 6-hr time-point data, in that, with only one exception (neonatal sample N3.6), each age group clustered separately from the others. In contrast, the nine 1-hr time-point samples correlated closely, with no apparent age-related differences. The Pearson correlation values (R values) used for the hierarchical clustering are shown in Figure C-1B. These results provide initial evidence that the vast majority

of LPS-induced genes are induced similarly in the three age groups.

Examination of the *Lm* data sets identified 865 annotated RefSeq genes that were induced by at least five-fold at the 2-hr or 6-hr time point in at least one sample, and that exhibited a transcript level exceeding four RPKM following induction. The hierarchical clustering results and the Pearson correlation values revealed even stronger correlations between age groups at each time point than were observed with the LPS data (Figure C-2). That is, although strong time-dependent clustering was observed, no consistent age-related differences were observed at any of the time points.

K-means cluster analysis of LPS- and *Lm*-induced genes

To extend the analysis of age-related differences in inducible gene expression, k-means clustering was used to define groups of genes that exhibited similar expression patterns among the three age groups and three time points. The k-means algorithm considers induction kinetics, induction magnitudes, and differences among age groups. Figure C-3A shows the results obtained when the 1147 LPS-induced genes (using the average expression values from the three independent samples analyzed for each age group and each time point) were assigned to one of ten distinct clusters. As expected on the basis of the hierarchical clustering, extensive similarities were apparent in the three age groups in almost all of the clusters. The similarities are also apparent in line graphs showing the average relative expression levels for all genes in a given cluster (Figure C-3B).

Only one cluster (Cluster I) was identified that showed substantial age-related differences (Figure C-3A,B). Genes in this cluster were generally expressed at a lower level in both unstimulated and LPS-stimulated monocytes from neonates in comparison to the young adult and older adult samples. Although the average induction magnitude for genes in this cluster was comparable among the age groups, the average expression level of these genes was significantly lower in neonates than in young adults at all three time points.

K-means clustering of the *Lm*-induced genes also revealed extensive similarities among

the three age groups (Figure C-4). Only one cluster (Cluster G) showed slightly reduced average expression in the neonatal and older adult samples in comparison to the young adult samples.

Analysis of genes exhibiting statistically significant expression differences

Because the clustering results described above revealed extensive similarities with limited age-related differences, we envisioned that meaningful insights would require the use of defined parameters to identify genes that exhibited the greatest differential expression. Toward this end, we first focused our attention on genes induced to a statistically significant extent ($p<0.01$) that also exhibited differential expression between neonates and young adults at a high level of statistical significance ($p<0.01$). Only 118 of the 1147 LPS-induced genes met these criteria.

The 118 genes (gene identities listed in Figure C-6) were separated into groups according to the time point at which their maximum mRNA level was observed (Figure C-5A: 1-hr peak expression for Groups I and II; 6-hr peak expression for Groups III-VI). The genes were then further grouped according to their expression level in neonates relative to their expression level in young adults (Figure C-5A, column 7). (For this calculation, the baseline and maximum expression levels in young adults were defined as 0% and 100%, respectively; the maximum expression level in neonates was then determined as a percentage relative to that range.) This analysis revealed 35 genes that exhibited enhanced expression in the neonatal samples (Groups I and III, lightest shade of purple) and 83 genes that exhibited reduced expression (Groups II, IV, V, and VI, three darker shades of purple). Group VI contains the 34 genes that exhibited the greatest difference between neonates and young adults. For these genes, the maximum LPS-induced mRNA level in neonates was less than 20% of the maximum level observed in young adults.

A parallel analysis with the *Lm* samples identified 123 genes (listed in Figure C-8) that were inducible and differentially expressed between neonates and young adults with a high level of statistical significance ($p<0.01$ for both induction and differential expression). Grouping

of these genes using the same strategy as above revealed 13 genes that were expressed more highly in neonates than young adults (Figure C-7A, Groups I and V) and 110 genes that were expressed more highly in young adults than neonates (Groups II-IV and VI-VIII). Forty-seven of these later genes exhibited mRNA levels in neonates that were less than 20% of the young adult levels (Groups IV and VIII).

A prominent role for IRF3 and Type I IFN signaling in the neonate-adult differences

To gain insight into the mechanisms responsible for differential gene expression in neonatal and young adult monocytes, we first examined the requirements for expression of the mouse orthologs of the differentially expressed genes. This analysis took advantage of a large number of RNA-seq data sets that have been generated in our laboratory using mouse bone marrow-derived macrophages stimulated with the Lipid A component of LPS. This collection of data sets includes kinetic analyses of lipid A-induced gene expression in macrophages from a variety of mutant mouse strains lacking key signaling molecules or transcription factors thought to be important for inducible transcription ([25] and unpublished results).

By examining the expression requirements for the mouse orthologs of the genes that were differentially expressed in human neonates and young adults, evidence was obtained that these genes frequently require the transcription factor IRF3 or Type I IFN receptor signaling. That is, many of the age-dependent differentially expressed genes were expressed at substantially reduced levels in *Irf3*$^{-/-}$ and/or *Ifnar*$^{-/-}$ macrophages stimulated with Lipid A.

To document the extent to which IRF3 and IFNAR signaling might contribute to the differential expression of LPS-induced genes in neonates and adults, human genes for which mouse orthologs could clearly be identified (114 of 118 genes; Figure C-5A, column 8, dark pink and red) were first separated from the small number of genes lacking obvious mouse orthologs (Figure C-5A, column 8, lightest pink). Then, the RNA-seq data sets were analyzed to identify genes that were both expressed (RPKM > 4 when maximally expressed) and induced (>5-fold) in both the human monocytes and wild-type mouse macrophages. The 38 genes that met these

criteria (Figure C-5A, column 8, red) were then evaluated for their dependence on IRF3 and IFNAR in mouse bone marrow-derived macrophages stimulated with Lipid A. The results revealed IRF3-dependence for 14 of the 16 genes in Group VI (Figure C-5A, column 9, dark blue if <10% of the wild-type expression level in *Irf3*$^{-/-}$ macrophages and light blue if 10-33% of the wild-type level in *Irf3*$^{-/-}$ macrophages). 14 of the 16 genes also exhibited reduced expression in *Ifnar*$^{-/-}$ macrophages (column 10). IRF3- and/or IFNAR-dependence was also observed for most Group V genes for which mouse orthologs were both expressed and induced in mouse macrophages (Figure C-5A).

As an independent strategy, a transcription factor binding site motif analysis was performed using the Pscan program [31] with the promoter regions of all genes in Groups I through VI. The goal of this analysis was to identify transcription factors whose binding sites are over-represented in the promoters of specific clusters of genes. The small number of transcription factors for which significant enrichment was observed are shown in Figure C-5B. Transcription factor binding motif enrichment generally was not observed for Groups I through V. However, highly significant enrichment of binding sites for IRF1, IRF2, STAT1, and a STAT2:STAT1 heterodimer was found at the promoters of Group VI genes (Figure C-5B). The IRF1 and IRF2 binding sites used by the Pscan program are similar to the experimentally defined consensus IRF3 binding motif [32], which is not assessed by Pscan. Importantly, IRF and STAT motifs were identified in the promoters of the vast majority of Group VI genes, including most genes whose mouse orthologs could not be examined for IRF3 and IFNAR dependence due to lack of inducible expression in both mice and humans (Figure C-5A, column 11).

Thus, both the functional analysis and motif analysis strongly support the hypothesis that reduced activation of IRF3- and IFNAR-dependent genes explains most gene expression differences between neonatal and adult monocytes. It is noteworthy that a previous study which documented reduced IRF3 activity in neonatal dendritic cells found that neonatal and adult cells were similarly responsive to IFNβ stimulation, suggesting that the reduced expression of IFNAR-

dependent genes is due to reduced IRF3 activity (resulting in reduced IFNβ expression) rather than a reduction in IFNAR signaling [15].

Consistent with the analysis of the LPS-induced genes, mouse orthologs of the human genes that exhibited differential expression upon *Lm* infection were generally found to exhibit IRF3- and/or IFNAR-dependence (Figure C-7A). Furthermore, binding sites for IRF1, IRF2, STAT1, and the STAT2:STAT1 heterodimer were greatly enriched in the Group VIII genes and to a lesser extent in Group VII genes (Figures 6A and 6B). Thus, although IRF3 is thought to be activated by different pathways in LPS-stimulated and *Lm*-infected cells [33,34], a common reduction in IRF3 activity is likely to be responsible for the strongest gene expression differences between neonatal and adult monocytes.

