Title

Highly Selective Changes in Chromatin Accessibility in Response to Pro-Inflammatory and Anti-Inflammatory Stimuli

Author

Thomas, Brandon Thomas

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Highly Selective Changes in Chromatin Accessibility in Response to Pro-Inflammatory and Anti-Inflammatory Stimuli

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

by

Brandon Thomas

2017

ABSTRACT OF THE DISSERTATION

Highly Selective Changes in Chromatin Accessibility in Response to Pro-Inflammatory and Anti-Inflammatory Stimuli

by

Brandon Thomas

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2017

Professor Siavash K. Kurdistani, Chair

A major challenge in understanding complex rheumatologic and inflammatory diseases has been to understand the relative contributions and influences of multiple cytokines and cell types to disease pathogenesis. Progress has been limited by our inadequate knowledge of the underlying mechanisms responsible for regulating inflammatory responses, both in circumstances of human disease, but also at a fundamental level of how transcription factors (TFs), chromatin, and diverse stimuli regulate inducible transcription in immune cells. We have utilized in-vitro models of mouse macrophage activation combined with hi-throughput sequencing in order to provide a high-resolution view of the inflammatory response. The long-term goal is to identify modes of regulation governing inducible gene expression with the expectation that the knowledge gained will provide insight into how the immune response can be manipulated in the setting of human disease. Here, we provide evidence for both broad trends governing inducible transcription as well as examples of highly specific forms of regulation occurring at individual genes. We further provide support for utilizing mouse models in relation to questions regarding human physiology, as well as address the role of Interluekin-10 in the inhibition of inducible chromatin at sites upstream of inhibited genes.

The dissertation of Brandon Thomas is approved.

Alex Hoffmann

Matteo Pellegrini

Kathrin Plath

Stephen Smale

Siavash K. Kurdistani, Committee Chair

University of California, Los Angeles

2017

In dedication to Heather and David Thomas;

for being my first and greatest teachers in life

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

FIGURES

TABLES

ACKNOWLEDGMENTS

I am fortunate to have met and worked with many talented and encouraging people throughout my graduate studies. This work is a culmination of many discussions, collaborations, and support from others, of which my thesis advisor, Stephen Smale, has been central to my development as a scientist, thinker, and communicator. Other important influences I would like to thank include all members of the Smale lab. I've had great conversations with all of you - physics, philosophy, politics, relationships, careers, history, and yes, occasionally science - I couldn't imagine a better working environment. In particular, I would like to thank Ann-Jay Tong for teaching me all the bioinformatics and wet-lab protocols during my research rotation and for inviting me to all the grad division happy hours. I would also like to thank all members of Peter Tontonoz's lab who made me feel welcome during collaborative experiments, in particular to Prashant Rajbhandari for letting me share in his excitement of metabolism and for the many thoughtful discussions.

I am fortunate to have found a good mentor early on, and I am grateful for Tony Krumm who first introduced me to research as an undergraduate student, taught me how to pipette, and continues to give guidance and support.

I would like to also thank Rebecca Virata; thank you for coming into lab with me on the weekends in Seattle, thank you for listening to and editing my research fellowship application at the Brooklyn Brewery, and thank you for dropping off burritos to lab after your many 30-hour shifts across the street. You have spoiled me.

Lastly, to my parents who early on identified and encouraged my interests in science, you have served as role models to me as parents, as teachers, as friends. Thank you.

Chapter 1 contains reprints of figures and tables from published manuscripts with permission by Elsevier (Order: 4080240549081, 4080480625299) for Figure 1-1 and Table 1-1 and by Nature Publishing group (Order: 4080271375579, 4080271375579) for Figure 1-2 and Figure 1-3.

Chapter Two is a reprint of an article published in PLoS One with permission via the Creative Commons Attribution (CC BY) license.

Lissner MM, Thomas BJ, Wee K, Tong AJ, Kollmann TR, Smale ST. Age-Related Gene Expression Differences in Monocytes from Human Neonates, Young Adults, and Older Adults. *PLoS One*. 2015 Jul 6;10(7):e0132061. doi: 10.1371/journal.pone.0132061. eCollection 2015. Conceived and designed the experiments: MML BJT KW TRK STS. Performed the experiments: MML BJT KW AJT. Analyzed the data: MML BJT AJT. Wrote the paper: MML BJT KW TRK STS.

Chapter Three is version of a manuscript titled “IL-10 Signaling Remodels Adipose Chromatin Architecture to Limit Thermogenesis and Energy Expenditure“ submitted to *Cell.* Prashant Rajbhandari, Cynthia Hong, Laurent Vergnes, Tamer Sallam, Bo Wang, Stephen Young, Karen Reue, Stephen Smale, and Peter Tontonoz were co-authors in this study. P.R. designed and performed most of the experiments. P.R. and B.J.T. performed animal studies and genome-wide sequencing studies. L.V., and K.R. performed mitochondria respiration analysis. P.R., B.J.T., and C.H. performed bone-marrow transplantation experiment. B.W. performed multiplex immunoassay. P.R. and T.S. performed metabolic chamber studies. S.G.Y., K.R., S.T.S, and P.T. designed experiments. P.R. and P.T. wrote the manuscript.

Appendix A is a reprint of an article published in Cell with permission by Elsevier (Order: 4055071224865):

Tong, A.-J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L.,

Barish, G.D., and Smale, S.T. (2016). A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. Cell 1–15. A.-J.T., X.L., B.J.T., and M.M.L. designed and performed most experiments and wrote the manuscript. M.R.B., M.D.S., A.L.A., and G.D.B performed the RelA ChIP-seq experiments, and S.T.S. designed experiments and wrote the manuscript.

This work was supported by the UCLA Microbial Pathogenesis Training Grant (T32AI007323) and the UCLA Medical Scientist Training Program (T32GM008042).

VITA

Education

2010	Bachelors of Science, Biochemistry, Cum Laude
	University of Washington – Seattle, Washington

Honors

2008-2009	Mary Gates Research Scholar
	University of Washington, Seattle, WA
2010	Biochemistry Departmental Honors
	University of Washington, Seattle, WA
2015-2017	Microbial Pathogenesis Training Grant Recipient
	University of California, Los Angeles, CA

Publications

Thomas BJ, Rubio ED, Krumm N, Broin PO, Bomsztyk K, Welcsh P, Greally JM, Golden AA, Krumm A. "Allele-specific transcriptional elongation regulates monoallelic expression of the IGF2BP1 gene" *Epigenetics Chromatin*. 2011, 4:14. doi: 10.1186/1756-8935-4-14.

Lissner MM*, **Thomas BJ***, Wee K*, Tong AJ, Kollmann TR, Smale ST. Age-Related Gene Expression Differences in Monocytes from Human Neonates, Young Adults, and Older Adults. *PLoS One*. 2015 Jul 6;10(7):e0132061. doi: 10.1371/journal.pone.0132061. eCollection 2015. (*Equal Contributions)

Tong AJ*, Liu X*, **Thomas BJ***, Lissner MM, Baker MR, Senagolage MD, Allred AL, Barish GD, Smale ST. A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. *Cell*. 2016 Mar 24;165(1):165-79. doi: 10.1016/j.cell.2016.01.020. Epub 2016 Feb 25. (*Equal Contributions)

Abstracts

Tong AJ, Liu X, **Thomas BJ**, Lissner M, and Smale ST. Understanding the transcriptional program during macrophage activation. Presented at the Cold Spring Harbor meeting on Gene Expression and Signaling in the Immune System. Cold Spring Harbor, NY, April 2014

Lissner M, **Thomas BJ**, Tong AJ, Fortuno III E, Wee K, Kollmann T, Smale ST. Age-dependent deficiencies in immune signaling pathways. Presented at the UCLA MSTP Research Conference. UC Los Angeles, CA, January 2014

Oral Presentations

UCLA-Caltech MSTP Tutorial Series, Jan. 25th 2016: "Understanding the contributions of chromatin and gene-specific mechanisms regulating inflammatory cascades" UC Los Angeles, CA.

MBIDP Student Seminar Series, February 10th, 2016: "Mechanisms of IL-10 inhibition and its role in adipose tissue biology" UC Los Angeles, CA.

Microbial Pathogenesis Training Program Trainee Seminar, June 15th 2016: "Understanding transcriptional responses to inflammatory stimuli and new insights into mechanisms of Interleukin-10 mediated gene repression" UC Los Angeles, CA.

UCLA-Caltech MSTP Research Conference, September 9th, 2016: "Identification of gene-specific mechanisms regulating transcription by non-conventional analysis of genomics data" UC Los Angeles, CA.

CHAPTER 1

Introduction

Gene Regulation Studies Using Mouse Cells in the Genomics Era

DNA

Life on earth depends upon the synthesis, maintenance, and propagation of deoxyribonucleic acid (DNA). DNA consists a sugar-phosphate backbone with one of four nitrogen containing chemical side chains; Adenine, Cytosine, Guanine, and Thymine (A,C,G,T). These side chains are attached to a sugar backbone and can be used to form a polymer consisting of alternating bases. The nitrogenous bases on one polymer of DNA interact via hydrogen bonding with nucleobases on a complementary strand of DNA in a specific arrangement, which requires Adenine to hydrogen bond with Thymine on the complementary strand (and Cytosine with Guanine). DNA exists stably as a dimer with a bend in the backbone, which results in a double helical structure (Franklin and Gosling, 1953; Watson and Crick, 1953; Wilkins et al., 1953), and localizes to the nucleus of eukaryotic cells where it closely associates with histone proteins. The histone/DNA complexes that result from close association and wrapping of ~150 base pairs are termed nucleosomes. Nucleosomes contribute to the higher-order structure of DNA within the cell and can facilitate or impede biological processes acting on the DNA, such as transcription, replication, and repair.

The phasing of A,T,G,C base pairs along a strand of DNA represents a method of information storage, whereby specific arrangements of sequences act as high-affinity interaction sites for cellular machinery. Some segments of DNA act as binding sites for proteins, other segments of DNA are interconverted to Ribonucleic acid (RNA), another form of information storage that acts as a template for the synthesis of other macromolecular structures. RNA sequences arise from their DNA counterparts through the process of transcription, where a protein with DNA binding and polymerase ability contacts the DNA and synthesizes RNA using one DNA strand as template. RNA polymerases covalently attach the same four nitrogenous bases used in DNA with the exception of Thymine, which exists in a demethylated form in RNA called Uracil (U). Once formed, processed, and released from the nucleus of eukaryotic cells, RNA can be used as template for the generation of amino acid containing polymers called

proteins. Proteins vary in their functions; some enable the structural support needed within cells, some are released outside of the cell to act as complex chemical messengers, and others are transported back within the nucleus to act as DNA- and RNA-polymerases to ensure the continued generation of more DNA and RNA species within the cell.

Transcriptional Regulation

RNA transcription is a process that occurs continually during the life cycle of a cell. Constitutive expression is needed for basal cellular metabolism and functioning, while inducible transcription is the result of the cell's ability to interact with and respond to the surrounding environment. This holds true in circumstances where a cell needs to mount a response to suit the specific needs brought on by the changing environment, but also during development and maturation of cells from a common stem cell progenitor. Muscle, skin, blood, and nerve cells all perform unique roles within the body, but all can be traced through a developmental lineage with a shared common origin. The gene expression programs have been established in each of these cells to best suit the needs and specific roles each plays in whole body homeostasis, with each 'gene program' providing the necessary proteins needed for carrying out cell-specific functions. Developmentally regulated expression on genes can feed-forward to further commit cells down a particular maturation pathway and the continued differentiation of a cell is directed by tissue specific expression of genes involved in regulating cellular identify. It is because of differential gene expression and strict regulation of transcriptional output that the diversity and uniqueness of cell types are able to emerge from a shared common origin, and therefore, is a pre-requisite for multicellular life.

Transcriptional responses following cellular stimulation or perturbation are targeted towards the generation of specific genes whose protein products may be used to help achieve homeostasis under a new environmental signal. This often arises from a stimulus-specific activation of DNA binding proteins with the ability to recruit RNA polymerases to the sites of

target genes. These proteins are called Transcription Factors (TFs), and while there are many varieties of TFs within the cell, nearly all biological responses rely on the cell's ability to convert environmental signals to productive gene transcription through the use of differential TF activation.

TFs often contain at least two sub-domains in their amino-acid sequence. One domain is used for DNA binding (DBD) and can be used for recognizing short DNA sequences, often with high affinity for specific DNA residues within the short stretch of DNA. The unique arrangement of DNA base pairs that preferentially associate with the DNA binding domain of a protein is referred to the 'motif', and protein binding can be predicted throughout the genome by comparing the motif strength found at any given locus in the genome and the computationally or biochemically determined protein-binding motif preferences of any TF of interest. Other domains of a TF can be specific for protein-protein interactions including dimerization with itself or with other proteins, a domain that can be used for regulation such as sites of potential ubiquitination, phosphorylation, or cleavage by proteases within the cell. These regulatory regions within the protein structure acts as sites that can be modified in response to an environmental signals and are able to convert the TF from an inactive state to an active state. TFs can be held in an inactive state by a variety of mechanisms. One common feature of stimulus responsive TFs are that they exist outside of the nucleus during basal conditions and are only transported into the nucleus following stimulation and subsequent protein modification. Phosphorylation, cleavage, and dimerization are all processes known to contribute to nuclear localization of TFs, but many other forms of TF regulation have been documented.

Chromatin regulation of transcription

Once an active TF has entered the nucleus it can make contact with the preferred DNA motif and has the ability to recruit other proteins that are needed for transcription to occur, including general transcription factors and RNA polymerase II (RNA Pol II). Chromatin configuration

within cells is highly dynamic, however, and if the target DNA sequence is located in an inaccessible region the activated TF may not be able to bind to facilitate transcription upon entering into the nucleus. Instead, before TF binding and transcription can occur, the inactive nucleosome must be modified and the underlying DNA sequence must be made accessible for protein binding (Li et al., 2007). Nucleosome remodeling, therefore, is an additional mechanism used by cells to control transcriptional responses to a stimulus and is a regulatory obstacle that must be overcome in order to TF binding to occur (Figure 1-1). Nucleosome remodeling can occur in a number of different ways (Calo and Wysocka, 2013). Some DNA binding proteins have the direct ability to recognize their motifs even when embedded within an inactive nucleosome structure. These proteins have been termed 'pioneer factors' for their ability to be among the first proteins to bind and begin the chromatin remodeling process so that subsequent binding from non-pioneering factors can be achieved. Alternatively, it is thought that if closed chromatin contains numerous binding sites for TF that can bind cooperatively, the non-pioneering TF can bind together, potentially beginning at sites most distal to the core of the nucleosome where motifs could be partially exposed and accessible to binding. Regardless of the initiating events that lead to initial protein binding, the chromatin can be acted upon by chromatin remodeling complexes that add modifications leading to a sustained open chromatin configuration, often times these chromatin changes remain long after the initiating stimulus leading to nucleosome remodeling is withdrawn from the cell's environment. Because chromatin changes are dynamic and maintained within the cell, they represent an additional layer of information storage and can further shape the cell's response in the future. Remodeled nucleosomes can rapidly recruit TF binding and transcriptional initiation, and sites marked by open chromatin from the cells initial encounter with a stimulus can be acted upon more rapidly in response to subsequent challenges with similar stimuli.