Low-level inflammation in older adults

To evaluate gene expression differences between young adults and older adults, we first used the strategy described above to identify differentially induced genes. This analysis revealed minimal differences in transcriptional induction (data not shown), suggesting that the pathways involved in the responses to LPS and *Lm* in monocytes from the two age groups are highly similar. Instead, the largest differences were observed when examining transcript levels for inducible genes prior to stimulation. Specifically, 189 LPS-induced genes (>5-fold induction magnitude; induction significance $p<0.01$; maximum induced transcript level >4 RPKM) exhibited transcript levels that were significantly different ($p<0.01$) in unstimulated cells from young adults in comparison to older adults (Figure C-9A; gene list in Figure C-10). For these 189 genes, Figure C-9A, column 7 shows the ratio of the unstimulated transcript level in older adults to that in younger adults (OA0/A0). In this figure, the genes are grouped on the basis of their time point of maximum expression, and the genes were then rank-ordered by the ratio of the unstimulated transcript level. This analysis revealed that a large majority of the differentially expressed genes are expressed at an elevated level in older adults (shown as shades of red, see vertical color scale at right). In fact, 52% of the differentially expressed genes exhibited

unstimulated transcript levels in older adults that were at least 3-fold higher than in young adults, whereas only 3% exhibited transcript levels that were at least 3-fold higher in young adults than in older adults. Similar results were observed in the *Lm* experiment (data not shown), but the number of genes showing differential expression was lower, probably because the unstimulated cells for the *Lm* experiment were cultured for 2 hrs prior to collection, whereas the unstimulated cells in the LPS experiment were collected without culturing.

Importantly, although relatively large differences in expression between young adults and older adults were observed in the unstimulated cells, the magnitudes of the differences were generally lower after stimulation. This is apparent in Figure C-9A, column 8 (max OA/max A), which shows the ratio between the maximum induced transcript levels in older adults versus young adults. Because the same color scale is used for columns 7 and 8, it is readily apparent that the transcript level ratios move toward 1 after stimulation for most genes that are differentially expressed prior to stimulation. Figure C-9B, which displays average transcript levels for all genes in Groups I and II, also shows that transcript levels in older adults were elevated to a greater extent prior to stimulation than after stimulation. Thus, an inflammatory state is readily apparent in unstimulated monocytes from older adults. This inflammatory state in unstimulated cells may influence transcript levels observed after stimulation or infection, but to a limited extent relative to the differences observed in the basal state.

Discussion

The diminished capacity of human neonates and older adults to mount an immune response to infectious agents has been well documented [1,2]. However, because of the complexity of the human immune system and limitations in the experimental approaches that are available for studying immune responses in humans, insights into the underlying mechanisms have been difficult to obtain. One starting point toward a mechanistic understanding can be characterized as reductionist, in which the goal is to first delineate age-related differences intrinsic to defined

immune cell types in an *ex vivo* setting, with subsequent experiments focusing on how these intrinsic differences contribute to clinical observations in the far more complex *in vivo* setting.

In this study, RNA-seq was used to examine the intrinsic response of blood monocytes to LPS stimulation and *Lm* infection. The improved dynamic range of the RNA-seq method in comparison to microarray methods [21] led to the expectation that the results might reveal extensive differences among the age groups. Given this expectation, the most striking finding is perhaps the extensive similarity in both constitutive and inducible gene expression. The results suggest that a single mechanism – variable induction of IRF3 – may be responsible for most and perhaps all differences between neonatal and young adult monocytes. Another defined mechanism, variable low-level inflammation prior to induction, may explain most of the differences between young adults and older adults.

Our results strengthen previous evidence that reduced IRF3 activity makes a major contribution to the deficient innate responses of neonates to infectious stimuli [15]. The previous study was performed with LPS-stimulated dendritic cells differentiated from cord blood or adult peripheral blood, whereas the current study was performed with freshly isolated monocytes stimulated with LPS or infected with *Lm*. In the previous study, a large number of IRF3- and Type 1 IFN-dependent genes were found to be expressed at reduced levels in neonates. The reduced expression of these genes was attributed to reduced IRF3 activity because the neonatal and adult cells responded similarly to direct stimulation with IFNβ. Reduced IRF3 activity would lead to a broad reduction in the expression of IFN-dependent genes because IRF3 is critical for the initial induction of *IFNB* transcription in LPS-stimulated cells.

Interestingly, the previous study found that IRF3 translocated to the nucleus similarly in neonatal and adult cells, and its *in vitro* DNA-binding activity was similarly induced [15]. However, its ability to bind endogenous target genes was reduced, suggesting that an additional event – possibly an additional post-translational modification – is needed for binding to target genes and may be reduced in neonatal cells. Of relevance, a separate study identified a major defect in IRF7 activation in neonatal plasmacytoid dendritic cells and, in this cell type, a defect

in nuclear translocation of IRF7 was observed in neonates [35]. An additional clue into the underlying mechanism is our finding of a similar deficiency in both LPS-stimulated and *Lm*-infected cells. LPS and *Lm* activate IRF3 via different signaling pathways – the TRIF pathway for LPS and the STING pathway for *Lm* [33,34] – suggesting that the reduced IRF3 activity in neonatal cells involves a mechanism that influences the activation of IRF3-dependent genes via both of these pathways.

In addition to elucidating the specific mechanism, it will also be important to understand why this difference exists between neonatal and adult monocytes. The simplest model is that neonatal monocytes are fundamentally different from adult monocytes and represent a developmentally distinct monocyte subtype. However, this model predicts that prominent gene expression differences would be observed prior to stimulation. The differentially expressed genes would be expected to include cell-surface markers that define different myeloid cell populations and genes that might help regulate IFN responses. Surprisingly, the expression profiles of the unstimulated monocytes from neonates and adults were remarkably similar (data not shown), with no large differences suggesting that they represent different myeloid subtypes, and no differences that would be predictive of the differential induction of IRF3-dependent genes.

One possible explanation for this apparent paradox is that the differences between neonatal and adult monocytes are due to the differential expression of micro-RNAs or long noncoding RNAs, which were not examined in this analysis. However, the differential expression or processing of non-coding RNAs would be expected to require the differential expression of transcription factors that regulate the non-coding RNA genes, or the differential expression of processing enzymes; these protein-coding genes would have been included in our analysis. Differences in alternative pre-mRNA splicing also were not examined in our analysis. Once again, differential splicing would be expected to require the differential expression of genes encoding splicing factors. A more likely possibility is that the pronounced

difference in the induction of IRF3-dependent genes is regulated by genes whose expression levels vary by only a small and statistically insignificant amount.

Because the RNA-seq profiles failed to provide evidence that the neonatal and adult cells represent developmentally distinct monocyte subtypes, the neonatal-adult differences may instead be due to environmental differences that act on the fully differentiated cells to influence their capacity to induce IRF3 activity. Such a mechanism would need to influence IRF3's capacity for induction for a prolonged time period, because the IRF3 difference has been observed in dendritic cells differentiated for several days *in vitro* [15]. This environmental difference may lead to small and stable differences in the expression of genes that regulate IRF3 activity. Alternatively, the neonatal microenvironment may alter the structure of chromatin at IRF3-dependent genes, resulting in a reduced capacity for IRF3 binding in response to a stimulus.

To summarize, the results of this study will help guide future efforts to understand the mechanisms responsible for the immune deficiencies observed in neonates and older adults. The results suggest that the intrinsic properties of blood monocytes are remarkably stable throughout life and vary to only a limited extent. The reduced capacity of neonatal monocytes to activate IRF3-dependent genes could play an important role in the deficient response of neonates to many microbial pathogens. Furthermore, the low-level inflammation that is readily apparent in monocytes from older adults could also influence anti-microbial responses. RNA-seq studies to quantitatively characterize intrinsic age-related differences in other innate and adaptive immune cell types should provide additional insights and should ultimately suggest strategies to enhance immune responses in deficient populations.

Acknowledgments

The authors are grateful to the Broad Stem Cell Research Center at UCLA for high-throughput sequencing. We also thank Stanislas Goriely, Ranjan Sen, and Philip Scumpia for helpful discussions.

Author Contributions

Conceived and designed the experiments: MML BJT KW TRK STS. Performed the experiments: MML BJT KW AJT. Analyzed the data: MML BJT AJT. Contributed to the writing of the manuscript: MML BJT KW TRK STS.

Figure Legends

Figure C-1. Hierarchical clustering of LPS-stimulated monocyte transcriptomes from human neonates, adults, and older adults. (A) RNA-seq experiments were performed with three independent human monocyte samples from cord blood (N), young adult peripheral blood (A), and older adult peripheral blood (OA) stimulated with LPS for 0, 1, and 6 hrs. Hierarchical clustering was performed with the 1147 genes found to be induced by at least 5-fold at the 1- or 6-hr time point in at least one sample and with an induced RPKM of at least 4 (genes smaller than 200 bp were also excluded from the analysis). Sample codes correspond to the age abbreviation followed by the sample number (1 through 3 for each age); the time point (0, 1, or 6 hr) is indicated after the period. Inducible transcriptomes exhibit strong time-dependent clustering, with much less age-dependent clustering. (B) Pearson correlation values (R) used for the hierarchical clustering in panel A are shown. Each time point from each sample was compared to every other sample and time point. R values are color-coded from low (green) to high (red). Samples on the X and Y axes are grouped first according to age group, then time point (0, 1, or 6), and then sample number (1-3).