Chromatin remodeling is facilitated by a number of proteins, often acting in concert as a multi-membered complex. Once remodeling complexes are recruited to DNA they can deliver

chemical modifications to the N-terminal tails of histone proteins at specific residues. Histone H3, as an example, contains a number of amino acid residues that can be modified with different chemical groups. Acetylation can occur at Lysine (K) at position 9, K27, and K36 by the histone-modifying enzyme GCN5, and these marks have been associated with regions demonstrating active transcription (Grant et al., 1999). These same sites can instead be methylated by a different group of histone modifiers including SETDB1, EZH2, and SET2 . Depending on the context and status of methylation (mono-, di-, or tri-methylation), these sites have been found associated with repression (H3K9me3) (Schultz et al., 2002), silencing (H3K27me) (Plath et al., 2002) or transcriptional elongation (H3K36me) (Krogan et al., 2003). Collections of histone marks have been documented and have been used to classify chromatin regions throughout the genome, and certain arrangements of histone marks are often used as surrogate markers for the chromatin configuration and accessibility to TF binding (Table 1-1). Alongside direct histone modification with chemical side-groups, the entire replacement of nucleosomes has been observed and depends upon an ATP-dependent activity of large remodeling complexes. One particularly well-studied ATP-dependent chromatin-remodeling complex is the Switch/Sucrose Non-Fermentable (SWI/SNF) complex that is conserved in bacteria, yeast, mice, and humans (Cairns et al., 1994; Côté et al., 1994; Peterson et al., 1994). This complex uses the high-energy molecule ATP to restructure and loosen the DNA/histone associations and allow for other TF and histone-modifying enzyme to act on the nucleosome.

Aside from protein-directed modification of chromatin structure, specific DNA content can directly shape nucleosome stability and therefore transcriptional regulation. The frequency of CG dinucleotides (CpG) observed in the genome is much less than the once per 16 bases that is expected if they were distributed randomly throughout the genome. The depletion of CpG content has been hypothesized to result from the deamination of methyl-cytosine to uracil and a stable mutation of CpG to TpG (Bird, 1980). There is, however, a preservation of CpG sequences in the promoter region upstream of many transcriptional start sites (TSSs). Upwards

of 70% of all gene promoters contain high CpG content near what would be expected if distributed at random, and nearly all housekeeping genes that are constitutively expressed in nearly all cell types contain numerous CpGs in close proximity to one another (Davuluri et al., 2001). Many CpG island promoters have been found to associate with properties of open and active chromatin including active histone modifications and accessible DNA. Indeed, even under conditions of low or no transcriptional activity, many CpG island promoters can be found pre-bound with a stalled RNA pol II (Hargreaves et al., 2009). There continues to be active investigation into determining whether the DNA has an intrinsic ability to direct nucleosome structure, or whether specific protein motifs are enriched in CpG islands and protein binding facilities the sustained open chromatin at these sites. Regardless of the precise determinants and mechanisms of achieving open chromatin at CpG islands, it has been shown that CpG content has the ability to directly prevent stable nucleosome formation and instead associates in regions of active chromatin (Ramirez-Carrozzi et al., 2009). Presumably, this allows rapid activation of CpG island promoters under a variety of contexts and further delays the activation of genes with low CpG content, which may exist in closed chromatin and require additional steps of chromatin remodeling for productive gene transcription to occur.

Immune cell regulation of gene transcription

While the DNA content of all cells within a human body are approximately equivalent, each cell type is responsible for completing specific roles and requires discrete sets of proteins to accomplish unique tasks. Individual cells from different lineages often function as a collection, with different gene expression programs established across different types of cells that together constitute a tissue. Blood can be considered as a tissue because it contains many different kinds of cells that posses specific functions. The differential gene expression programs that allow for unique functioning are achieved through developmental expression and regulation of key TFs. These developmental TFs turn on specific target genes through a combination of

direct recognition of the DNA target sequences, interactions with other co-factors, and the overall chromatin structure of the host cell.

Analogous to how many different cells work together within a given tissue, many tissues and organs can work together to constitute a system. A system is a highly specialized collection of cells, organs, and tissues distributed across the body dedicated towards a common goal. The immune system is responsible for surveying the body for signs of foreign invasion and pathogenic substances. Furthermore, immune cells contain TFs that, once becoming activated, are able to interact with DNA near genes encoding inflammatory and microbiocidal proteins to promote their transcription. The immune system's ultimate role, therefore, is to be able to distinguish host cells from those of another origin (self vs non-self). If an immune cell detects a foreign substance it determines to be potentially pathogenic, it must either eradicate the threat individually and/or recruit other cell types with an enhanced ability to resolve the threat. Innate immune cells are a type of immune cell that are considered the 'first line' defense against invading pathogens. These cells patrol the body at many environmental:host interfaces, such as the skin, the intestinal tract, the oral mucosa, and lungs. Innate immune cells also circulate in the blood and can be recruited to sites of injury or inflammation. Adaptive immune cells, including B-lymphocytes and T-lymphocytes, contribute to long lasting immunity through their ability to proliferate into cells with high affinity receptors against particles present on the invading pathogens. While distinct entities, innate and adaptive immune cells work together to protect the host against invading threats. Immune cells are able to communicate with one another through the transcription and synthesis of secreted proteins including cytokines and chemokines that activate specific signaling pathways once detected by other cells.

Toll-like receptor 4 and pattern recognition

The ability to sense and respond to foreign molecules and stress signals is of central importance to the protection conferred by innate immune cells. Cells are able to sense pathogen

associated molecular patterns (PAMPs) through one of their Pattern Recognition Receptors (PRRs). One well-studied family of PRRs are the Toll-like receptor family (TLR), of which TLR4 is present on the cell surface and has high affinity for a PAMP found the gram negative bacterial cell wall component, lipopolysaccharide (LPS). TLR4 sits on the outer membrane of many immune cells and complexes with other proteins to form a signaling receptor complex (Park et al., 2009) (Figure 1-2). Myeloid Differentiation Factor 2 (MD2) and TLR4 form a heterodimer that interacts with LPS bound to CD14. This complex allows for efficient extracellular recognition of LPS and intracellular dimerization of the Toll Il-1 Receptor (TIR) domains on the TLR4 receptor. These domains are highly conserved across multiple animal and plant species, highlighting their important role in transmitting environmental signals to intracellular pathways. On the inside of the cell, a number of other accessory proteins associate with the TLR4 receptor complex and amplify the bacterial signal across many signaling pathways. Myeloid differentiation primary response gene 88 (MYD88) and TIR-domain-containing adapter-inducing interferon-B (TRIF) are two important adapter proteins used for TLR4 signaling. MYD88 activation leads to a rapid activation of kinases and ubiquitin ligases that feed into the activation of MAP kinases and NF-κB signaling. The MYD88-independent activation following TLR4 stimulation signals through TRIF and leads to a partial restoration of NF-κB activation, but with delayed kinetics due to the requirement of endocytosis and activation at the endosome. TRIF signaling is unique to MYD88 in that it leads to a potent activation of IRF3, which binds to a different set of target genes important for anti-viral immunity through the expression of type-I interferons.

Of all known TLRs, which include over 10 in human and 12 in mice, TLR4 is the only one to lead to the activation of both MYD88- and TRIF-dependent signaling pathways. This is one reason for why TLR4 has received attention for study; its dimerization and signaling results in the activation of a large range of TFs that can enter the nucleus and drive transcription of target genes. Of the many TFs that enter the nucleus to drive inducible transcription, members of the NF-κB and IRF family have been well studied. RelA, or p65, is an NF-κB family member that

localizes to the cytoplasm under resting conditions where it exists as a dimer with p50 and complexes with inhibitor NF-κB family members (IκBs). The IκBs mask the nuclear localization domain of RelA which prevents its shuttling into the nucleus. Upon stimulation, Inhibitor Kappa kinases (IKKs) become active and phosphorylate IκB. This promotes the ubiquitination and degradation of IκB and allows active NF-κB dimers to enter the nucleus and drive expression. The MYD88-dependent pathway is a stronger induced of NF-κB signaling because it promoters a greater activation of the IKKs. The TRIF-dependent pathway leads to activation of IRF3, which also remains in the cytoplasm until becoming phosphorylated and enters the nucleus to bind its distinct set of target motifs.

Significance of understanding gene regulation

The transcriptional response following PAMP recognition results in the induction of hundreds of genes; some genes encode proteins that promote inflammation while others communicate with other cells of the immune system to aide in polarization and development of the adaptive response. The immune response must be precisely regulated; it would be detrimental to the host, for example, if immune-stimulated cytokines were released from cells following weak stimulation by non-pathogenic stimuli. Similarly, poor responses to potential pathogens could result in lack of clearance and ultimate organ destruction or death of the host. The impressive control of immune activation has been of great interest to immunologists, biochemists, and physicians over the years and there has been a growing interest in understanding innate immune cell function since the first discovery of phagocytic antigen presenting cells over 100 years ago (Cavaillon, 2011) (Figure 1-3). Researchers have an immense interest in better understanding precisely how transcriptional responses are best tailored for a given stimulus so that we may subsequently manipulate and repurpose the system for our own goals. Furthermore, this knowledge could be used to explain cellular mechanisms responsible for disease pathogenesis. Examples of this can be seen throughout the history of

medicine, a specific example of advances in vaccination technologies emphasizes this point.

Initial vaccine therapies involved inoculation with low doses of live related pathogens of lower virulence. The immunological “training” on less virulent strains of pathogenic stimuli allowed for innate and adaptive immune cells to successfully resolve the threat and produce protective antibodies which could circulate and provide protection against repeat, or similar exposures. Cowpox inoculation was the first vaccine and was used to immunize against smallpox virus in the late 1700s. The name “vaccine” derives from the latin “vacca” meaning cow, highlighting the role of cowpox virus in providing immunologic protection. By exposing individuals to cowpox virus, protective immunity was built against both cowpox and smallpox. Similar efforts have been made across many human pathogens. Polio vaccines currently used today come in two forms, a live-attenuated version as well as a killed wild type version. There are pros and cons for each vaccine type; the live attenuated version can induce a much stronger immunity against the virus, but can become virulent itself in some individuals. The killed vaccine is considered safer, but fails to mount an adaptive immune protective response at mucosal sites in the body. This failure to induce a strong immunity in the safer, killed vaccine is due to a lack of strong induction of a specific transcriptional program in either the antigen-presenting cell, or a lack of strong transcriptional response in the adaptive immune cell. The knowledge of which types of transcriptional responses necessary for mounting protection via the adaptive immune system has led to work investigating the use of vaccine adjuvants, which are stimulatory compounds included with vaccine preparations to allow for targeted immune activation. It has been shown, as an example, that innate immune dendritic cells (DCs) can be used as a target cell for delivery of antigens to be used for vaccination. DCs challenged with Hepatitis B virus surface antigens are able to mount a protective immune response when co-infected with a modified LPS molecule with enhanced activation of the TRIF-dependent TLR4 signaling pathway. This vaccine may be protective due to the activation of the IRF3 TF, which enables a robust activation of type-I interferon and downstream anti-viral immunity. Without the

knowledge of basic mechanisms of TLR4 induced gene expression programs, the rational design of vaccine adjuvants would be impossible (Levitz and Golenbock, 2012; Nabel, 2013). While there have been success stories such as these through history, there is much more to be learned about the basic mechanisms responsible for coordinating effective immune responses.

Another example of clinical observations being addressed with focused mechanistic questions in the lab includes the problem observed in preterm and neonatal infants being unable to combat infectious diseases as well as older aged babies (Kollmann et al., 2017). The immune system is established early in life, but continues to change through childhood and into adulthood. It has been known that specific pathogens pose a greater threat to neonates, and that vaccine responses tend to be reduced during early life. It has been hypothesized that the diminished immune response is due to immaturity of immune cells, a programmed anti-inflammatory immune repertoire to enable early bacterial seeding of commensal sites in the skin and gut, as well as a chemical effect from hormones synthesized during fetal development. Regardless of the biological or evolutionary explanation for these findings, the clarity as to why this occurs at the cellular level only came about once researchers interested in understanding immune cells responses began to compare immune activation in neonatal cells compared to older individuals in the lab. It was shown that vaccine responses could be recovered to adult-like levels in young children with the use of adjuvant stimuli included in the vaccine (Siegrist, 2001). These results seemed to indicate that there was a particular hypo-responsiveness to stimuli in neonates compared to adults. Later studies comparing similar cell populations across different age groups were able to determine that interferon responses were diminished in neonates, while expression of the anti-inflammatory cytokine Interluekin-10 was enhanced and IRF3 activation was diminished (Aksoy et al., 2007). These findings have been confirmed by multiple groups and will be a topic of discussion in chapter 2 of this thesis.

The beauty that comes from investigating basic questions regarding gene regulation is that the knowledge gained will likely be highly conserved across species and hold broad

implications spanning many organ systems. The 'rules' learned from such studies will allow for a richer understanding of disease pathogenesis and overall cellular physiology which will be essential for moving forward as we use these rules for understanding how to best manipulate cells and systems to achieve desired outcomes that promote human health and achieve whole body homeostasis.

Gene regulation studies in the pre-genomics era

Our general interest lies in better understanding how transcription is coordinated within cells following environmental changes. Our specific interest is focused on the transcriptional cascade induced by a well-studied pattern recognition receptor, Toll-like receptor 4 (TLR4), in mouse bone marrow- derived macrophages (BMDMs). The general rationale for the studies in our lab is that a reductionist approach toward understanding how a well-defined cell type mounts a transcriptional response will be essential for the long-term goals of understanding how the response differs in other cell types, how targeted inhibitors of inflammation achieve therapeutic activities, and how technological advances in biology can be utilized for addressing mechanistic questions. The methods that we use to address these questions are considerably different than the methods originally used to study mechanisms of gene expression. Early work that focused on understanding principles of gene regulation heavily relied upon the study of individual model genes that became potently activated in response to bacterial challenge. This was because many of the technologies and experimental techniques at the time were neither sensitive to detect low levels of expression or modest changes following stimulation, nor high-throughput to be able to monitor multiple genes simultaneously. Investigators from many fields used single-gene studies to guide mechanistic work because it was believed that a detailed analysis of a single gene would provide a comprehensive view of how the gene is regulated and provide a framework that could be used to study other model genes. Emphasis was placed on identifying which regulatory sites both up- and downstream of the transcriptional start site were being

utilized for DNA-binding, as well identifying which DNA-binding factors were associated at these sites. Similarly, after mapping and characterizing the key regulatory domains, mutagenesis could be conducted to artificially change the genetic sequences of the regulatory sites and assay which individual nucleotide sequences were necessary for productive gene transcription.