Figure C-2. Hierarchical clustering of *Lm*-infected monocyte transcriptomes from human neonates, adults, and older adults. (A) RNA-seq experiments were performed with three independent human monocyte samples from cord blood (N), young adult peripheral blood (A), and older adult peripheral blood (OA) infected with *Lm* for 0, 2, and 6 hrs. Hierarchical clustering was performed with the 865 genes found to be induced by at least 5-fold at the 2- or 6-hr time point in at least one sample and with an induced RPKM of at least 4 (genes smaller than 200 bp were also excluded from the analysis). Sample codes correspond to the age abbreviation followed by the sample number (1 through 3 for each age); the time point (0, 2, or 6 hr) is indicated after the period. Inducible transcriptomes exhibit strong time-dependent clustering, with much less age-dependent clustering. (B) Pearson correlation values (R) used for the hierarchical clustering in panel A are shown. Each time point from each sample was compared

to every other sample and time point. R values are color-coded from low (green) to high (red). Samples on the X and Y axes are grouped first according to age group, then time point (0, 2, or 6), and then sample number (1-3).

Figure C-3. Analysis of LPS-induced genes in monocytes by K-means cluster analysis. (A) The 1147 genes that exceeded 200 bp in length, exhibited an RPKM of at least 4 in one sample, and were induced by LPS by at least 5-fold in the same sample were divided into 10 clusters by k-means cluster analysis, which considers similarities in transcript levels for each gene across all 27 samples (3 age groups, 3 samples for each age group, and 3 time points for each sample). The three independent samples are shown in parallel for each age group. Colors indicate the percentile of the relative expression level (based on the log-transformed mean-centered RPKM for each gene), as indicated at the bottom. (B) The average relative transcript levels for genes within each cluster are shown for each age group (neonates, blue diamonds; young adults, red squares; older adults, green triangles).

Figure C-4. Analysis of *Lm*-induced genes in monocytes by K-means cluster analysis. (A) The 865 genes that exceeded 200 bp in length, exhibited an RPKM of at least 4 in one sample, and were induced by *Lm* infection by at least 5-fold in the same sample were divided into 10 clusters by k-means cluster analysis, which considers similarities in transcript levels for each gene across all 27 samples (3 age groups, 3 samples for each age group, and 3 time points for each sample). The three independent samples are shown in parallel for each age group. Colors indicate the percentile of the relative expression level (based on the log-transformed mean-centered RPKM for each gene), as indicated at the bottom. (B) The average relative transcript levels for genes within each cluster and are shown for each age group (neonates, blue diamonds; young adults, red squares; older adults, green triangles).

Figure C-5. Genes that exhibit the greatest expression deficit in LPS-stimulated cord blood monocytes in comparison to adult monocytes are regulated by IRF3 and/or Type I IFNs. (A) LPS-induced genes exhibiting statistically significant differential expression in neonates and adults (n = 118) were grouped according to the time point at which their maximum transcript levels were observed (1 or 6 hrs). They were then grouped according to their relative maximum transcript levels in cord blood (neonates) versus young adults. Induced genes with a higher maximum transcript level in neonates than young adults (with statistically significant differential expression) are included in classes I (1-hr peak) and III (6-hr peak) (column 7). Genes exhibiting a maximum transcript level in neonates that was 50-100% of the young adult transcript level (but with statistically significant differential expression) are included in class IV (no genes with peak transcript levels at 1-hr fit this criterion). Genes exhibiting a maximum transcript level in neonates that was 20-50% of the young adult transcript level are in classes II (1-hr) and V (6-hr). Genes with a maximum transcript level in neonates below 20% of the young adult transcript level are in class VI. Columns 1-6 show the relative transcript levels (based on the log-transformed mean-centered RPKM) for these 118 classified genes in all samples and all time points from both neonates and young adults. Column 8 indicates genes that lack obvious mouse orthologs (lightest pink), genes that contain mouse orthologs that are either not expressed or not induced in mouse bone marrow-derived macrophages (dark pink), and genes containing mouse orthologs that are both expressed and induced by LPS (red). Columns 9 and 10 show relative expression of the mouse ortholog of the human gene in Lipid A-stimulated macrophages from $IRF3^{-/-}$ and $IFNAR^{-/-}$ mice, respectively (see blue scale at right). Note that these columns are only relevant for genes shown in red in Column 8. Column 11 indicates genes with promoters that contain an IRF1 transcription factor binding motif between -450 and +50 bps relative to the transcription start site. (B) Enrichment of transcription factor binding sites determined using the Pscan program is shown for each gene class from panel A. Color intensity is proportional to the negative log(p-value).

Figure C-6. LPS-induced genes exhibiting statistically significant differences in transcript levels in cord blood and young adult monocytes. An expanded version of Figure C-5A is shown, which includes the identities of the LPS-induced genes that are differentially expressed in cord blood and young adult monocytes. RefSeq IDs and gene names are shown for human genes and their mouse orthologs.

Figure C-7. Genes that exhibit the greatest expression deficit in *Lm*-stimulated cord blood monocytes in comparison to adult monocytes are regulated by IRF3 and/or Type I IFNs. (A) *Lm*-induced genes exhibiting statistically significant differential expression in neonates and young adults (n = 123) were grouped according to the time point at which their maximum transcript levels were observed (2 or 6 hrs). They were then grouped according to their relative maximum transcript levels in cord blood (neonates) versus young adults. Induced genes with a higher maximum transcript level in neonates than young adults (with statistically significant differential expression) are included in classes I (2-hr peak) and V (6-hr peak) (column 7). Genes exhibiting a maximum transcript level in neonates that was 50-100% of the young adult transcript level (but with statistically significant differential expression) are included in classes II (2-hr) and VI (6-hr). Genes exhibiting a maximum transcript level in neonates that was 20-50% of the young adult transcript level are in classes III (2-hr) and VII (6-hr). Genes with a maximum transcript level in neonates below 20% of the young adult transcript level are in classes IV (2-hr) and VIII (6-hr). Columns 1-6 show the relative transcript levels (based on the log-transformed mean-centered RPKM) for these 123 classified genes in all samples and all time points from both neonates and young adults. Column 8 indicates genes that lack obvious mouse orthologs (lightest pink), genes that contain mouse orthologs that are either not expressed or not induced in mouse bone marrow-derived macrophages (dark pink), and genes containing mouse orthologs that are both expressed and induced by LPS (red). Columns 9 and 10 show relative expression of the mouse ortholog of the human gene in Lipid A-stimulated macrophages from $IRF3^{-/-}$ and $IFNAR^{-/-}$ mice, respectively (see blue scale at right). Note that these columns are

only relevant for genes shown in red in Column 8. Column 11 indicates genes with promoters that contain an IRF1 transcription factor binding motif between -450 and +50 bps relative to the transcription start site. (B) Enrichment of transcription factor binding sites determined using the Pscan program is shown for each gene class from panel A. Color intensity is proportional to the negative log(p-value).

Figure C-8. *Lm*-induced genes exhibiting statistically significant differences in transcript levels in cord blood and young adult monocytes. An expanded version of Figure C-7 is shown, which includes the identities of the *Lm*-induced genes that are differentially expressed in cord blood and young adult monocytes. RefSeq IDs and gene names are shown for human genes and their mouse orthologs.

Figure C-9. Elevated expression of a broad range of inflammatory genes prior to stimulation of freshly isolated monocytes from older adults. (A) LPS-induced genes exhibiting differential basal expression between adults and older adults (n = 189) are grouped according to maximum mRNA level. Columns 7 and 8 show the ratio of transcript levels between older adults and young adults before stimulation and at maximum transcript levels, respectively. (B) The average relative transcript levels within each cluster and for each age are shown.

Figure C-10. LPS-induced genes that exhibit statistically significant differences in basal transcript levels in monocytes from young and older adults. An expanded version of Figure C-9A is shown, which includes the identities of LPS-induced genes that are differentially expressed in unstimulated young and older adult monocytes. Human RefSeq IDs and gene names are shown.

Figure C-1. Hierarchical clustering of LPS-stimulated monocyte transcriptomes from human neonates, adults, and older adults.