The experimental approaches for monitoring gene expression in the pre-genomics era included reporter gene assays where the promoter region of a gene of interest was placed in front of a transgene that contained a marker that could be measured, often times the readout would be a fluorescent signal from either recombinant fluorescent proteins or from the enzymatic activity of the transgene after being incubated with substrate. These studies were valuable for the ability to artificially generate multiple mutant promoter constructs to assay for specific DNA base pairs required for generating the fluorescent signal. The disadvantages for this system were that the transfection of recombinant DNA on a plasmid backbone or viral infection of the DNA would be introduced at unknown numbers within the cell in a genomic context unlike the native gene under study. The quantitative nature of this assay is often limited.

Northern blots attempt to measure the endogenous gene in a native chromatin context by first collecting all RNA species within the cell, then subjecting them to gel electrophoresis which will separate the RNA by size. After transferring the RNA to another membrane, the RNA is incubated with a radiolabeled oligonucleotide that share complimentary sequences to the gene one wants to study. After washing away unhybridized probe, radiometric evaluation of gene expression can be detected with x-ray film. The sensitivity achieved with northern blot assays was much greater than reporter gene detection, and it has the additional advantage of being able to uncover size-separated isoforms of genes as well as genes that share homology to the target if the oligo was designed for this purpose.

An alternative method for measure difference in nucleic acid content was developed from knowledge of primer based hybridization and polymerase-dependent extension occurring over cycling temperatures from the same reaction tube in a continuous fashion. This method,

polymerase chair reaction (PCR) could amplify massive amounts of targeted DNA from very low starting material and was soon adopted for the use of making semi-quantitative measurements of starting material concentrations. When coupled to fluorescent imaging of DNA products after each cycle of melting, hybridization, and extension, one is able to determine the cycle number needed to reach a threshold value or detection limit. This cycle number can be compared across samples to estimate amounts of DNA present. This semi-quantitative approach has been, and will continue to be, an important tool for scientists as it provides a relatively low cost approach for measuring gene expression in a semi-throughput manner.

The main disadvantages with the assays discussed above with a focus on studying a few model genes is that one must know which gene they want to study before performing the experiment. These approaches are similarly not an attractive option for those wishing to survey many genes under the same experimental condition. Importantly, even if investigators were able to fully describe all modes of regulation occurring at any individual gene, they were unable to determine how frequently these modes of regulation were being utilized by the cell genome-wide. It was essentially impossible to address these issues with the current technologies of the time, and yet, model gene studies provided the evidence for nearly all of the fundamental aspects of transcriptional regulation that were later confirmed with more comprehensive genome-wide sequencing techniques now common in the genomics era.

Gene regulation studies in the genomics era

Technology continues to drive the advancement of understanding in the biological sciences. The introduction of DNA sequencing technologies in the 1970s ushered in an era in molecular biology that has been known as the genomics revolution as it allowed investigators to determine the precise DNA and RNA sequences of any sample of interest. The ability to sequence DNA allowed for an expanded understanding of a gene's context within the genome, identification of distal regulatory sequences, document mutational diversity across samples, and infer function

through analysis of conservation across species. Not only did it allow for a greater opportunity to interrogate genome-level biological questions, it allowed for a high-throughput method to perform experiments that were already commonplace. For researchers interested in understanding questions relating to gene expression, there were now methods available for simultaneously monitoring of hundreds or even thousands of genes. DNA microarrays were borne out of the data generated from DNA sequencing, not to mention the technological advances in fields of micro-engineering and chemical synthesis. In this method, short oligonucleotides could be synthesized and attached in groups along a surface. DNA or complimentary DNA generated from RNA from a sample of interest could then labeled and allowed to hybridize to the surface-attached DNA. If the sample of interest contained sequences complimentary to a spot on the surface it would bind and remain attached giving off a signal in the area of binding. Subsequent imaging of the entire plate would demonstrate areas of strong binding, which would reflect high abundance of transcript/DNA in the sample preparation. The locations of strong hybridization could then be referenced back to the known DNA sequences that were originally attached to the plate allowing researchers to make abundance measurements across many genes/locations simultaneously.

Commercial hi-throughput sequencers superseded microarray technologies due to the higher sensitivity of detection using chemical sequencing technology in place of the fluorescent hybridization based quantitation required by microarrays. Sequencers also provided a level of de-novo discovery that was difficult to obtain with microarrays. Array based methods require the knowledge of sequences to be studied, while chemical sequencing machines could reconstruct the transcript of DNA identity without the need of a reference during the procedure. Reference genomes were often required for the downstream analysis, but the sequencing based methods allowed for greater ease in identification of novel insertion, deletion, and polymorphic mutations in samples. Additionally, like northern blot analysis, sequencing allowed for the detection of splice variants in a much more hi-throughout manner. Sequencing technologies continue to

evolve, and the price per sample sequenced continues to drop as improvements are made in sample preparation, sequencing chemistries, and imaging improves. Not only does this improve throughput, quality, and sensitivity, it also carries the advantage of being able to monitor gene activities from relatively small numbers of naïve cells taken directly from within the body, as compared to earlier methods of studying transcription (reporter genes) which required heavy manipulation of cell lines in culture.

With the explosion of available data in gene expression studies with the advent of hi-throughput sequencing, bioinformatics techniques have been introduced to allow for efficient mapping of reads to the genome as well as categorized genes based on their expression profiles across samples and conditions. If groups of genes display similar expression patterns under a tested condition, they can be grouped together with the assumption that an underlying mechanism is responsible for their coordinated expression. This method of clustering is sometimes referred to as a 'network analysis' and often requires the grouping of hundreds or thousands of genes in order to perform statistical tests when making comparisons across different groups. One comparison that can be done, for example, is to hypothesize that if distinct clusters of genes are co-regulated there should be evidence of similar TF motifs embedded within the promoter or local region of DNA that is responsible for coordinating the genes activity. This sort of motif analysis can compute the observed frequency of motif matches across genes within the same group and compare the number of motif instances in one group compared to a background frequency which would be expected by random. Similar comparisons can be made based on the functional annotation of genes belonging to each cluster using gene ontology analysis, and eventually hypothesis that emerge from network analysis can be testing by modifying the system, predicting the response, and adjusting the classification of genes based on their observed behaviors in new contexts.

Philosophies on scientific inquiry and the scientific method

Network analyses such as these have been incredibly useful for investigators who have sought information about a system or process using an unbiased genome-wide approach. Because data-collection is the first experiment done in many network analysis studies, it allows the investigator to approach the data without having a particular hypothesis in mind prior to performing the experiment. One can simply collect measurements and investigate the potential relationships between variables in the system and predict how the system will change in response to further perturbation. This data-driven research approach is in contrast with the popular scientific approach of hypothesis-driven work that requires extensive background on established theories. In hypothesis driven research, investigators attempt to infer associations from theory in context with new ideas. They then test those inferences using the scientific method, which consists of designing well-controlled experiments with results dependent on individual variables under the control of the investigator. The results obtained can either support or refute the original hypothesis and will contribute to or refine the approved working theory. It is through this method that many important insights have been made in science, but the advent of massively parallel sequencing has allowed resurgence in data-driven research, and this has been met with some hesitation among scientists. The hesitation can arises in many ways. For one, data-driven work often heavily relies upon statistics to drive associations, and statistics can be both misleading as well as ill applied for the situation at hand (Nuzzo, 2014). Secondly, data-driven approaches that study genome-wide behaviors become extremely complicated and require mathematical algorithms to facilitate understanding, but the assumptions that must be made in generating the algorithms can introduce bias themselves. Lastly, because data-driven work usually transitions from the wet-lab (or experimental lab) to the dry-lab (computer based research), it has the potential to end with the proposal of a new hypothesis to explain the observed data. If the new hypothesis never returns back to the wet-lab to be testing under new conditions, it has the potential to gain traction in a scientific field without being adequately addressed using the scientific method.

As more and more data-driven work gets published in the biomedical sciences, it will be important to transition back towards hypothesis driven work that is brought about by associations discovered with data-driven research. The two philosophies are not at odds with one another, and the scientific method is still an essential tool for work being done in the genomics era, but the advances in technology will continue to shape the approach that is used to conduct scientific investigation and it is important to understand the limitations and obstacles facing investigators using a primarily data-driven approach (Haufe, 2013).

On the choice of a model system

While our ultimate goal is directed at achieving knowledge useful for understanding and treating human disease, there are a number of limitations in using human cells as a research tool and make mouse cells an attractive model system to use. Before addressing the advantages of using mouse cells for the questions we are interested in investigating, it is worth briefly discussing the difficulties we have in predicting which kinds of studies will hold the greatest impact on human health and whether this should be a factor in determining which kinds of studies should be done in order to achieve our long-term, human-centric goals. As an example, the first reports on the discovery of DNA ligase enzymes began in 1961 by a group whose research was focused on understanding how viruses infect bacteria, without an immediate goal in understanding any human physiology or therapeutic endpoint. In the years following, many groups were able to show that bacteriophages infecting *E.Coli* bacteria had enzymes that would attach different strands of DNA together (Meselson and Weigle, 1961). The discovery of DNA ligases enabled researchers from all fields of biology to use this as a tool for performing molecular biology cloning, and this enzyme continues to be used today as a required step in the generation of samples for hi-throughput sequencing. Arguably, the advancement of molecular diagnostics and human disease screening has been transformed by the discovery of these and other related enzymes originally discovered by labs using model organisms far removed from

human physiology. It is because of a primarily basic science interest that we have advanced our ability to perform diagnostic and therapeutic work in translational studies.

Similarly, the most recent technology famed for holding the potential to revolutionize modern medicine did not come from research groups studying how human cells behave, rather the discovery and innovative work that resulted in CRISPR-Cas9 mediated DNA mutagenesis began in the archaen *Haloferax mediterranei* during the early 1990s (Mojica et al., 1993). A series of follow up studies occurring over the next two decades resulted in the observation that Cas9-induced DNA cleavage was a bacterial conserved adaptive immune response to viral infection. The understanding of this ancient process led to the repurposing for directed mutagenesis in mammalian cells, a process that until this discovery was laborious and inefficient. It is quite possible that the great advances in human therapeutics of our generation will be due to work that was focused on understanding how bacteria respond to viral infection. It remains important to acknowledge that we cannot reliably predict where the next breakthroughs will occur, instead we should value the work being done across multiple disciplines and to share this knowledge in a way that facilitates collaboration and innovation.

Mouse macrophage model for understanding inducible transcription

Mice are a valuable model system in part due to the ability to generate large amounts of genetically identical cells that can be grown under specific conditions that allow for high reproducibility between experiments. Mouse studies also benefit from the wealth of inbred strains that lack individual genes to allow for studies on gene function. We use mouse macrophages as a model for the reasons presented above, as well as because many of the signaling pathways and immune responses are conserved in humans. This simplified system allows us to precisely control the experimental conditions in order to investigate the immune response in great detail. Most of our studies have focused on the transcriptional cascade induced by a well-studied pattern recognition receptor, TLR4, in mouse bone marrow- derived

macrophages (BMDMs). Many of the induced genes are present at very low or undetectable levels before stimulation, yet they increase transcription hundreds or thousands of fold after the cell is stimulated which allows us to study the steps required for all stages of inducible transcription. Specifically, our lab has recently described an RNA sequencing (RNA-seq) approach involving nascent RNA transcript purification from chromatin as a means of monitoring active transcription in lipid A-stimulated macrophages. This approach has revealed, with high-resolution, on-going transcription across clusters of genes displaying similar magnitudes and kinetics of induction. These co-expressed genes have been further scrutinized with computational analyses of promoter sequences and subsequent functional approaches to identify TFs that may be responsible for their coordinated expression. The fine resolution of chromatin RNA-seq has provided the level of detail necessary for deconstruction of the regulatory logic governing target gene activation, and this approach can be easily adapted to a variety of cell types and experimental settings, and would be extremely difficult to perform in human cells.

Chromatin dynamics during transcriptional responses

Previous work from our lab and others has resulted in an advanced understanding of how DNA, chromatin, and inducible TFs work together to shape the immune response following stimulation. Of the approximately thousand genes that are significantly induced above 2-fold in BMDM, a small proportion shows robust activation exceeding 10-fold. Of the > 10-fold induced genes, many are induced very rapidly in response to stimulation, often within the first hour. The early kinetics of activation for these genes are consistent with them being the direct targets of inducible TFs, and subsequent experiments conducted in the presence of a mammalian antibiotic, cyclohexamide (CHX), demonstrated no effect for the majority of the rapidly induced genes. These genes are termed primary response genes (PRGs) because they do not require any new proteins to be synthesized to become active, thus their CHX resistance. The other

inducible genes that failed to be induced in the presence of CHX need new proteins to be made in order to become activated and tend to be induced later during the course of stimulation are are termed secondary response genes (SRGs). Further evaluation revealed that PRGs more frequently contained CpG islands within their promoters, while this feature was less common in SRGs. Furthermore, SRGs were much more likely to require the presence of the chromatin remodeling complex containing Brg1/Brm, as siRNA knockdown of these subunits resulted in diminished expression for many SRGs and a much smaller number of PRGs. Of the PRGs that were affected by Brg1/Brm1 knockdown, all but two contained low CpG density within their promoters. The association of high CpG content at primary response genes lead to the hypothesis that the intrinsic DNA properties of CpG island promoters inhibits stable nucleosome assembly and allows for open chromatin and resistance to Brg1/Brm knockdown. Most SRGs would be expected to have stable nucleosome assembly due to their low CpG content. This hypothesis was investigated using an in-vitro assembly reaction where it was shown that high CpG content promoters had less affinity for associating with recombinant histone octomers while low CpG content sequences more efficiently assembled into nucleosomes.

While the general trend is true that most PRGs contain CpG islands and do not require chromatin remodeling enzymes for their maximal expression, there were a number of PRGs where this trend did not hold true. Many of these genes are known targets of IRF3 activation, but it remained unknown to what extent IRF3 could contribute to chromatin remodeling at these sites. A restriction enzyme accessibility assay was used to monitor chromatin changes at a few of the PRGs with dependence on chromatin remodeling and IRF3. BMDM from mice lacking IRF3 (*Irf3-/-)* were unable to remodel the promoters of *Ccl5* and *Ifit1,* implicating IRF3 in the regulation of these genes through its effects on chromatin remodeling (Ramirez-Carrozzi et al., 2009).