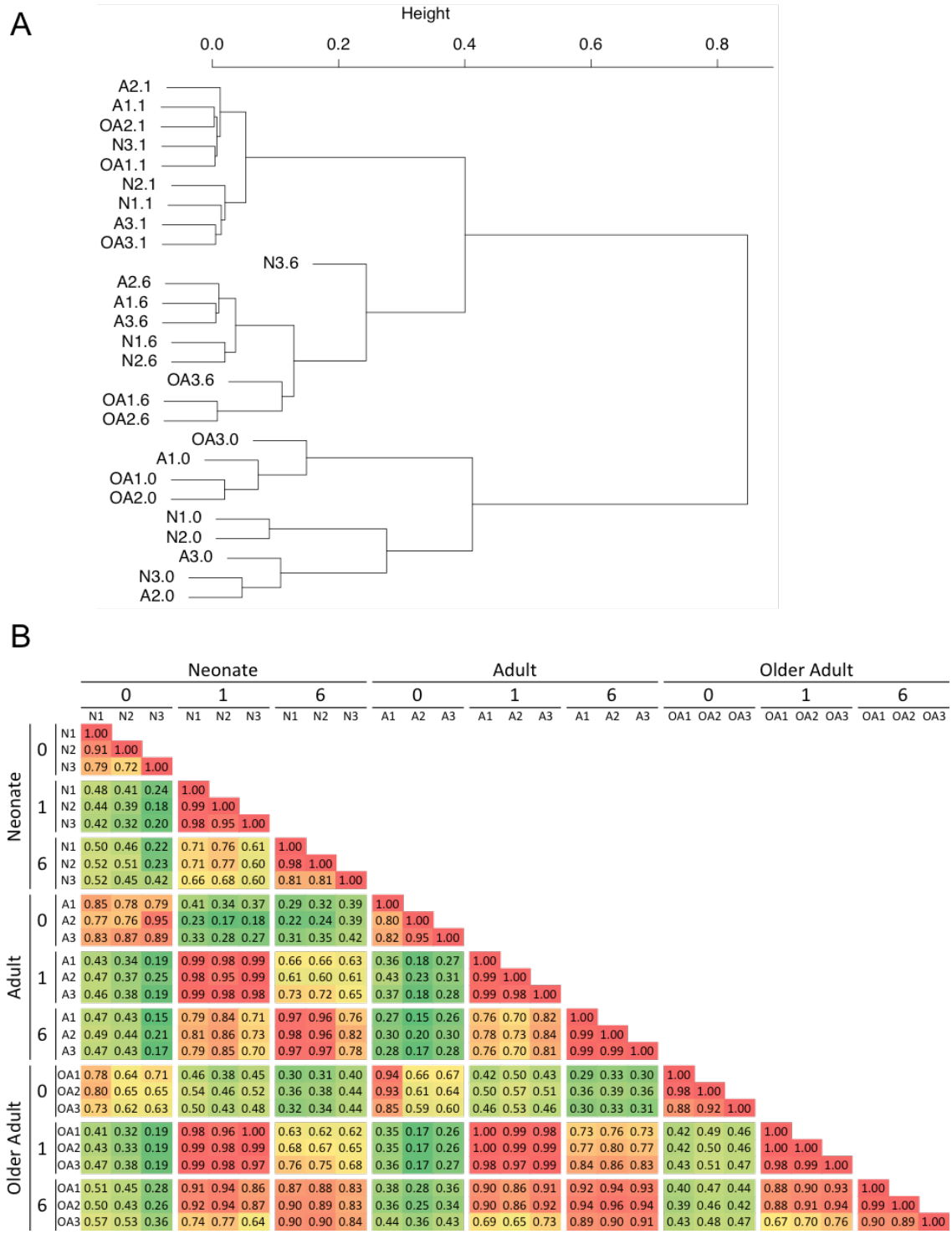


Figure C-2. Hierarchical clustering of *Lm*-infected monocyte transcriptomes from human neonates, adults, and older adults.

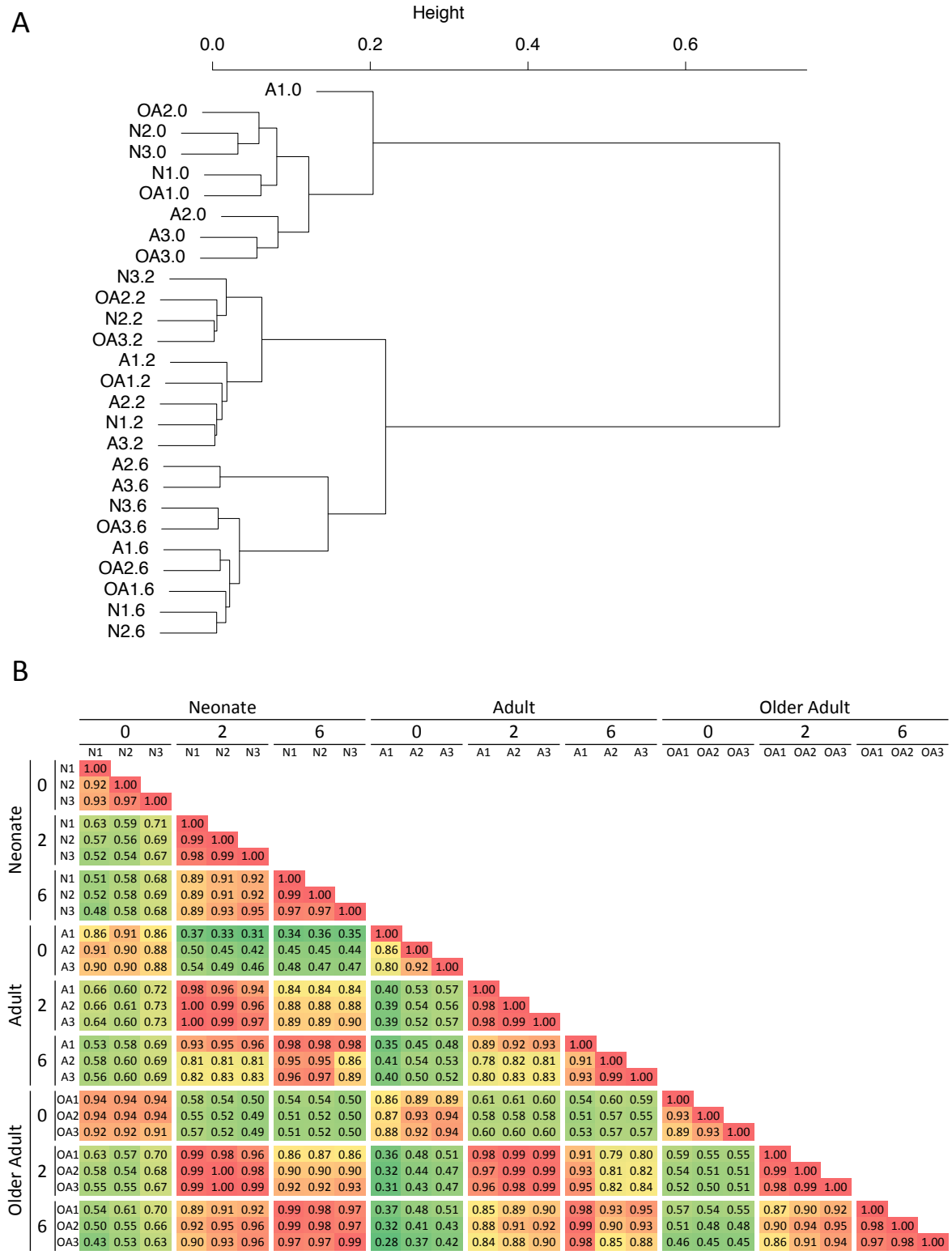


Figure C-3. Analysis of LPS-induced genes in monocytes by K-means cluster analysis.

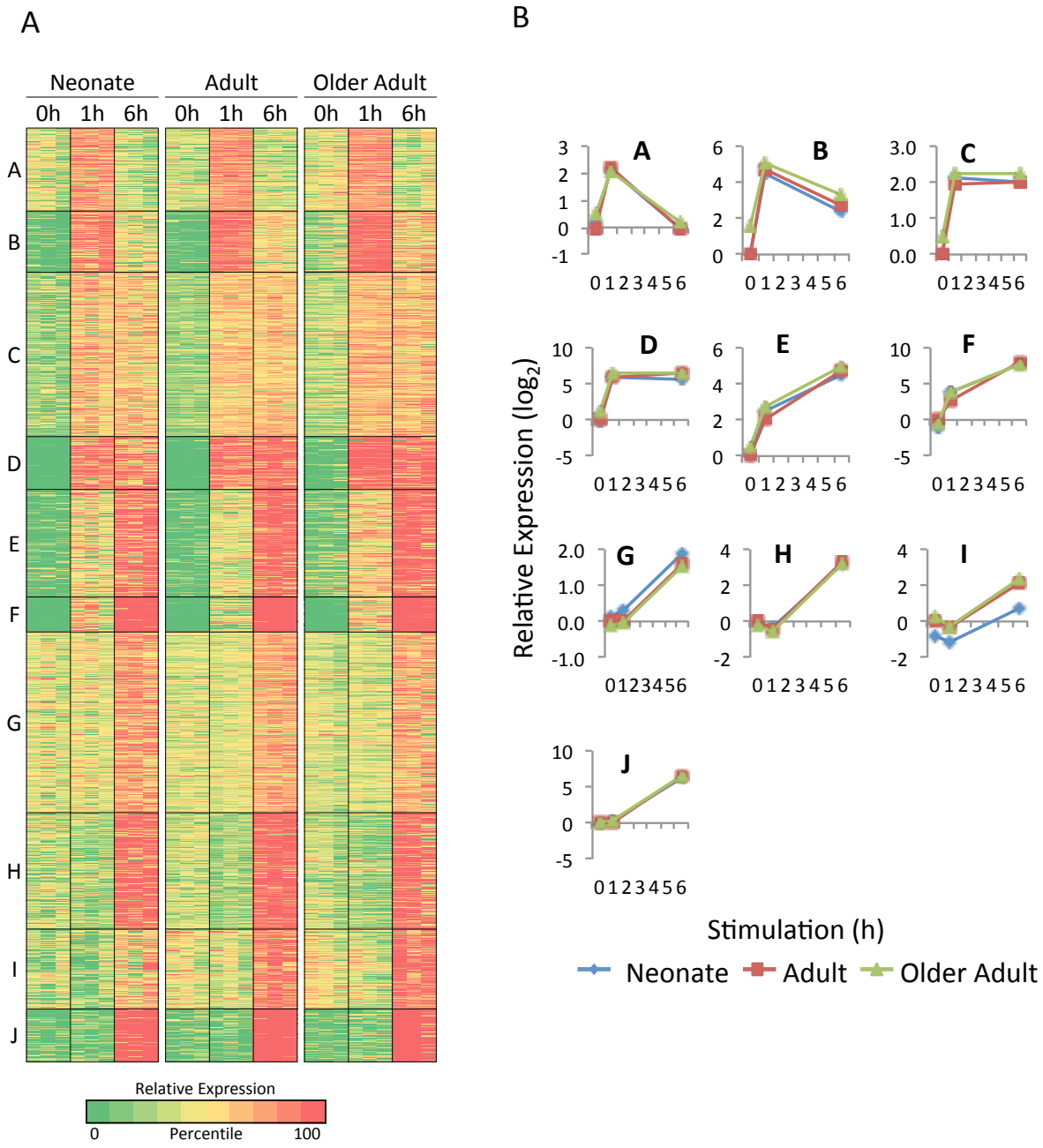


Figure C-4. Analysis of *Lm*-induced genes in monocytes by K-means cluster analysis.

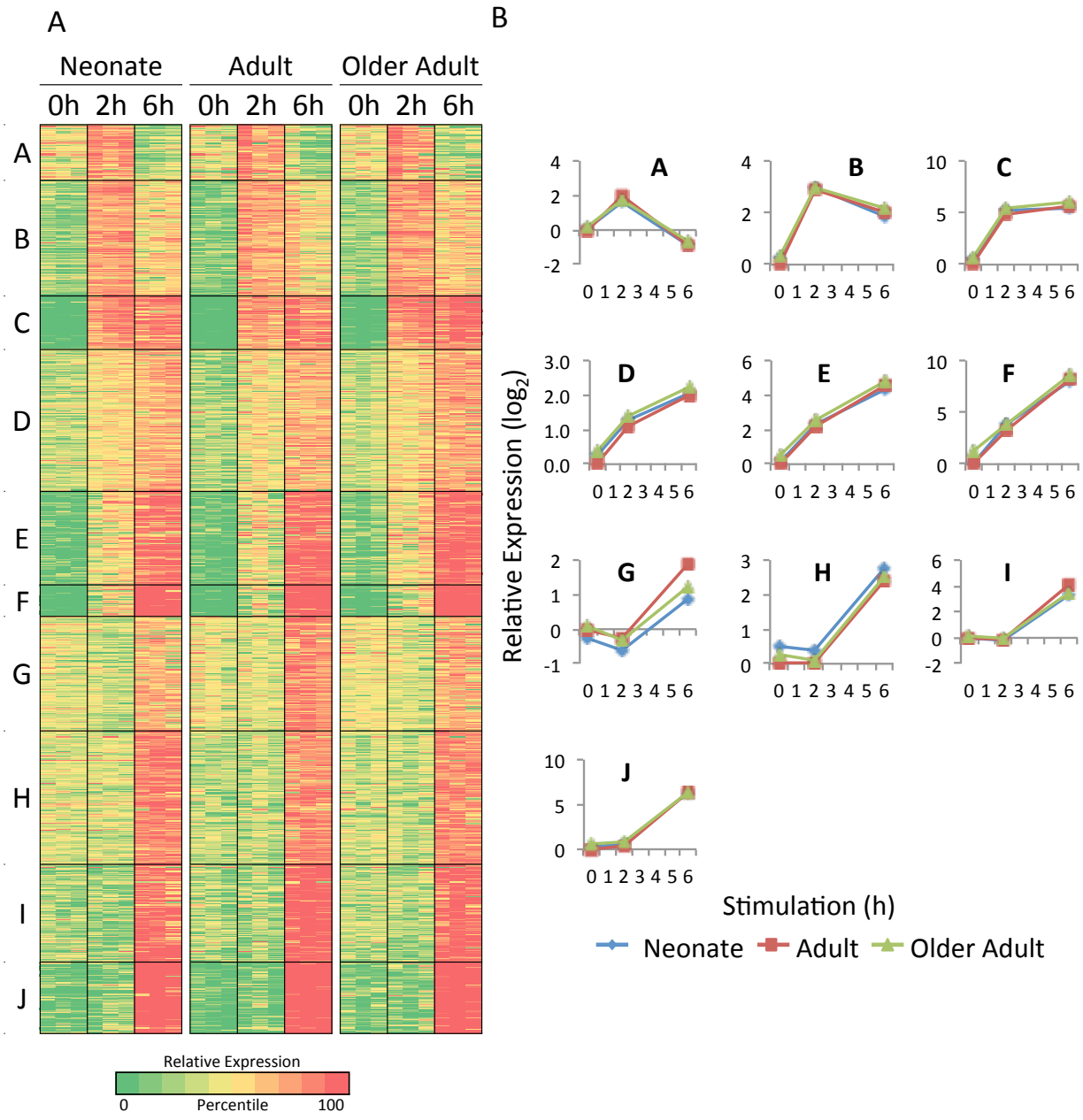


Figure C-5. Genes that exhibit the greatest expression deficit in LPS-stimulated cord blood monocytes in comparison to adult monocytes are regulated by IRF3 and/or Type I IFNs.