More recently, chromatin-associated RNA (caRNA) sequencing was performed every 5-minutes during the first hour of the lipid A response in order to obtain a fine kinetic resolution of

the activation kinetics of the most potently induced genes (Tong et al., 2016). While the kinetic clustering was of use in grouping together large groups of genes displaying similar expression profiles, there was considerable heterogeneity of expression differences within genes classified together as co-regulated. Similarly, when promoter motif analysis were done in the different classified groups, significant enrichments could be observed, but manual inspection of genes classified together often revealed some genes with strong consensus sites while others had low strength motifs, indicating multiple mechanisms of activation may be used to regulate genes belonging to the same kinetic classification. For this reason, a gene-centric approach of classification was used in lieu of a statistical classification scheme. For this analysis, gene promoters were scored based on the presence of strong TF motifs in combination with chromatin immunoprecipitation (ChIP) sequencing of TFs to identify sites of binding. Both motif strength and motif score were used to infer functional binding events, and genes sharing similar TF dependence for activation were grouped together. Further support of the classifications came from genetic deletions of key TFs (*RELA, IRF3, MYD88, TRIF)* as well as through chemical inhibition of signaling pathways and alternative stimuli (Map Kinase inhibitors, TLR2 ligand). The gene-centric analysis resulted in the classification of only 5 genes that showed strong dependence on both RelA and IRF3 TFs with strong motifs that were spaced within 55 base pairs of one another. All of this would suggest that the 5 genes should display similar activation profiles since they are likely co-regulated by the same two TFs, but the kinetic profiles displayed heterogeneity in terms of when these genes become activated. *Ccl5* displayed the slowest activation kinetics, and consistent with this, RelA binding to the promoter was delayed until 30 minutes following stimulation, but was bound to all other target genes by 15 minutes. This result suggested that RelA binding was dependent on remodeling by IRF3, but it remained unknown if remodeling was a unique characteristic among the 5 genes within this classification, and whether the other 4 genes within this gene group similarly required IRF3 for remodeling and subsequent RelA binding. A genome-wide accessibility assay was performed to investigate

these possibilities using the Assay for Transposase Accessible Chromatin (ATAC)-Sequencing, which provides information regarding open chromatin during the lipid A stimulation time course. Surprisingly, of the 132 primary response genes classified, the 5 within the NFKB-IRF3 dependent class showed a distinct ability to remodel their promoters. To address whether the binding of RelA to the remodeled promoters required IRF3, ChIP-qPCR was performed on wild type cells and cells lacking IRF3. Remarkably, RelA binding was significantly decreased at the *Ccl5* promoter, and partially reduced at the *Ifnb1* promoter at late time points. These results indicate that even though small groups of genes may require the same two TFs, unique mechanisms may exist and require gene-centric approaches in the analysis of genomics data.

The entire analysis discussed above is highlighted in Appendix A of this thesis. It is important to note that while this analysis resulted in a classification scheme that can be used to guide and inform future work, the complete picture of mechanisms responsible for coordinating the LPS response is far from complete. The gene-centric approach from above relied upon a close examination of promoter elements to guide classification of regulatory factors required for transcription. This has been a powerful approach for understanding direct and immediate regulators, but gene regulation strategies extend far beyond the core promoter elements. It has been known since the early 1908s that distal DNA sequences in a plasmid construct could influence the behavior of genes in an orientation- and distance-independent manner (Banerji et al., 1981). These “enhancer” elements have been found embedded in mammalian genomes, where they associate with marks of open chromatin. Enhancers are thought to regulate genes via the recruitment of TFs that can interact with the target gene promoter via looping, tracking, or through protein-protein interactions. An exciting result of the prior analysis was the promoter remodeling was a rare event for PRGs, but it remains to be seen how frequently enhancer elements near PRGs remodel and how inducible remodeling genome-wide could be used to control gene regulation. Future analyses regarding these questions are discussed in chapter 4 of this thesis.

Negative regulation of inflammatory signaling

While it is critical for cells to be able to induced pro-inflammatory cytokines after an encounter with bacterial or viral product, it is perhaps even more important to be able to shut off inflammatory pathways once no longer needed. Indeed, many human diseases arise from an overactive immune response, and a large proportion of clinical drugs are used to limit excessive inflammation. Our cells have evolved ways to prevent sustained inflammation, and many inducible cytokines have anti-inflammatory properties. Since its original discovery over 25 years ago, the cytokine synthesis inhibitory factor (CSIF or Interleukin 10, IL10) has received much attention due to its potent inhibitory effects on the innate and adaptive immune system. In vitro studies have documented IL10's ability to inhibit the synthesis of *Ifng*, *Tnf*, and *Il12b* from Th1 cells while having limited effects on Th2-related cytokine production, indicating specificity in its inhibitory actions (Fiorentino et al., 1989). Furthermore, IL-10 functions in a non-redundant fashion in vivo, as genetic ablation of IL-10 protein leads to inflammatory bowel disease in both mice and humans (Glocker et al., 2009). IL-10 signaling involves a Janus Kinase / Signal Transducer of Activated Transcription (JAK/STAT) cascade where the IL10 receptor maintains constitutively associated JAK1 and TYK2 tyrosine kinases. These kinases become activated when IL-10 binds to the receptor and STAT3 is recruited and phosphorylated by the tyrosine kinases. Activated STAT3 proteins dimerize and enter the nucleus to drive transcription. The immediate targets of IL-10 induced STAT3 include many genes with putative roles as negative regulators of the immune response. Mice with a conditional knockout of STAT3 in macrophages fail to respond to IL-10 treatment, suggesting the STAT3 is essential for transmitting the IL-10 signal for subsequent gene repression (Takeda et al., 1999). Furthermore, early reports have indicated that STAT3 acts indirectly through the synthesis of new proteins, as IL-10's inhibition of selected target genes was abolished when cells were pre-incubated with CHX (Aste-Amezaga et al., 1998). Because of the observations that STAT3 is required and acts indirectly

for IL-10's inhibitory actions, the mechanisms involved in IL-10 signaling and gene inhibition have been largely focused on the identification of early STAT3 target genes that have the potential to feed forward as negative regulators (El Kasmi et al., 2007). These studies have been difficult to perform due to potential redundancies among IL10-inducible negative regulators. As an alternative strategy for identifying mechanism of IL-10 mediate gene repression, we have attempted to first survey the degree of inhibition across gene groups previously classified according to a gene-centric approach. The findings that emerged from an analysis of IL10-regulated genes are discussed in greater detail in chapters 3 and 4 of this thesis.

Figure Legends

Figure 1-1: Chromatin Regulation in Transcription

While it is critical for cells to be able to induced pro-inflammatory cytokines after an encounter with bacterial or viral product, it is perhaps

Figure 1-2: Structure of TLR4:MD2:LPS Interactions

Top (a) and side (b) view of the TLR4:MD2 dimer with LPS bound. The lipid A component of LPS is colored in red.

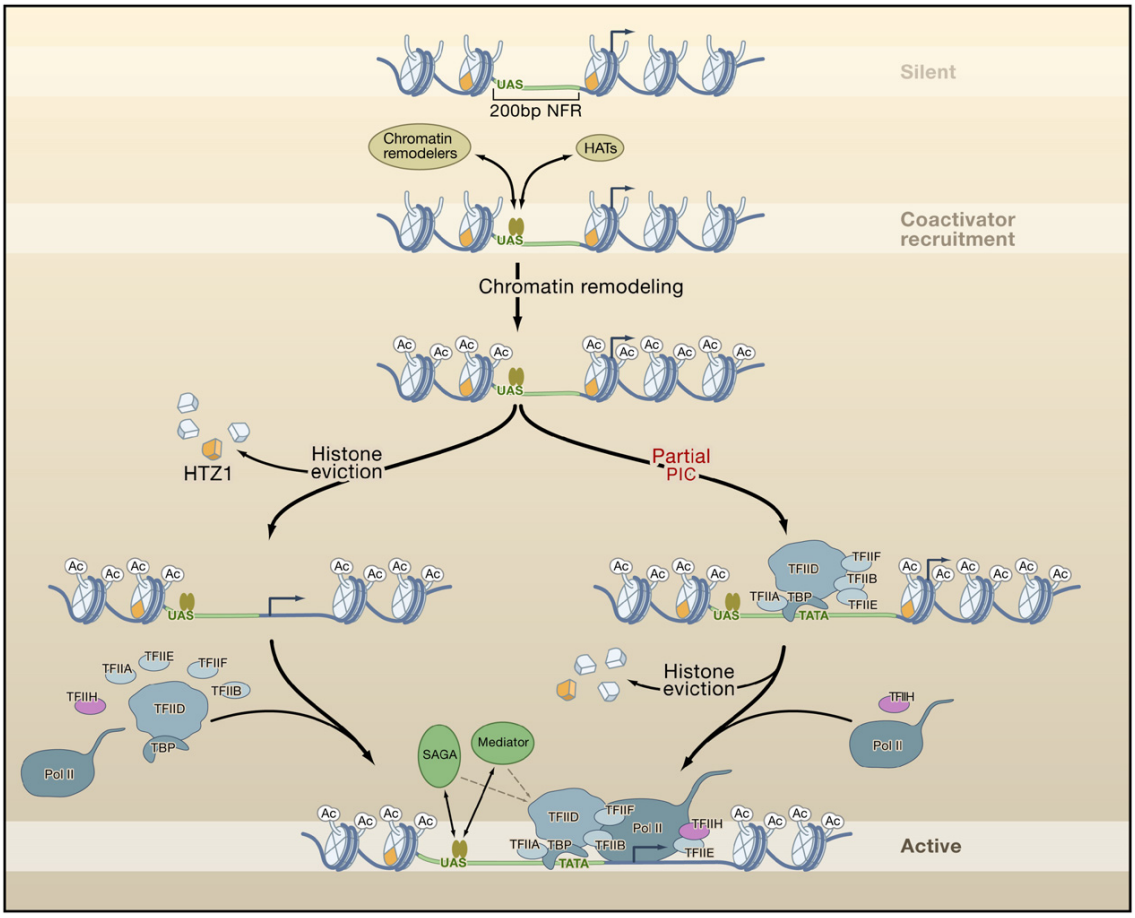
Figure 1-3: Phagocytosis in Response to Injury

Drawings reproduced from Metchnikoff during a time course of cellular injury depicting the recruitment of phagocytes to a site of inflammation.

Table 1-1: Patterns of Histone Modification and Influence on Transcription

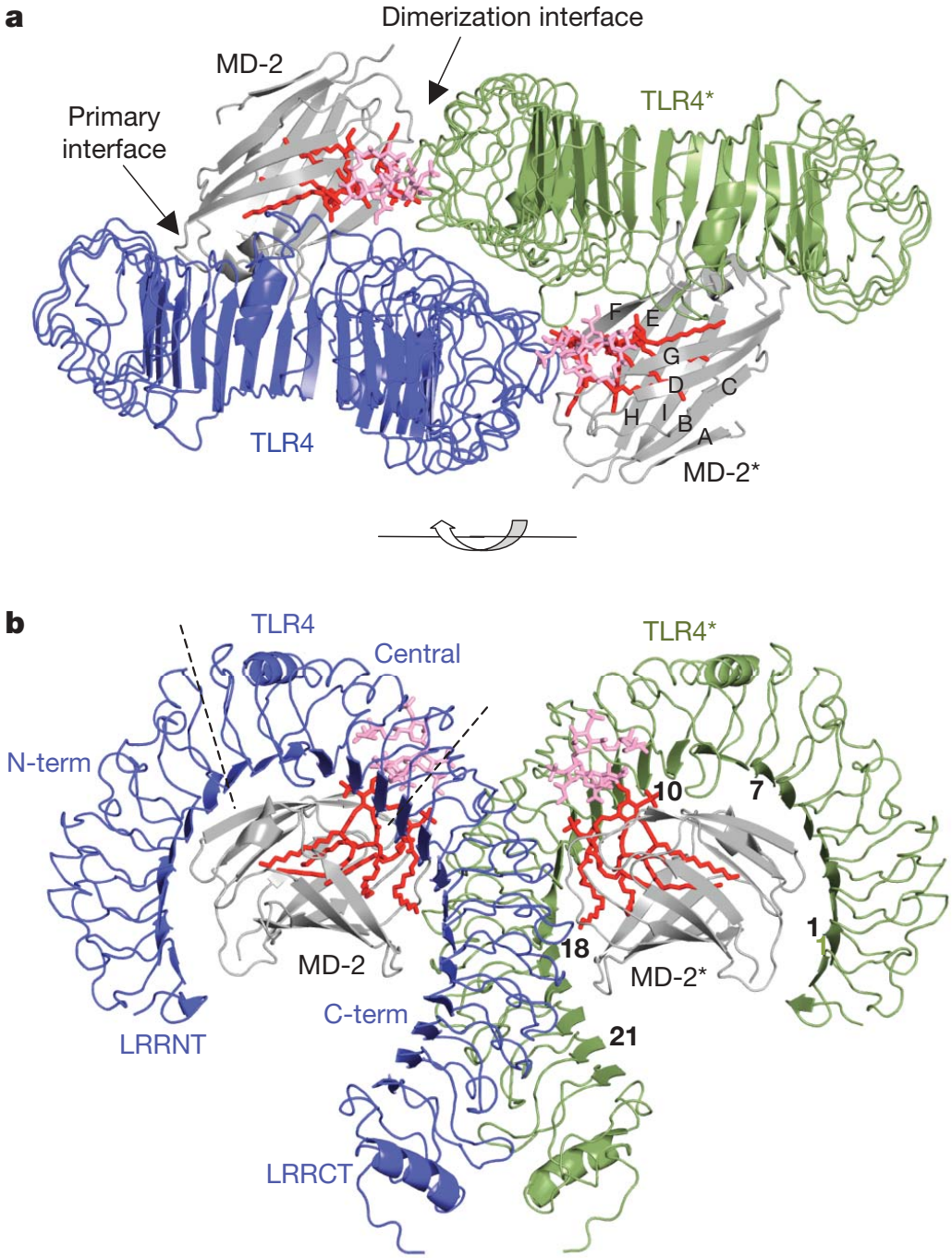
Histone modifications are listed with the enzymes responsible for generating modifications, for recognizing modifications, and their association with transcription.

Figure 1-1: Chromatin Regulation in Transcription



(Li et al., 2007)

Figure 1-2: Structure of TLR4:MD2:LPS Interactions



(Park et al., 2009)

Figure 1-3: Phagocytosis in Response to Injury

Nature Reviews | Molecular Cell Biology

(Tauber, 2003)

Table 1-1: Patterns of Histone Modification and Influence on Transcription

Modifications	Position		Enzymes				Recognition Module(s)[a]	Functions in Transcription
			S. cerevisiae	*S. pombe*	*Drosophila*	Mammals		
Methylation	H3	K4	Set1	Set1	Trx, Ash1	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	PHD, Chromo, WD-40	Activation
		K9	n/a	Clr4	Su(var)3-9, Ash1	Suv39h, G9a, Eu-HMTase I, ESET, SETBD1	Chromo (HP1)	Repression, activation
		K27				E(Z)	Ezh2, G9a	Repression
		K36	Set2			HYPB, Smyd2, NSD1	Chromo(Eaf3), JMJD	Recruiting the Rpd3S to repress internal initiation
		K79	Dot1			Dot1L	Tudor	Activation
	H4	K20		Set9	PR-Set7, Ash1	PR-Set7, SET8	Tudor	Silencing
Arg Methylation	H3	R2				CARM1		Activation
		R17				CARM1		Activation
		R26				CARM1		Activation
	H4	R3				PRMT1	(p300)	Activation
Phosphorylation	H3	S10	Snf1				(Gcn5)	Activation
Ubiquitination	H2B	K120/123	Rad6, Bre1	Rad6		UbcH6, RNF20/40	(COMPASS)	Activation
	H2A	K119				hPRC1L		Repression
Acetylation	H3	K56					(Swi/Snf)	Activation
	H4	K16	Sas2, NuA4		dMOF	hMOF	Bromodomain	Activation
	Htz1	K14	NuA4, SAGA					Activation

[a] The proteins that are indicated within the parentheses are shown to recognize the corresponding modifications but specific domains have yet to be determined.

(Li et al., 2007)

References

Aksoy, E., Albarani, V., Nguyen, M., Laes, J., Wit, D. De, Willems, F., and Goldman, M. (2007). Brief definitive report Interferon regulatory factor 3-dependent responses to lipopolysaccharide are selectively blunted in human cord blood. Med. Immunol. *109*, 2887–2894.

Aste-Amezaga, M., Ma, X., Sartori, A., and Trinchieri, G. (1998). Molecular Mechanisms of the Induction of IL-12 and Its Inhibtion by IL-10. J. Immunol. *160*, 5963–5944.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a B-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. *8*, 1499–1504.

Cairns, B.R., Kim, Y.J., Sayre, M.H., Laurent, B.C., and Kornberg, R.D. (1994). A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. Proc. Natl. Acad. Sci. U. S. A. *91*, 1950–1954.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Cavaillon, J.-M. (2011). The historical milestones in the understanding of leukocyte biology initiated by Elie Metchnikoff. J. Leukoc. Biol. *90*, 413–424.

Côté, J., Quinn, J., Workman, J.L., and Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science *265*, 53–60.

Davuluri, R. V, Grosse, I., and Zhang, M.Q. (2001). Computational identification of promoters and first exons in the human genome. Nat. Genet. *29*, 412–417.

Fiorentino, D.F., Bond, M.W., and Mosmann, T.R. (1989). Two types of mouse T helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones. J. Exp. Med. *170*, 2081–2095.

Franklin, R.E., and Gosling, R.G. (1953). Molecular Configuration in Sodium Thymonucleate. Nature *171*, 740–741.

Glocker, E.O., Kotlarz, D., Boztug, K., Gertz, E.M., Schaffer, A.A., Noyan, F., Perro, M., Diestelhorst, J., Allroth, A., Murugan, D., et al. (2009). Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N. Engl. J. Med. *361*, 2033–2045.

Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. J. Biol. Chem. *274*, 5895–5900.

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

Haufe, C. (2013). Why do funding agencies favor hypothesis testing? Stud. Hist. Philos. Sci. Part A *44*, 363–374.

El Kasmi, K.C., Smith, A.M., Williams, L., Neale, G., Panopolous, A., Watowich, S.S., Hacker, H., Foxwell, B.M.J., and Murray, P.J. (2007). Cutting Edge: A Transcriptional Repressor and Corepressor Induced by the STAT3-Regulated Anti-Inflammatory Signaling Pathway. J. Immunol. *179*, 7215–7219.

Kollmann, T.R., Kampmann, B., Mazmanian, S.K., Marchant, A., and Levy, O. (2017). Protecting the Newborn and Young Infant from Infectious Diseases: Lessons from Immune Ontogeny. Immunity *46*, 350–363.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., et al. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol. Cell. Biol. *23*, 4207–4218.

Levitz, S.M., and Golenbock, D.T. (2012). Beyond empiricism: Informing vaccine development through innate immunity research. Cell *148*, 1284–1292.

Li, B., Carey, M., and Workman, J.L. (2007). The Role of Chromatin during Transcription. Cell *128*, 707–719.

Meselson, M., and Weigle, J.J. (1961). Chromosome Breakage Accompanying Genetic Recombination in Bacteriophage*†. Proc. Natl. Acad. Sci. U. S. A. *47*, 857–868.

Mojica, F.J.M., Juez, G., and Rodriguez-Valera, F. (1993). Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol. Microbiol. *9*, 613–621.

Nabel, G.J. (2013). Designing tomorrow's vaccines. N. Engl. J. Med. *368*, 551–560.
Nuzzo, R. (2014). Statistical errors: P values, the "gold standard" of statistical validity, are not as reliable as many scientists assume. Nature *506*, 150–152.

Park, B.S., Song, D.H., Kim, H.M., Choi, B., Lee, H., and Lee, J. (2009). The structural basis of lipopolysaccharide recognition by the TLR4 – MD-2 complex. Nature *458*, 1191–1195.

Peterson, C.L., Dingwall, A., and Scott, M.P. (1994). Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. Proc. Natl. Acad. Sci. U. S. A. *91*, 2905–2908.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., dela Cruz, C.C.,

P., O.A., Panning, B., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science (80-.). *298*, 1039–1043.

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.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J. (2002). SETDB1: A novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. *16*, 919–932.

Siegrist, C.A. (2001). Neonatal and early life vaccinology. Vaccine *19*, 3331–3346.

Takeda, K., Clausen, B.E., Kaisho, T., Tsujimura, T., Terada, N., Förster, I., and Akira, S. (1999). Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of stat3 in macrophages and neutrophils. Immunity *10*, 39–49.

Tauber, A.I. (2003). Metchnikoff and the phagocytosis theory. Nat. Rev. Mol. Cell Biol. *4*, 897–901.

Tong, A.-J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. Cell 1–15.

Watson, J.D., and Crick, F.H.C. (1953). Molecular structure of nucleic acids. Nature *171*, 737–738.

Wilkins, M.H.F., Stokes, a R., and Wilson, H.R. (1953). Molecular structure of deoxypentose nucleic acids. Nature *171*, 738–740.

CHAPTER 2

Age-Related Gene Expression Difference in monocytes from Human Neonates, Young Adults, and Older Adults

PLOS | ONE

RESEARCH ARTICLE

Age-Related Gene Expression Differences in Monocytes from Human Neonates, Young Adults, and Older Adults

Michelle M. Lissner[1☯], Brandon J. Thomas[1☯], Kathleen Wee[2☯], Ann-Jay Tong[1], Tobias R. Kollmann[2‡*], Stephen T. Smale[1‡*]

1 Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, Los Angeles, California, United States of America, **2** Division of Infectious and Immunological Diseases, Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* tkollmann@cw.bc.ca (TRK); smale@mednet.ucla.edu (STS)

Citation: Lissner MM, Thomas BJ, Wee K, Tong A-J, Kollmann TR, Smale ST (2015) Age-Related Gene Expression Differences in Monocytes from Human Neonates, Young Adults, and Older Adults. PLoS ONE 10(7): e0132061. doi:10.1371/journal.pone.0132061

Editor: Degui Zhi, University of Alabama at Birmingham, UNITED STATES

Received: March 2, 2015

Accepted: June 9, 2015

Published: July 6, 2015

Data Availability Statement: All RNA sequencing files are available at the Gene Expression Omnibus (GEO) database repository (www.ncbi.nlm.nih.gov/geo) under accession number GSE60216. All other relevant data are within this paper and its Supporting Information files.

Funding: This study was funded by National Institutes of Health (www.nih.gov) grants R01GM086372 (S.T.S.), P50AR063030 (S.T.S.), T32GM008042 (B.J.T.), and T32CA009120 (A.J.T.), and also by grants from the Canadian Institutes of Health Research (www.cihr-irsc.gc.ca) (K.W., T.R.K.),

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 more 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

Burroughs Wellcome Fund (www.bwfund.org) (T.R.K.), British Columbia Children's Hospital Foundation (www.bcchf.ca) (T.R.K.), and the Michael Smith Foundation for Health Research (www.msfhr.org) (T.R.K.). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

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

This study was specifically approved by the Research Ethics Board of the University of British Columbia (Protocol H13-00347). Informed written consent from all enrolled adult participants or the next of kin, care givers, or guardians on the behalf of the cord blood participants involved in our study was obtained for all study participants. Animal research for this study was specifically approved by the UCLA Chancellor's Animal Research Committee (Protocol 1999-073-53E).

Samples of cord blood from healthy, full-term elective Caesarean sections without labor and samples of healthy young adult (ages 19–45) and older adult (aged 65 and older) peripheral blood were collected directly into sodium heparin-containing vacutainers (BD Biosciences). Mononuclear cells were isolated by density gradient centrifugation within two hours of blood collection to avoid alterations of cell properties [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°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 immediately stimulated with LPS (10 ng/ml) (InvivoGen tlrl-eblps) for 0, 1, and 6 hrs, or were infected with *Lm* at MOI = 5 for 0, 2, and 6 hrs. These time points were selected on the basis of pilot experiments, which showed that they capture the first and second major waves of gene activation in response to LPS stimulation and *Lm* infection. Wild-type (WT) *Lm* strain 10403s was provided by Dr. D. Portnoy (University of California, Berkeley, CA) and grown as described [23]. For the *Lm* experiment, uninfected cells (referred to as 0-hr time point) were collected after culturing without *Lm* for 2 hrs; in contrast, the unstimulated cells in the LPS experiment were collected immediately after isolation.

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 and healthy older adults. In addition, neonatal monocytes were obtained from umbilical cord blood samples; cord blood is known to consist almost exclusively of neonatal cells [32]. Monocytes from three individuals of each age group were stimulated with LPS or infected with *Lm*. For the LPS experiments, samples were collected 0, 1, and 6 hrs post-stimulation. For the *Lm* experiments, samples were collected 0, 2, and 6 hrs post-infection. 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 (Fig 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 Fig 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

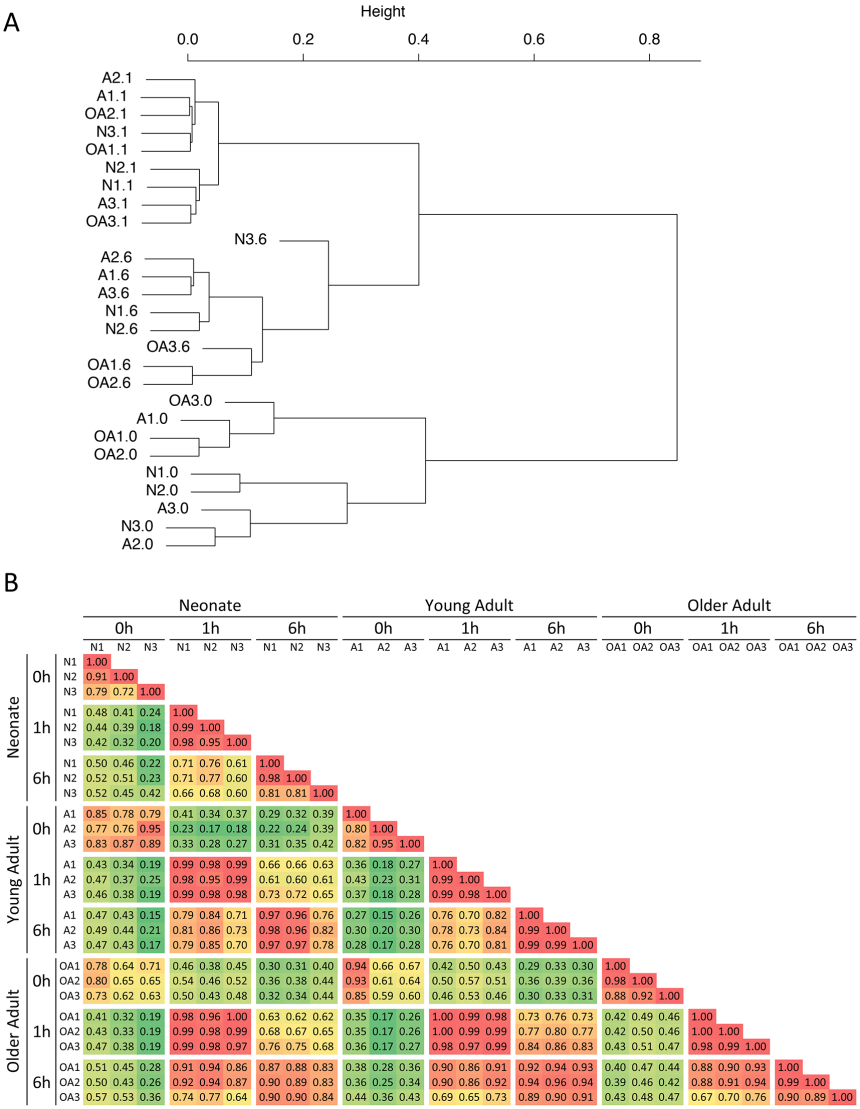


Fig 1. Hierarchical clustering of LPS-stimulated monocyte transcriptomes from human neonates, adults, and older adults. (A) RNA-seq experiments were performed with three independent human monocyte samples from cord blood (N), young adult peripheral blood (A), and older adult peripheral blood (OA) stimulated 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).

doi:10.1371/journal.pone.0132061.g001

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 (Fig 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. Fig 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 (Fig 3B).

Only one cluster (Cluster I) was identified that showed substantial age-related differences (Fig 3). 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 (Fig 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 S1 Fig) were separated into groups according to the time point at which their maximum mRNA level was observed (Fig 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 (Fig 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

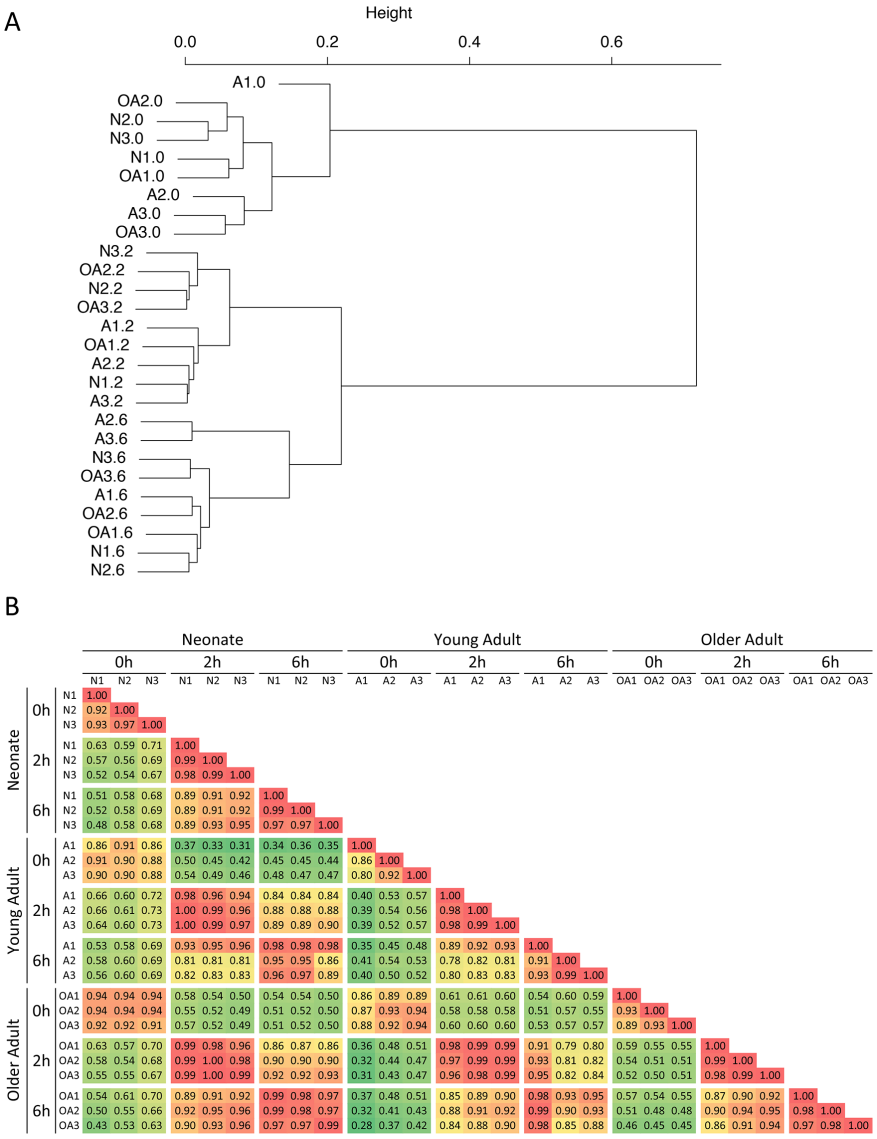


Fig 2. Hierarchical clustering of *Lm*-infected monocyte transcriptomes from human neonates, adults, and older adults. (A) RNA-seq experiments were performed with three independent human monocyte samples from cord blood (N), young adult peripheral blood (A), and older adult peripheral blood (OA) infected 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).