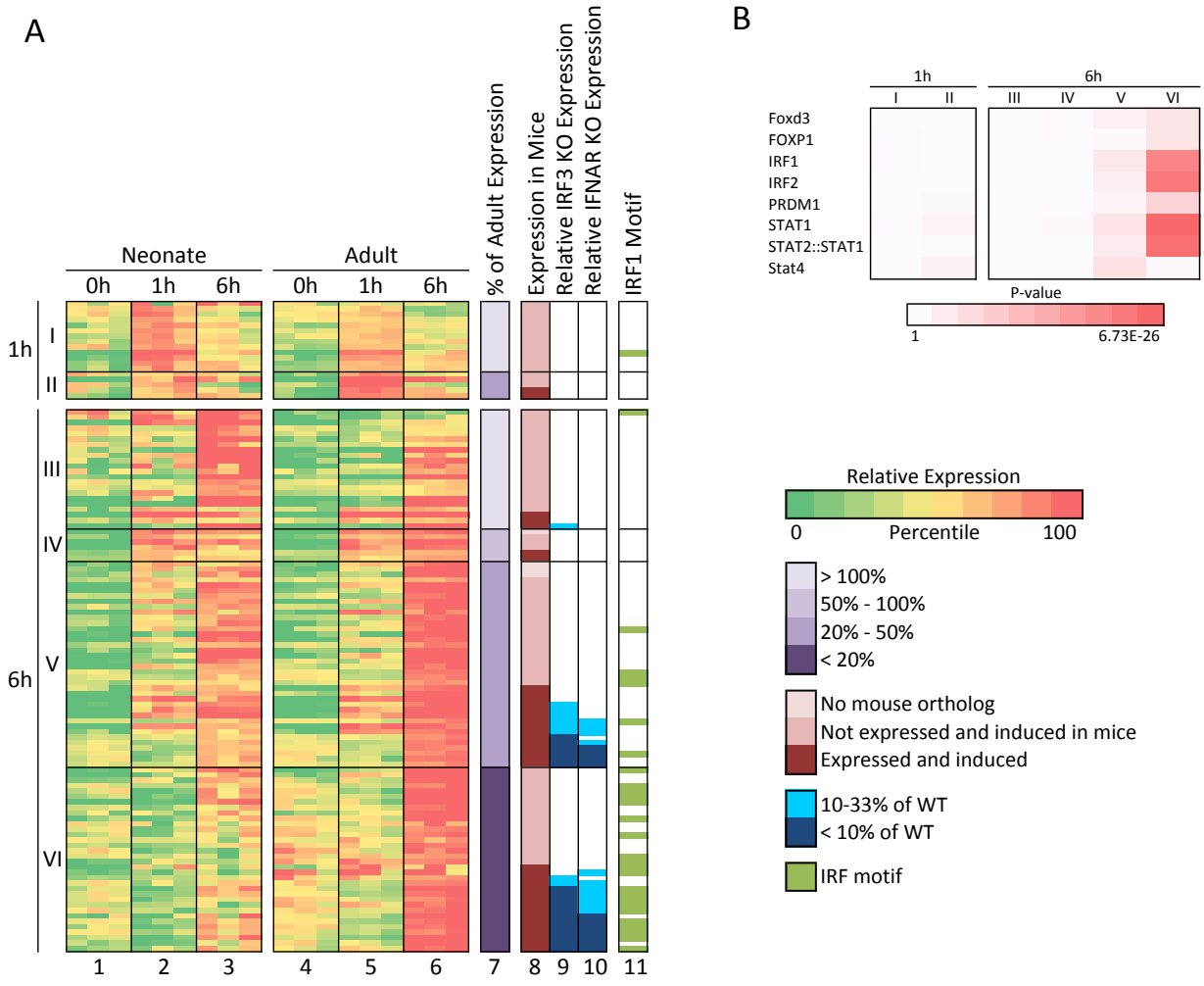


Figure C-6. LPS-induced genes exhibiting statistically significant differences in transcript levels in cord blood and young adult monocytes.

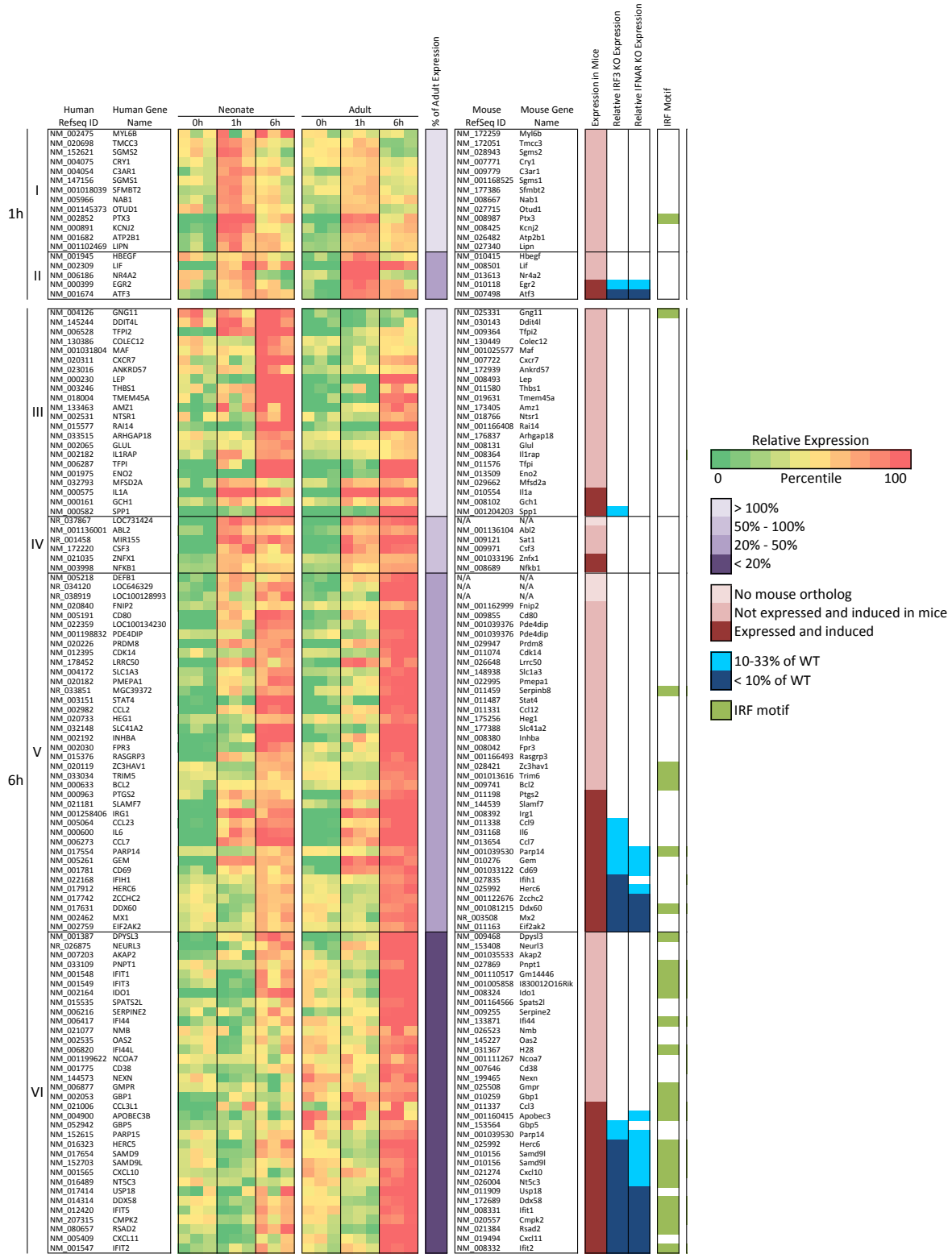


Figure C-7. Genes that exhibit the greatest expression deficit in *Lm*-stimulated cord blood monocytes in comparison to adult monocytes are regulated by IRF3 and/or Type I IFNs.

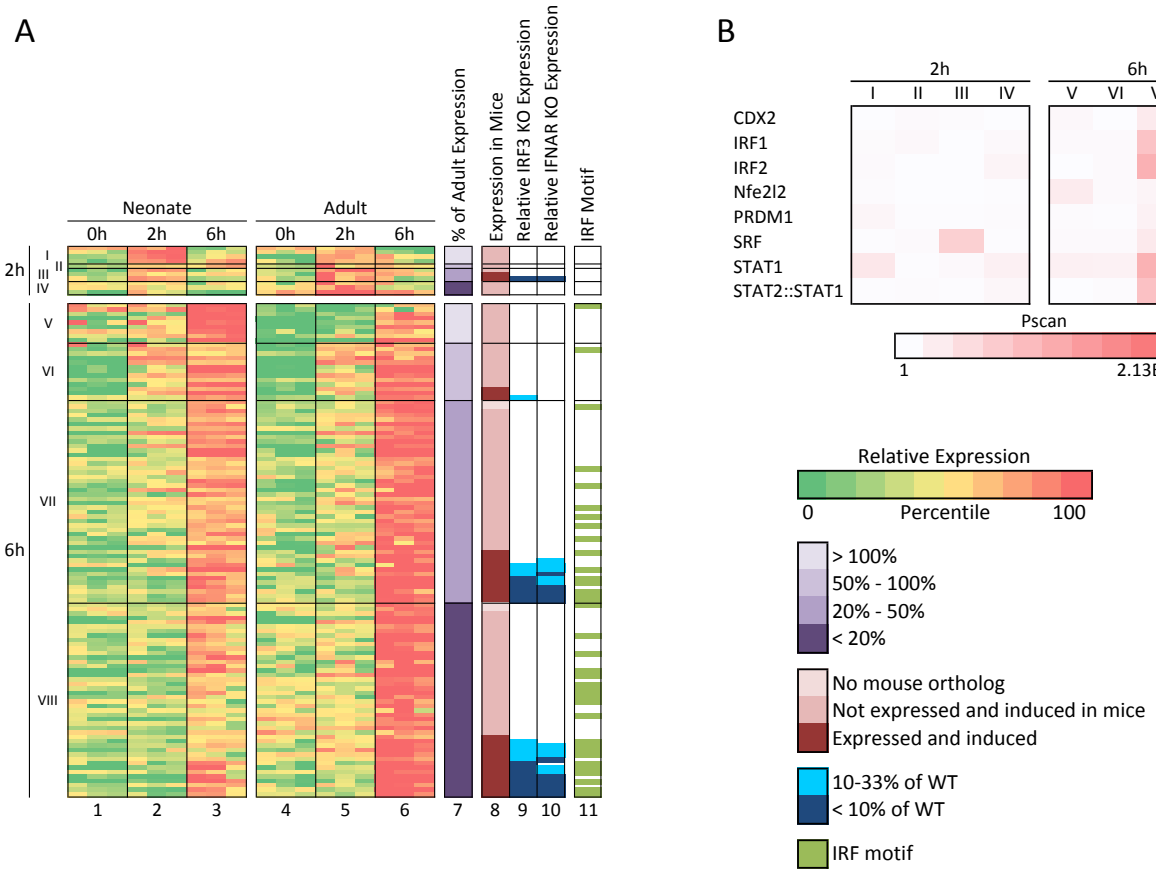


Figure C-8. *Lm*-induced genes exhibiting statistically significant differences in transcript levels in cord blood and young adult monocytes.