doi:10.1371/journal.pone.0132061.g002

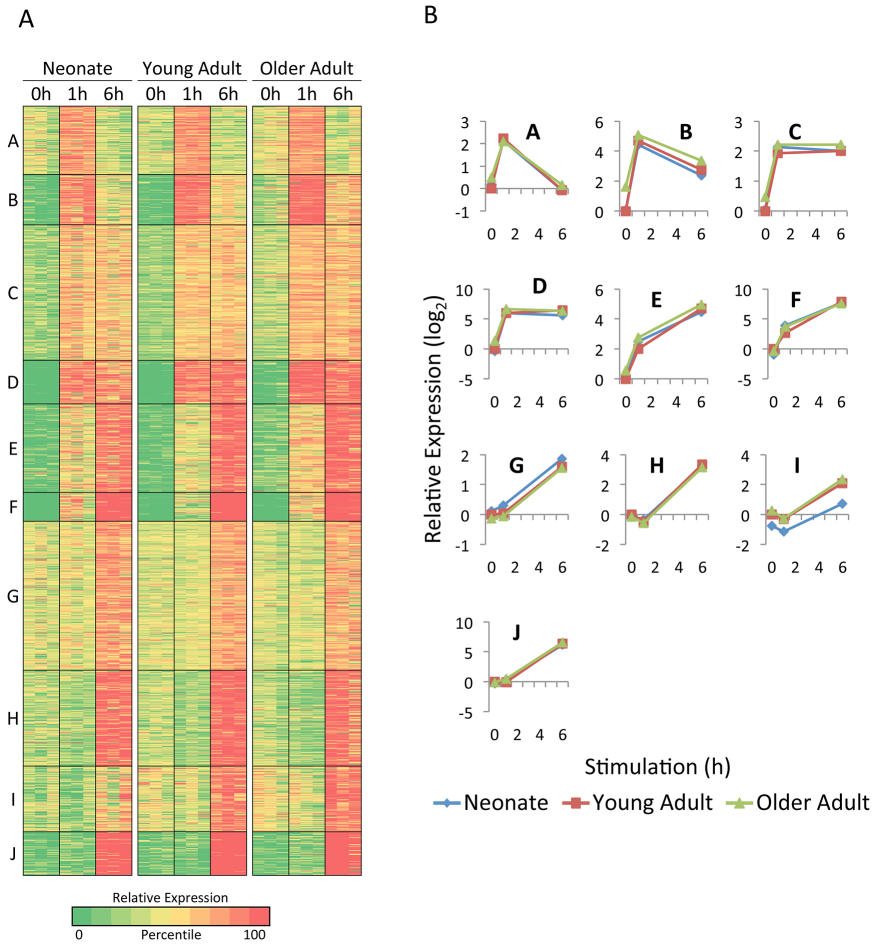


Fig 3. Analysis of LPS-induced genes in monocytes by K-means cluster analysis. (A) The 1147 genes that exceeded 200 bp in length, exhibited an RPKM of at least 4 in one sample, and were induced by LPS by at least 5-fold in the same sample were divided into 10 clusters by k-means cluster analysis, which 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).

doi:10.1371/journal.pone.0132061.g003

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 S2 Fig) 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 (Fig 6A, Groups I and V) and 110 genes that were expressed

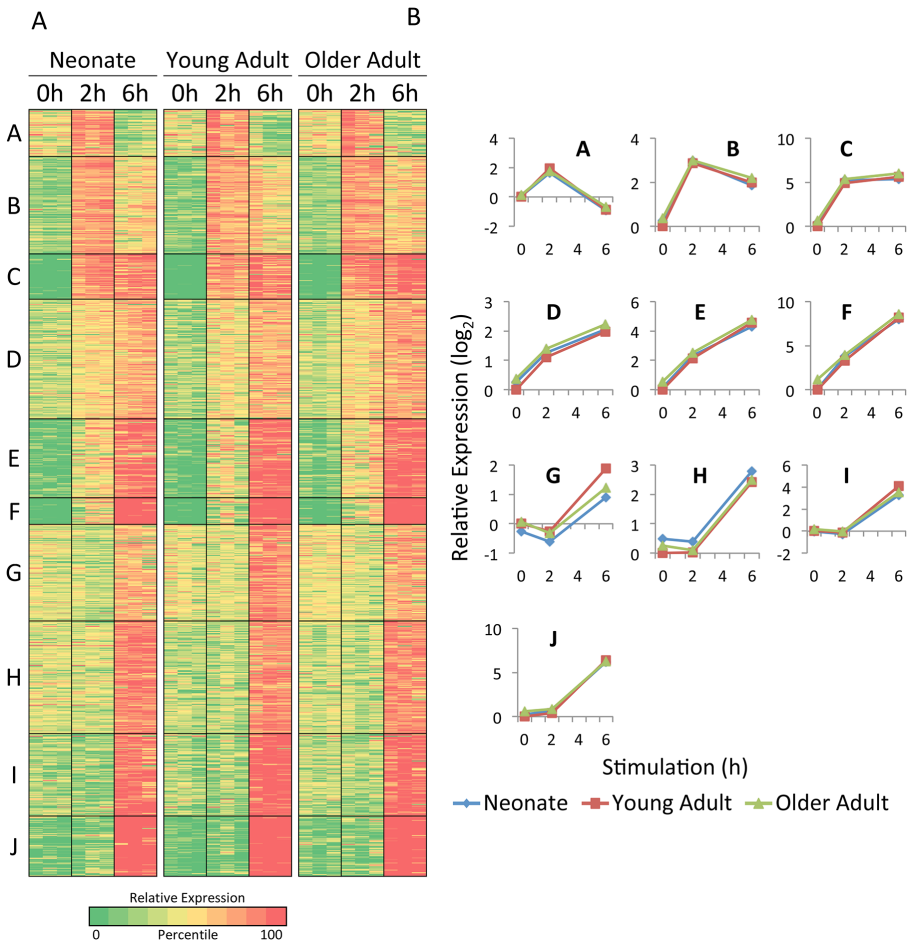


Fig 4. Analysis of *Lm*-induced genes in monocytes by K-means cluster analysis. (A) The 865 genes that exceeded 200 bp in length, exhibited an RPKM of at least 4 in one sample, and were induced by *Lm* infection by at least 5-fold in the same sample were divided into 10 clusters by k-means cluster analysis, 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).

doi:10.1371/journal.pone.0132061.g004

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).

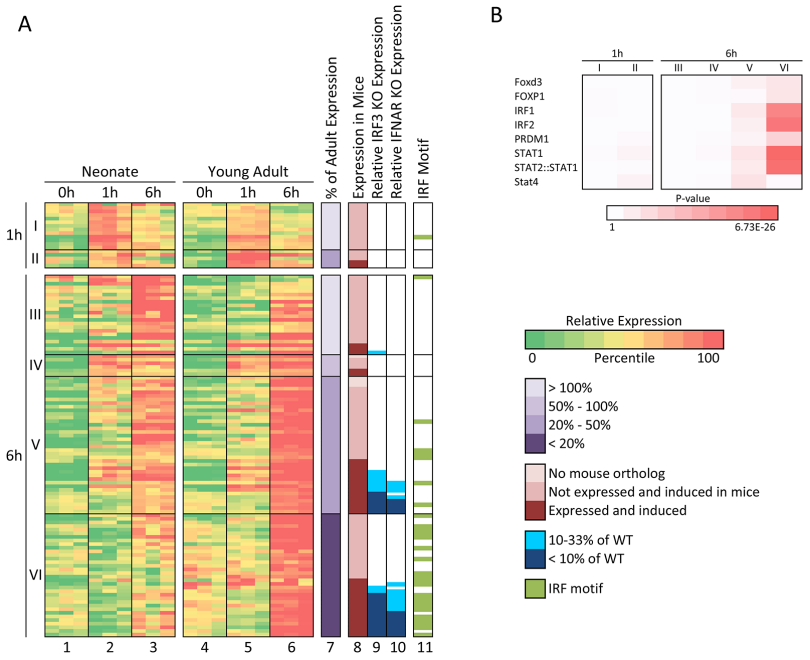


Fig 5. Genes that exhibit the greatest expression deficit in LPS-stimulated cord blood monocytes in comparison to adult monocytes are regulated by IRF3 and/or Type I IFNs. (A) LPS-induced genes exhibiting statistically significant differential expression in neonates and adults (n = 118) were grouped 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. The differential expression of six of these genes was confirmed by quantitative RT-PCR (data not shown). 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).

doi:10.1371/journal.pone.0132061.g005

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).

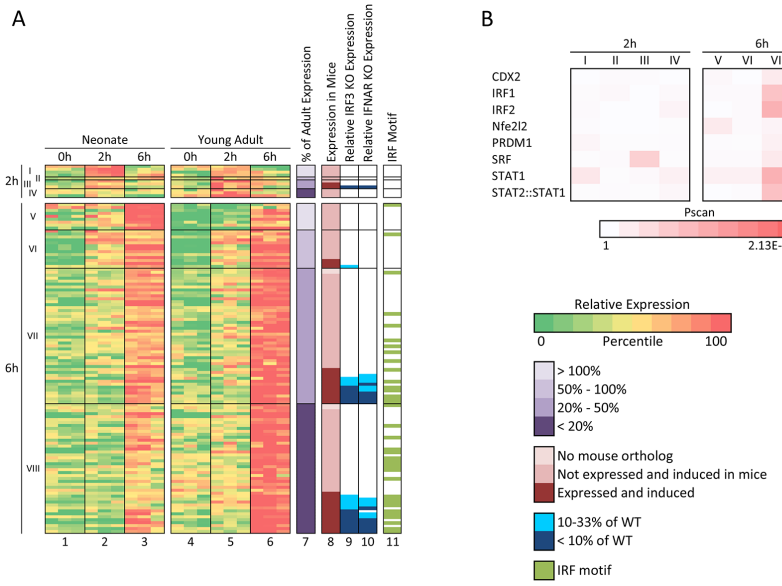


Fig 6. Genes that exhibit the greatest expression deficit in *Lm*-stimulated cord blood monocytes in comparison to adult monocytes are regulated by IRF3 and/or Type I IFNs. (A) *Lm*-induced genes exhibiting statistically significant differential expression in neonates and young adults (n = 123) were 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).

doi:10.1371/journal.pone.0132061.g006

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; Fig 5A, column 8, dark pink and red) were first separated from the small number of genes lacking obvious mouse orthologs (Fig 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 (Fig 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 (Fig 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 (Fig 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 Fig 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 (Fig 5B). The IRF1 and IRF2 binding sites used by the Pscan program are similar to the experimentally defined consensus IRF3 binding motif [33], 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 (Fig 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 (Fig 6A). 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 (Fig 6A and 6B). Thus, although IRF3 is thought to be activated by different pathways in LPS-stimulated and *Lm*-infected cells [34,35], 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 smaller differences in transcriptional induction than were observed when comparing the neonatal and young adult profiles (data not shown), suggesting that the pathways involved in the responses to LPS and *Lm* in monocytes in young adults and older adults 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 (Fig 7A; gene list in S3 Fig). For these 189 genes, Fig 7A, column 7 shows the ratio of the unstimulated transcript level in older adults to that in younger adults

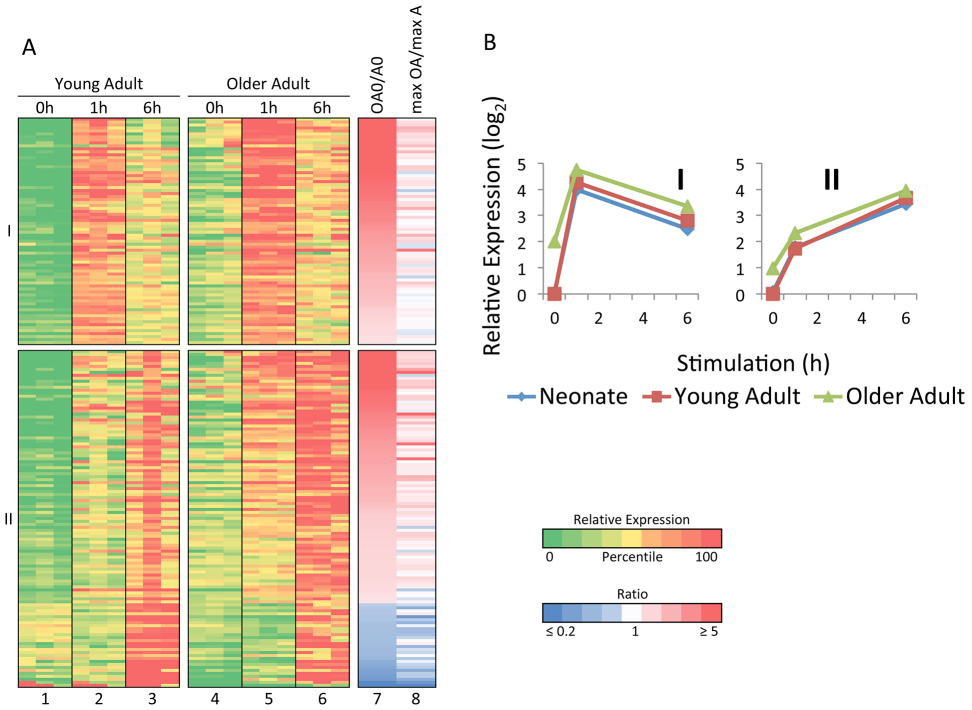


Fig 7. Elevated expression of a broad range of inflammatory genes prior to stimulation of freshly isolated monocytes from older adults. (A) LPS-induced genes exhibiting differential basal expression between adults and older adults (n = 189) are grouped according to maximum mRNA level. Columns 7 and 8 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.

doi:10.1371/journal.pone.0132061.g007

(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 unstimulated cells from 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 Fig 7A, 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 many genes that are

differentially expressed prior to stimulation. Fig 7B, 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. It is important to note, however, that although the differences between young adults and older adults are smaller after stimulation, many of these inflammatory genes are still expressed slightly higher after stimulation in the older adult samples than in the young adult samples. This differential expression causes the older adult 6-hr time point samples to cluster separately from the young adult and neonatal 6-hr samples in Fig 1A, although the separation is smaller that than observed with the unstimulated samples (see Fig 1A). 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 [36]. 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* [34,35]–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 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.

Supporting Information

S1 Fig. LPS-induced genes exhibiting statistically significant differences in transcript levels in cord blood and young adult monocytes. An expanded version of Fig 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.
(TIF)

S2 Fig. *Lm*-induced genes exhibiting statistically significant differences in transcript levels in cord blood and young adult monocytes. An expanded version of Fig 6A 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.
(TIF)

S3 Fig. LPS-induced genes that exhibit statistically significant differences in basal transcript levels in monocytes from young and older adults. An expanded version of Fig 7A 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.
(TIF)

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. Wrote the paper: MML BJT KW TRK STS.

References

1. Prabhudas M, Adkins B, Gans H, King C, Levy O, Ramilo O, et al. Challenges in infant immunity: implications for responses to infection and vaccines. Nat Immunol. 2011; 12: 189–194. doi: 10.1038/ni0311-189 PMID: 21321588
2. Kollmann TR, Levy O, Montgomery RR, Goriely S. Innate immune function by Toll-like receptors: distinct responses in newborns and the elderly. Immunity. 2012; 37: 771–783. doi: 10.1016/j.immuni.2012.10.014 PMID: 23159225
3. Levy O, Goriely S, Kollmann TR. Immune response to vaccine adjuvants during the first year of life. Vaccine. 2013; 31: 2500–2505. doi: 10.1016/j.vaccine.2012.10.016 PMID: 23085363
4. De Wit D, Tonon S, Olislagers V, Goriely S, Boutriaux M, Goldman M, et al. Impaired responses to toll-like receptor 4 and toll-like receptor 3 ligands in human cord blood. J Autoimmun. 2003; 21:277–281. PMID: 14599853
5. De Wit D, Olislagers V, Goriely S, Vermeulen F, Wagner H, Goldman M, et al. Blood plasmaytoid dendritic cell responses to CpG oligodeoxynucleotides are impaired in human newborns. Blood. 2004; 103: 1030–1032. PMID: 14504106
6. Wilson CB, Kollmann TR. Induction of antigen-specific immunity in human neonates and infants. Nestle Nutr Workshop Ser Pediatr Program. 2008; 61: 183–195. doi: 10.1159/0000113493 PMID: 18196952

7. Iliodromiti Z, Anastasiadis A, Varras M, Pappa KI, Siristatidis C, Bakoulas V, et al. Monocyte function in the fetus and the preterm neonate: immaturity combined with functional impairment. Mediators Inflamm. 2013; 2013: 753752. doi: 10.1155/2013/753752 PMID: 23690669

8. Goriely S, Vincart B, Stordeur P, Vekemans J, Willems F, Goldman M, et al. Deficient IL-12(p35) gene expression by dendritic cells derived from neonatal monocytes. J Immunol. 2001; 166: 2141–2146. PMID: 11160266

9. Vanden Eijnden S, Goriely S, De Wit D, Goldman M, Willems F. Preferential production of the IL-12 (p40)/IL-23(p19) heterodimer by dendritic cells from human newborns. Eur J Immunol. 2006; 36: 21–26. PMID: 16342235

10. Lavoie PM, Huang Q, Jolette E, Whalen M, Nuyt AM, Audibert F, et al. 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. 2010; 202: 1754–1763. doi: 10.1086/657143 PMID: 20977341

11. Corbett NP, Blimkie D, Ho KC, Cai B, Sutherland DP, Kallos A, et al. Ontogeny of Toll-like receptor mediated cytokine responses of human blood mononuclear cells. PLoS One. 2010; 5: e15041. doi: 10.1371/journal.pone.0015041 PMID: 21152080

12. Fortin CF, McDonald PP, Lesur O, Fulop T Jr. Aging and neutrophils: there is still much to do. Rejuvenation Res. 2008; 11: 873–882. doi: 10.1089/rej.2008.0750 PMID: 18847379

13. Plackett TP, Boehmer ED, Faunce DE, Kovacs EJ. Aging and innate immune cells. J Leukoc Biol. 2004; 76: 291–299. PMID: 15039467

14. Goriely S, Van Lint C, Dadkhah R, Libin M, De Wit D, Demonte D, et al. A defect in nucleosome remodeling prevents IL-12(p35) gene transcription in neonatal dendritic cells. J Exp Med. 2004; 199: 1011–1016. PMID: 15051764

15. Aksoy E, Albarani V, Nguyen M., Laes JF, Ruelle JL, De Wit D, et al. Interferon regulatory factor 3-dependent responses to lipopolysaccharide are selectively blunted in cord blood cells. Blood. 2007; 109: 2887–2893. PMID: 17138826

16. van Duin D, Mohanty S, Thomas V, Ginter S, Montgomery RR, Fikrig E, et al. Age-associated defect in human TLR-1/2 function. J Immunol. 2007; 178: 970–975. PMID: 17202359

17. Mold JE, Venkatasubrahmanyam S, Burt TD, Michaelsson J, Rivera JM, Galkina SA, et al. Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans. Science. 2010; 330: 1695–1699. doi: 10.1126/science.1196509 PMID: 21164017

18. Krow-Lucal ER, Kim CC, Burt TD, McCune JM. Distinct functional programming of human fetal and adult monocytes. Blood. 2014; 123: 1897–1904. doi: 10.1182/blood-2013-11-536094 PMID: 24518760

19. Krow-Lucal ER, McCune JM. Distinct functional programs in fetal T and myeloid lineages. Front Immunol. 2014; 5: 314. doi: 10.3389/fimmu.2014.00314 PMID: 25071775

20. Pettengill M, Robson S, Tresenriter M, Millan JL, Usheva A, Bingham T, et al. Soluble ecto-5'-nucleotidase (5'-NT), alkaline phosphatase, and adenosine deaminase (ADA1) activities in neonatal blood favor elevated elevated extracellular adenosine. J Biol Chem. 2013; 288: 27315–27326. doi: 10.1074/jbc.M113.484212 PMID: 23897810

21. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008; 18: 1509–1517. doi: 10.1101/gr.079558.108 PMID: 18550803

22. Kollmann TR, Mailman T, Bortolussi R. Listeriosis. In: Feigin R, Cherry J, Demmler-Harrison G, Kaplan S, editors. Feigin and Cherry's Textbook of Pediatric Infectious Diseases. 7th ed. Philadelphia: Elsevier Saunders; 2013: pp. 1329–1335.

23. Way SS, Kollmann TR, Hajjar AM, Wilson CB. Cutting edge: protective cell-mediated immunity to Listeria monocytogenes in the absence of myeloid differentiation factor 88. J Immunol. 2003; 171: 533–537. PMID: 12847214

24. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell. 2009; 138: 114–128. doi: 10.1016/j.cell.2009.04.020 PMID: 19596239

25. Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell. 2012; 150: 279–290. doi: 10.1016/j.cell.2012.05.043 PMID: 22817891

26. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods. 2010; 7: 709–715. doi: 10.1038/nmeth.1491 PMID: 20711195

27. Goecks J, Nekrutenko A, Taylor J, Team G. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010; 11: R86. doi: 10.1186/gb-2010-11-8-r86 PMID: 20738864

28. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28: 511–515. doi: 10.1038/nbt.1621 PMID: 20436464

29. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004; 20: 1453–1454. PMID: 14871861

30. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11: R106. doi: 10.1186/gb-2010-11-10-r106 PMID: 20979621

31. 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. 2009; 37: W247–52. doi: 10.1093/nar/gkp464 PMID: 19487240

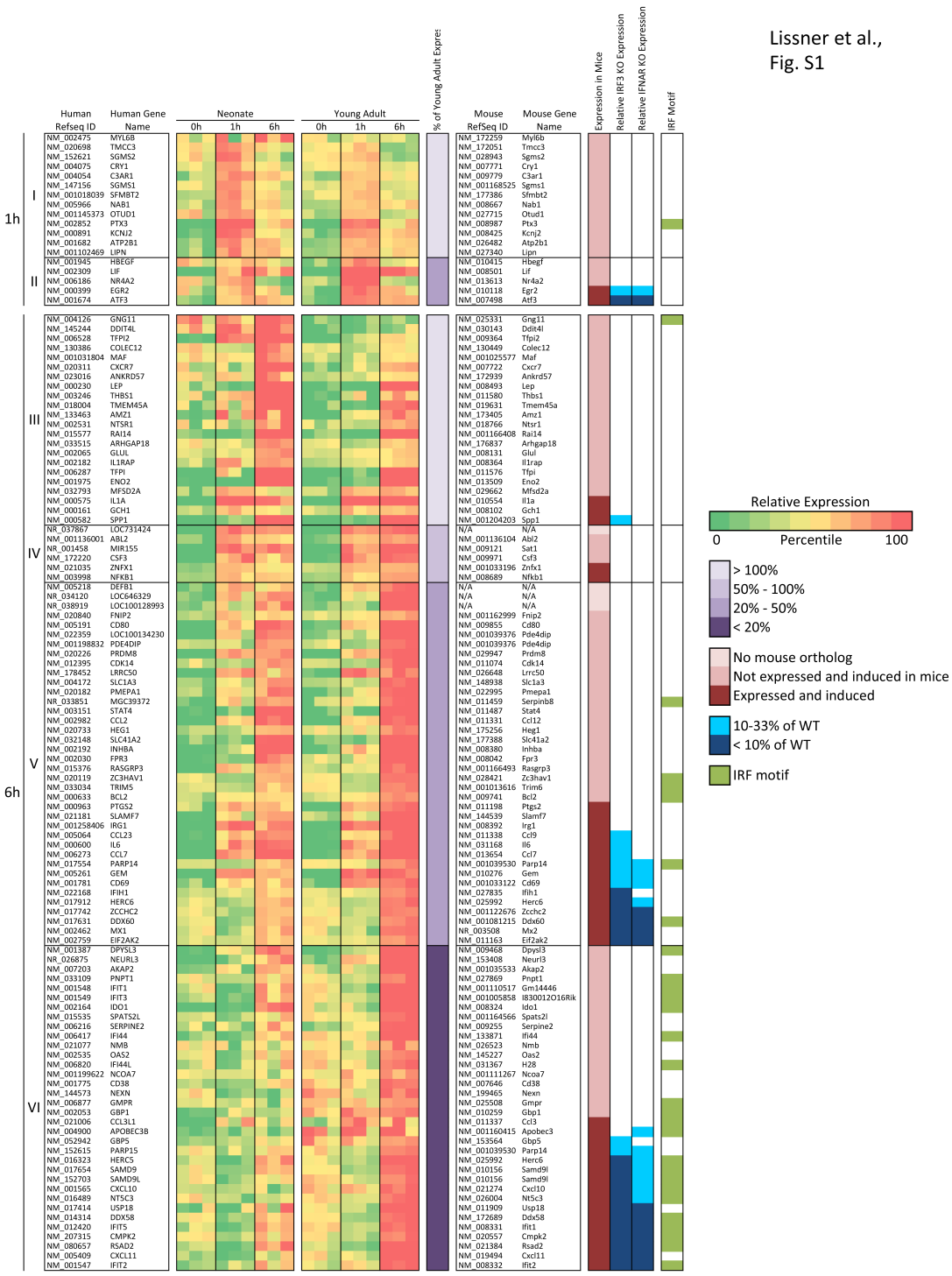
32. Hall JM, Lingenfelter P, Adams SL, Lasser D, Hansen JA, Bean MA. Detection of maternal cells in human umbilical cord blood using fluorescence in situ hybridization. Blood. 1995; 86: 2829–2832. PMID: 7545474

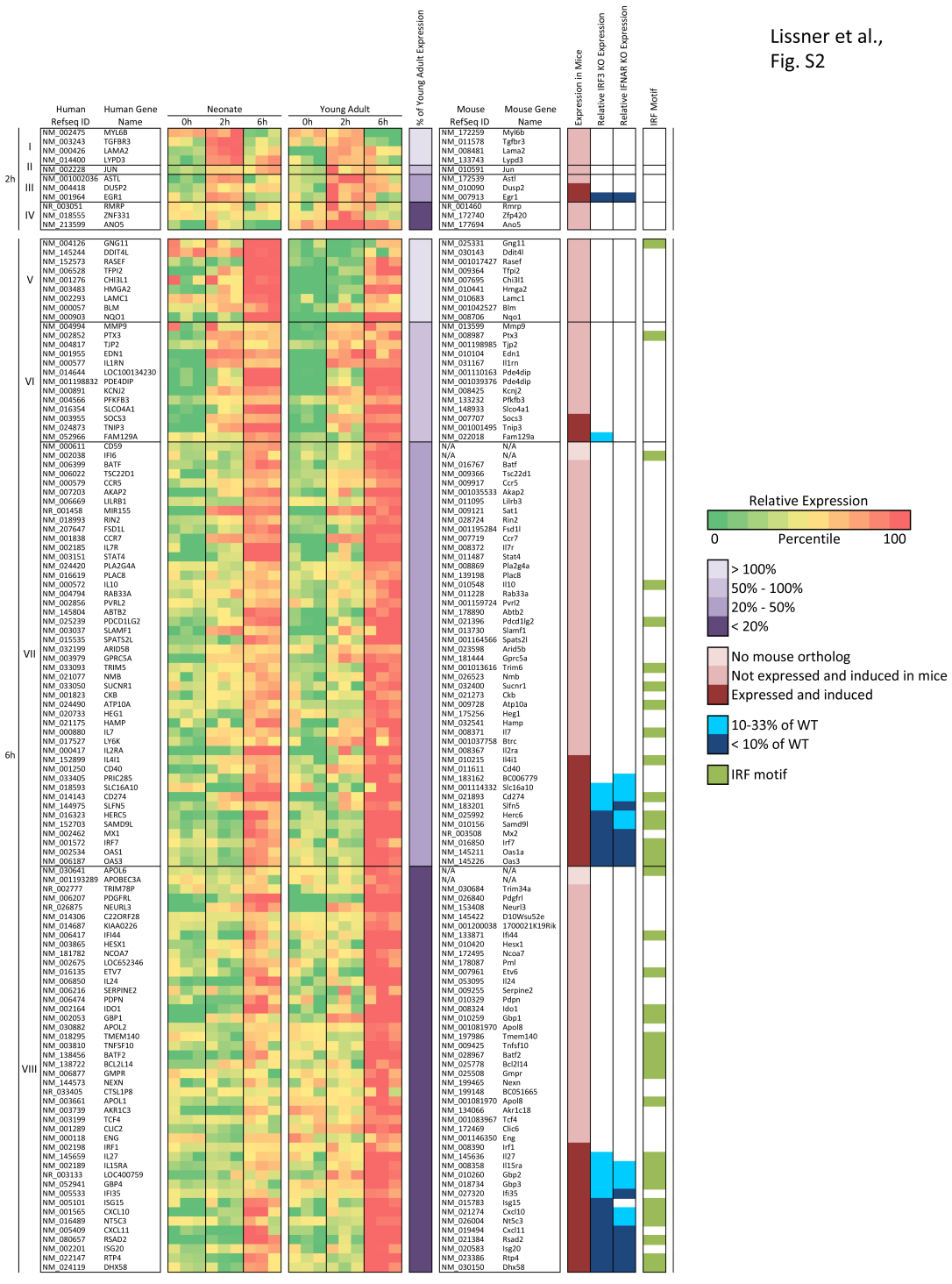
33. Lin R, Genin P, Mamane Y, Hiscott J. Selective DNA binding and association with the CREB binding protein coactivator contribute to differential activation of alpha/beta interferon genes by interferon regulatory factors 3 and 7. Mol Cell Biol. 2000; 20: 6342–6353. PMID: 10938111