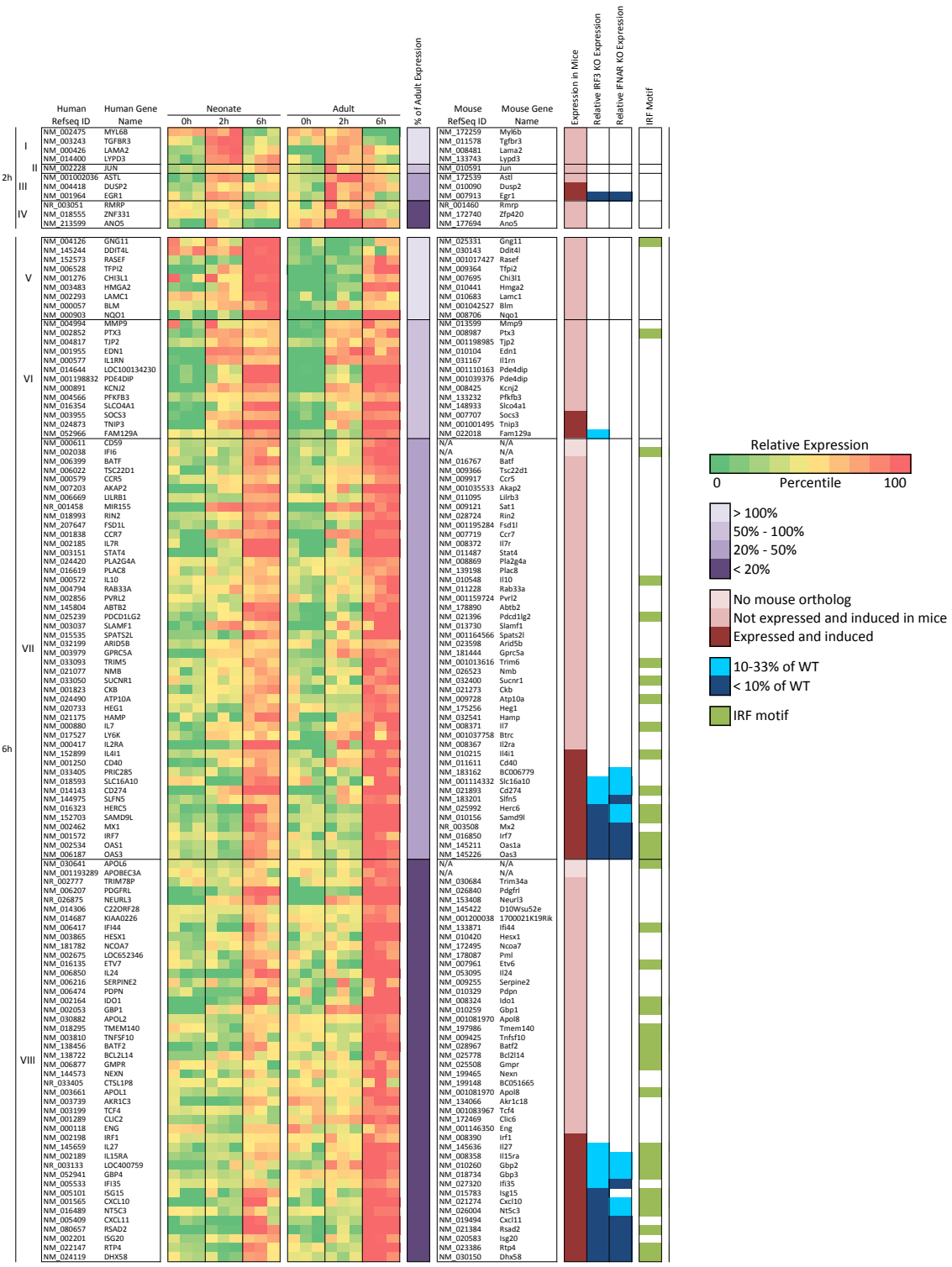


Figure C-9. Elevated expression of a broad range of inflammatory genes prior to stimulation of freshly isolated monocytes from older adults.

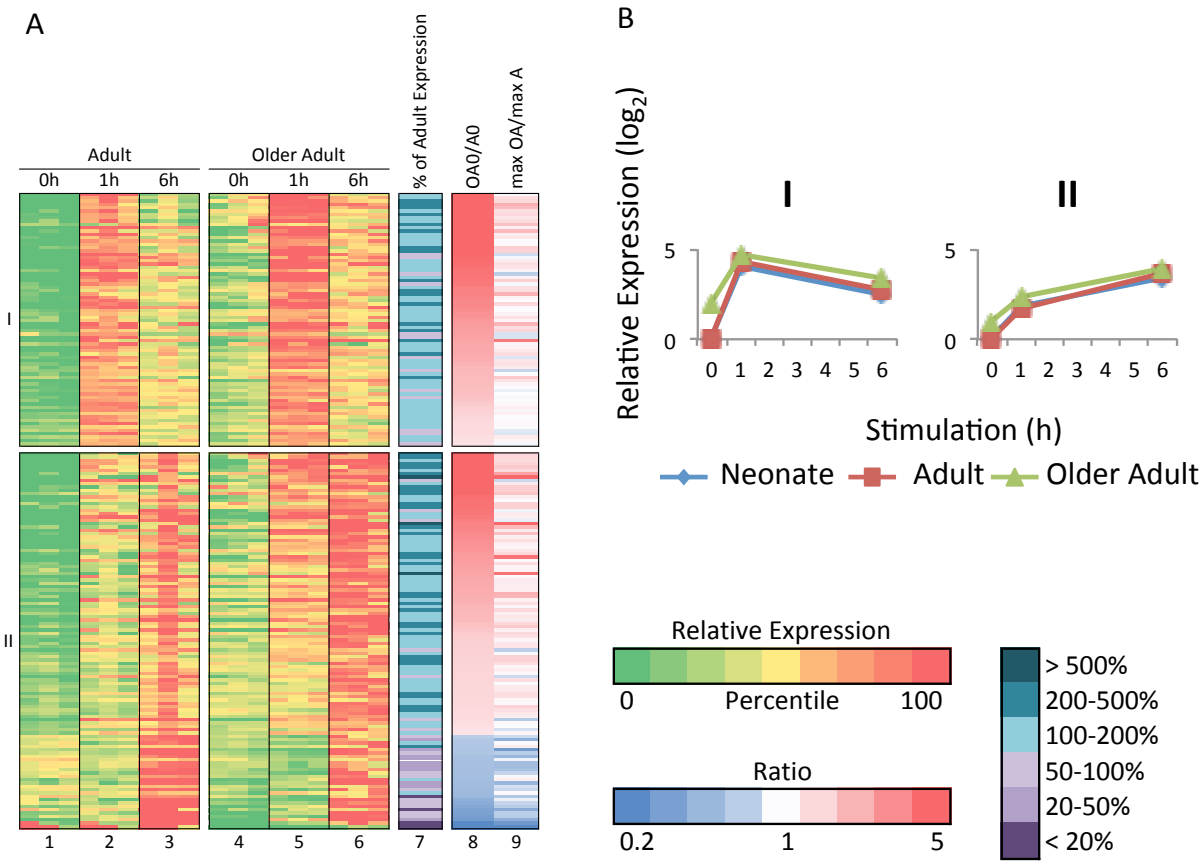
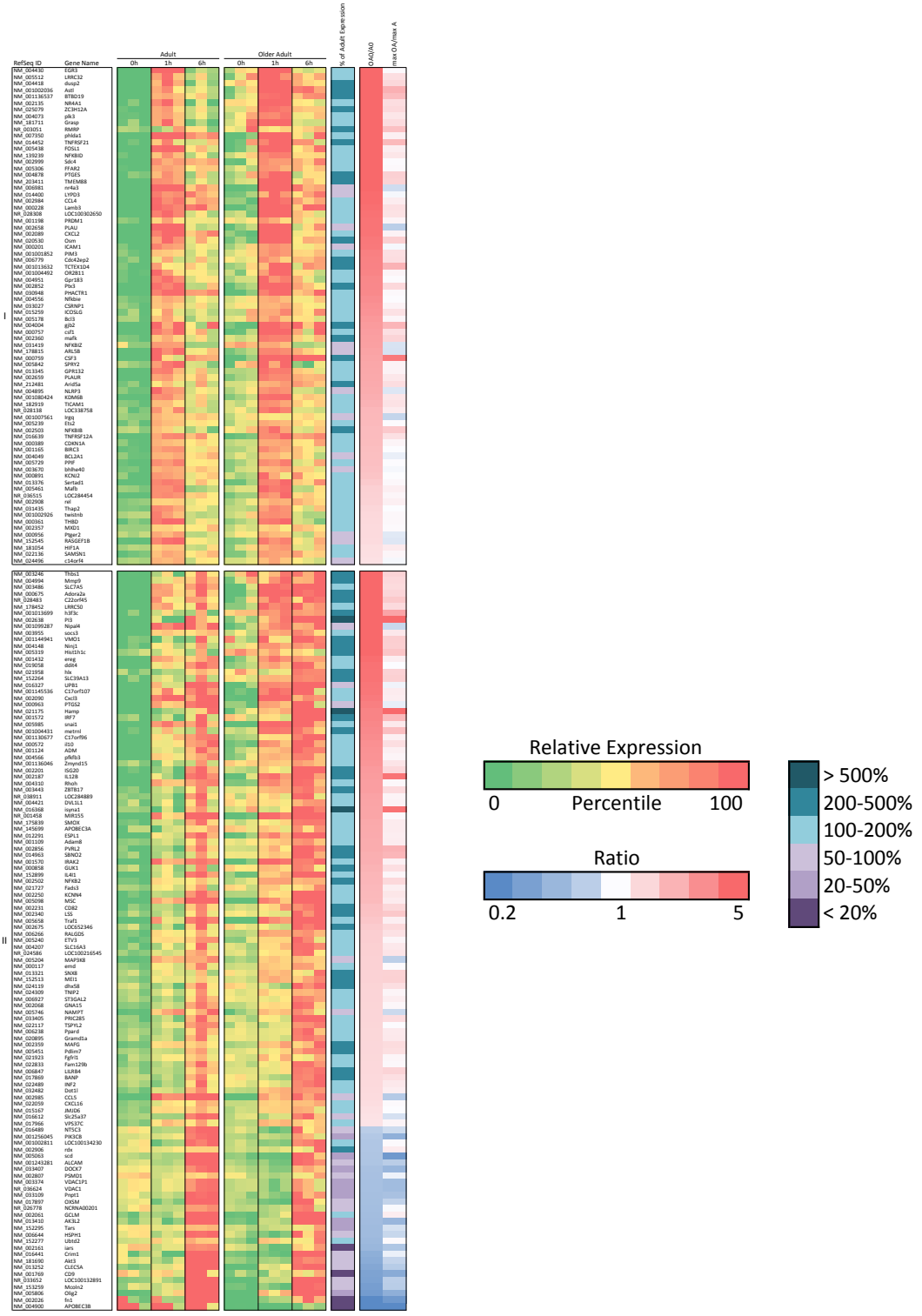


Figure C-10. LPS-induced genes that exhibit statistically significant differences in basal transcript levels in monocytes from young and older adults.



Works Cited

1. Prabhudas M, Adkins B, Gans H, King C, Levy O, et al. (2011) Challenges in infant immunity: implications for responses to infection and vaccines. Nat Immunol 12: 189-194.

2. Kollmann, TR, Levy, O, Montgomery, RR, Goriely, S (2012) Innate immune function by Toll-like receptors: distinct responses in newborns and the elderly. Immunity 37: 771-783.

3. Levy O, Goriely S, Kollmann TR (2013) Immune response to vaccine adjuvants during the first year of life. Vaccine 31: 2500-2505.

4. De Wit D, Tonon S, Olislagers V, Goriely S, Boutriaux M, et al. (2003) Impaired responses to toll-like receptor 4 and toll-like receptor 3 ligands in human cord blood. J Autoimmun 21:277-281.

5. De Wit D, Olislagers V, Goriely S, Vermeulen F, Wagner H, et al. (2004) Blood plasmaytoid dendritic cell responses to CpG oligodeoxynucleotides are impaired in human newborns. Blood 103: 1030-1032.

6. Wilson CB, Kollmann TR (2008) Induction of antigen-specific immunity in human neonates and infants. Nestle Nutr Workshop Ser Pediatr Program 61: 183-195.

7. Iliodromiti Z, Anastasiadis A, Varras M, Pappa KI, Siristatidis C, et al. (2013) Monocyte function in the fetus and the preterm neonate: immaturity combined with functional impairment. Mediators Inflamm 2013: 753752.

8. Goriely S, Vincart B, Stordeur P, Vekemans J, Willems F, et al. (2001) Deficient IL-12(p35) gene expression by dendritic cells derived from neonatal monocytes. J Immunol 166: 2141-2146.

9. Vanden Eijnden, S, Goriely S, De Wit D, Goldman M, Willems F (2006) Preferential production of the IL-12(p40)/IL-23(p19) heterodimer by dendritic cells from human newborns. Eur J Immunol 36: 21-26.

10. Lavoie PM, Huang Q, Jolette E, Whalen M, Nuyt AM, et al. (2010) Profound lack of interleukin (IL)-12/IL-23p40 in neonates born early in gestation is associated with an increased risk of sepsis. J Infect Dis 202: 1754-1763.

11. Corbett NP, Blimkie D, Ho KC, Cai B, Sutherland DP, et al. (2010) Ontogeny of Toll-like receptor mediated cytokine responses of human blood mononuclear cells. PLoS One 5: e15041.

12. Fortin CF, McDonald PP, Lesur O, Fulop T Jr (2008) Aging and neutrophils: there is still much to do. Rejuvenation Res 11: 873-882.

13. Plackett TP, Boehmer ED, Faunce DE, Kovacs EJ (2004) Aging and innate immune cells. J Leukoc Biol 76: 291-299.

14. Goriely S, Van Lint C, Dadkhah R, Libin M, De Wit D, et al. (2004) A defect in nucleosome remodeling prevents IL-12(p35) gene transcription in neonatal dendritic cells. J Exp Med 199: 1011-1016.

15. Aksoy E, Albarani V, Nguyen M., Laes JF, Ruelle JL, et al. (2007) Interferon regulatory factor 3-dependent responses to lipopolysaccharide are selectively blunted in cord blood cells. Blood 109: 2887-2893.

16. van Duin D, Mohanty S, Thomas V, Ginter S, Montgomery RR, et al. (2007) Age-associated defect in human TLR-1/2 function. J Immunol 178: 970-975.

17. Mold JE, Venkatasubrahmanyam S, Burt TD, Michaelsson J, Rivera JM, et al. (2010) Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans. Science 330: 1695-1699.

18. Krow-Lucal ER, Kim CC, Burt TD, McCune JM (2014) Distinct functional programming of human fetal and adult monocytes. Blood 123: 1897-1904.

19. Krow-Lucal ER, McCune JM (2014) Distinct functional programs in fetal T and myeloid lineages. Front Immunol 5: 314.

20. Pettengill M, Robson S, Tresenriter M, Millan JL, Usheva A, et al. (2013) Soluble ecto-5'-nucleotidase (5'-NT), alkaline phosphatase, and adenosine deaminase (ADA1) activities in neonatal blood favor elevated elevated extracellular adenosine. J Biol Chem 288: 27315-27326.

21. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18: 1509-1517.

22. Kollmann TR, Mailman T, Bortolussi R (2013) Listeriosis. In Feigin and Cherry's Textbook of Pediatric Infectious Diseases. 7th ed. Philadelphia, Elsevier Saunders. Eds: Feign R, Cherry J, Demmler-Harrison G, Kaplan S.

23. Way SS, Kollmann TR, Hajjar AM, Wilson CB (2003) Cutting edge: protective cell-mediated immunity to Listeria monocytogenes in the absence of myeloid differentiation factor 88. J Immunol 171: 533-537.

24. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, et al. (2009) A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell 138: 114-128.

25. Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, et al. (2012) Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell 150: 279-290.

26. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, et al. (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 7: 709-715.

27. Goecks J, Nekrutenko A, Taylor J, Team G (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11: R86.

28. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511-515.

29. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics 20: 1453-1454.

30. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.

31. Zambelli F, Pesole G, 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: W247-52.

32. Lin R, Genin P, Mamane Y, Hiscott J (2000) Selective DNA binding and association with the CREB binding protein coactivator contribute to differential activation of alpha/beta interferon genes by interferon regulatory factors 3 and 7. Mol Cell Biol 20: 6342-6353.

33. Witte CE, Archer KA, Rae CS, Sauer JD, Woodward JJ, et al. (2012) Innate immune pathways triggered by Listeria monocytogenes and their role in the induction of cell-mediated immunity. Adv Immunol 113: 135-156.

34. Liu S, Cai X, Wu J, Cong Q, Chen X, et al. (2015) Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induced IRF3 activation. Science, epub ahead of print.

35. Danis B, George TC, Goriely S, Dutta B, Renneson J, et al. (2008) Interferon regulatory factor 7-mediated responses are defective in cord blood plasmacytoid dendritic cells. Eur J Immunol 38: 507-517.