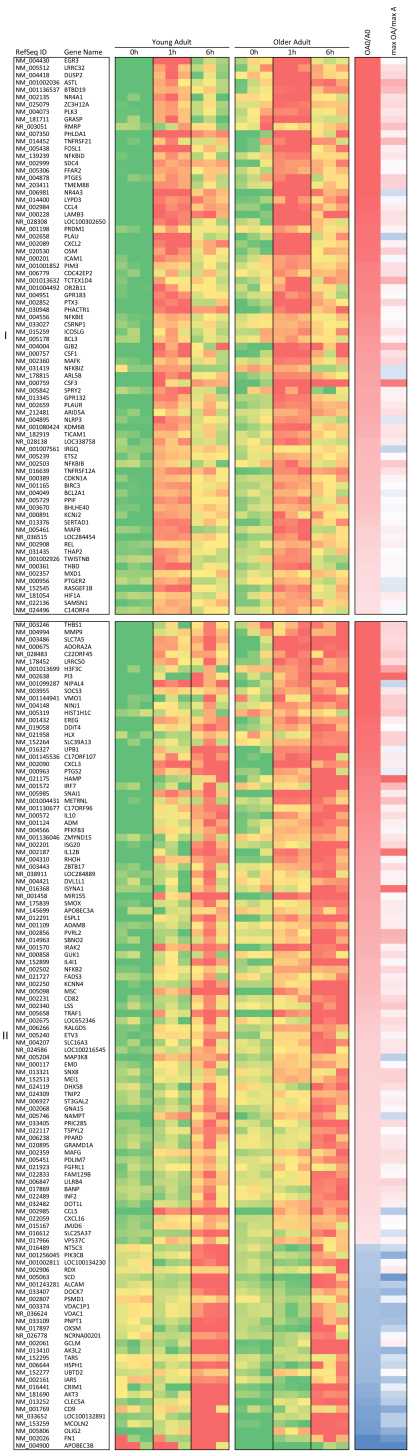
34. Witte CE, Archer KA, Rae CS, Sauer JD, Woodward JJ, Portnoy DA. Innate immune pathways triggered by Listeria monocytogenes and their role in the induction of cell-mediated immunity. Adv Immunol. 2012; 113: 135–156. doi: 10.1016/B978-0-12-394590-7.00002-6 PMID: 22244582

35. Liu S, Cai X, Wu J, Cong Q, Chen X, Li T, et al. Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induced IRF3 activation. Science. 2015; 347: 1217.

36. Danis B, George TC, Goriely S, Dutta B, Renneson J, Gatto L, et al. Interferon regulatory factor 7-mediated responses are defective in cord blood plasmacytoid dendritic cells. Eur J Immunol. 2008; 38: 507–517. doi: 10.1002/eji.200737760 PMID: 18200500


Lissner et al.,
Fig. S1
Human
Refseq ID
Human Gene
Name
Neonate
0h
1h
6h
Young Adult
0h
1h
6h
% of Young Adult Expres
Mouse
RefSeq ID
Mouse Gene
Name
Expression in Mice
Relative IRF3 KO Expression
Relative IFNAR KO Expression
IRF Motif
1h
6h
I
II
III
IV
V
VI
Relative Expression
0
Percentile
100
> 100%
50% - 100%
20% - 50%
< 20%
No mouse ortholog
Not expressed and induced in mice
Expressed and induced
10-33% of WT
< 10% of WT
IRF motif


Lissner et al.,
Fig. S2
Human RefSeq ID
Human Gene Name
Neonate
Young Adult
0h
2h
6h
% of Young Adult Expression
Mouse RefSeq ID
Mouse Gene Name
Expression in Mice
Relative IRF3 KO Expression
Relative IFNAR KO Expression
IRF Motif
2h
6h
I
II
III
IV
V
VI
VII
VIII
Relative Expression
0
Percentile
100
> 100%
50% - 100%
20% - 50%
< 20%
No mouse ortholog
Not expressed and induced in mice
Expressed and induced
10-33% of WT
< 10% of WT
IRF motif



Lissner et al.,
Fig. S3

CHAPTER 3

IL-10 Signaling Remodels Adipose Chromatin Architecture to Limit Thermogenesis and Energy Expenditure

Abstract

Signaling pathways that promote adipose tissue thermogenesis are well characterized, but the physiologic limiters of energy expenditure are largely unknown. Here we show that loss of the anti-inflammatory cytokine IL-10 improves insulin sensitivity, protects against diet-induced obesity, and elicits the browning of white adipose tissue. Mechanistic studies define bone marrow cells as the source of the IL-10 signal and mature adipocytes as the target cell type mediating these effects. IL-10 receptor alpha is highly enriched in mature adipocytes and is induced in response to cold, obesity and aging. ATAC-seq and RNA-seq reveal that IL-10 represses the transcription of thermogenic genes in adipocytes by altering chromatin accessibility and inhibiting ATF and PGC-1alpha recruitment to key enhancer regions. These findings identify the IL-10 axis as a critical and potentially targetable regulator of thermogenesis, and expand our understanding of the links between inflammatory signaling and adipose tissue function in the setting of obesity.

Introduction

Mammals contain specialized types of adipose tissue, including white adipose tissue (WAT) and brown adipose tissue (BAT) (Attie and Scherer, 2009; Rosen and Spiegelman, 2014). WAT stores energy in times of nutritional excess and its dysfunction contributes to metabolic disorders such as type 2 diabetes (Camp et al., 2002; Guilherme et al., 2008). BAT is specialized to dissipate stored chemical energy in the form of heat (Seale et al., 2009; Townsend and Tseng, 2014). BAT mass inversely correlates with body mass index and has been ascribed a potential anti-obesity function (van Marken Lichtenbelt et al., 2009; Virtanen et al., 2009). Recent studies have revealed the inherent plasticity of WAT and its ability to activate thermogenesis upon exposure to cold, thiazolidinediones, and hormonal stimuli (Ohno et al., 2012; Tiraby and Langin, 2003; Ye et al., 2013). A subpopulation of cells in inguinal WAT

(iWAT) known as “beige” cells expresses uncoupling protein 1 (UCP1) and carries out thermogenesis (Lee et al., 2012; Rosenwald et al., 2013; Wu et al., 2012). UCP1 is crucial for thermogenesis in both brown and beige adipocytes, and control of its activity contributes to regulation of body weight and energy balance (Feldmann et al., 2009).

Adipose tissue is composed of a heterogeneous cell population that includes adipocytes, endothelial cells, preadipocytes, and immune cells. Immune-adipose crosstalk has pronounced effects on the expansion and the thermogenic activation of beige adipose tissue. Recent studies have highlighted the importance of anti-inflammatory (type II) cytokines in promoting adipose thermogenesis. Production of IL4/IL13 by eosinophils upon stimulation by cold or exercise stimulates adipose tissue macrophages to secrete catecholamines and activate thermogenesis (Qiu et al., 2014; Rao et al., 2014). Activation of type 2 innate lymphoid (ILC2) cells by IL-33 acts via IL4Rα in preadipocytes to promote beige fat biogenesis (Lee et al., 2015). Furthermore, IL-33 was shown to license adipocytes for uncoupled respiration by regulating splicing of UCP1 (Odegaard et al., 2016). An ILC2-positive population was recently identified in human WAT, and activation of ILC2 by IL-33 was reported to stimulate production of a methionine- enkephalin peptide that directly acts on adipocytes to increase UCP1 expression (Brestoff et al., 2015). Despite these advances, endogenous pathways that antagonize thermogenesis in adipose tissue are relatively poorly understood.

IL-10 is a type II cytokine with anti-inflammatory properties and its loss is associated with autoimmune pathologies (Couper et al., 2008; Hutchins et al., 2013). IL-10 is secreted by multiple immune cells, including macrophages, dendritic cells, B cells, and T cells (Saraiva and O'Garra, 2010). It signals through a receptor complex of IL-10Rα and IL-10Rβ to trigger the activation of signal transducer and activator of transcription 3 (STAT3) (Moore et al., 2001). STAT3 is known to be essential for the anti-inflammatory activity of IL-10 (Lang et al., 2002), but the mechanisms by which IL-10 regulates gene expression remain very poorly understood. The biological effects of IL-10 are multifaceted, and it can simultaneously turn on and off specific

gene programs. Mechanisms proposed to underlie the anti-inflammatory actions of IL- 10 include: regulation of mRNA and protein stability, post-translational modification, microRNA-mediated RNA-interference, and transcriptional repression through intermediary proteins (El Kasmi et al., 2007; Hammer et al., 2006; Kuwata et al., 2003; Rossato et al., 2012; Schaljo et al., 2009; Smith et al., 2011). However, inhibition of transcription is widely believed to be the primary mode of IL10 anti-inflammatory actions (Murray, 2005; Murray and Smale, 2012)

The function of IL-10 in inflammation is well established, but its role in adipose biology and energy homeostasis is largely unknown. Some studies have suggested that IL-10 might function to create an anti-inflammatory milieu in adipose tissue by promoting the activity of anti-inflammatory M2 macrophages (Fujisaka et al., 2009; Gao et al., 2013; Hong et al., 2009; Lumeng et al., 2007; Nakata et al., 2016; Xie et al., 2014). However, IL10 is dispensable for M2 macrophage activation, and IL-10 loss-of-function studies have not supported an anti-obesity role for IL-10 (den Boer et al., 2006; Mauer et al., 2014; Miller et al., 2011). Furthermore, ablation of IL-10 from the hematopoietic system does not cause insulin resistance (Kowalski et al., 2011). One complicating factor in the study of the metabolic effects of IL-10 is the propensity of the *Il10*–/– mice to develop colitis and systemic inflammation (Kuhn et al., 1993). Spontaneous colitis in *Il10*–/– mice is dependent on several factors, including age, strain, and vivarium conditions; thus, these conditions may have a major impact on systemic metabolism in *Il10*–/– mice (Keubler et al., 2015).

Here we report the delineation of an unexpected role for IL-10 signaling in directing transcriptional responses in adipose tissue that limit thermogenesis. Inhibition of IL-10 action promotes energy expenditure and protects mice from diet-induced obesity. We show that bone marrow-derived IL-10 acts on adipocytes directly via IL-10Rα to repress thermogenic gene expression by altering the chromatin landscape at ATF-binding transcriptional regulatory regions. These findings challenge prior assumptions regarding the relationship between

inflammatory signaling pathways and adipose tissue function and provide insight into physiological control of thermogenesis that could inform future therapy for metabolic diseases.

Results

Ablation of IL-10 protects mice from diet-induced obesity

To begin to dissect the role of IL-10 in systemic metabolic homeostasis we analyzed young *Il10–/–*mice (10 weeks of age) on a C57B/6 background. Importantly, IL-10 deficiency on this genetic background is associated with a relatively low incidence of colitis (Keubler et al., 2015). There was no overt evidence of systemic inflammation in *Il10–/–* mice, and they had comparable body weights, serum cytokine profiles, and colon morphology to WT controls (Figure 3-8 A-D). However, on visual inspection, iWAT from *Il10–/–* mice appeared redder than that from WT littermate controls. Histological analysis revealed smaller adipocytes and increased numbers of cells with multilocular lipid droplets in iWAT from *Il10–/–* mice (Figure 3-1A). We further found that *Il10–/–* mice exhibited markedly improved glucose tolerance and had lower serum triglyceride levels (Figure 3-1B and 3-8E) compared to controls. Interestingly, ablation of IL-10 had no influence on the total macrophage or activated M1 macrophage populations in adipose tissues (Figure 3-8F).

We next addressed how this phenotype progressed as *Il10–/–* mice aged. Chow-fed *Il10–/–* mice at 8 months of age were grossly leaner than WT controls and had less total body mass and fat mass (Figure 3-1 C-E and Figure 3-8G). Furthermore, the size and weight of individual adipose depots (iWAT and eWAT) were reduced compared to controls. Livers of *Il10–/–* mice also appeared to be protected from hepatic steatosis (Figure 3-1D). We further assessed how loss of IL- 10 would affect the development of diet-induced obesity. WT and *Il10–/–* mice of 10 weeks of age were fed a high-fat-diet (HFD; 60% calories from fat) for 6 weeks. After this regimen, *Il10– /–* mice were grossly much leaner and gained markedly less weight than WT mice (Figure 3-1F and 3-1G). MRI analysis of body composition confirmed reduced

body fat in *Il10–/–* mice compared to WT controls with no difference in lean mass (Figure 3-1H). *Il10–/–* mice were also protected from diet-induced hepatic steatosis, and the size and weight of individual adipose depots were reduced compared to WT mice (Figure 3-1G-1I, 3-9C and 3-9D). Mice lacking IL-10 had improved glucose tolerance and insulin tolerance as revealed by IP glucose and insulin tolerance tests, respectively (Figures 3-1J,K). Despite the well-established role of IL-10 as a limiter of inflammatory responses, we observed no evidence of acute or chronic inflammation in *Il10–/–* mice. With the exception of the loss of IL-10 and an increase in GCSF, they had comparable levels of serum pro-inflammatory cytokines, and colon morphology and histology revealed no apparent signs of colitis (Figure 3-9 A-C). Thus, the improved metabolic phenotype of *Il10–/–* mice could not be linked with the development of colitis or obvious systemic inflammation.

Increased energy expenditure and mitochondrial respiration in IL-10-deficient mice

Adipose tissue is an important determinant of systemic energy balance. Given that the changes in adipose tissue morphology in *Il10–/–* mice were consistent with increased thermogenic activity, we probed the influence of IL-10 expression on energy homeostasis. Chow-fed mice at ten weeks of age were individually housed in metabolic chambers for 72 hours. Oxygen consumption rate (VO2) and energy expenditure (EE) were higher in *Il10–/–* mice compared to WT controls (Figure 3-2 A-C). Food consumption was also markedly increased in *Il10–/–* mice (Figure 3-2D). Metabolic cage studies performed on an independent cohort of high-fat diet-fed mice revealed similar elevations in VO2, and EE in mice lacking IL-10 (Figure 3-2 G-H). We next examined if the increase in EE in *Il10–/–* mice might reflect altered mitochondrial activity. We isolated mitochondria from iWAT from WT and *Il10–/–* mice and measured rates of oxygen consumption (OCR) to assess differences in mitochondrial respiration between the genotypes. Mitochondrial respiration was measured with substrate present (basal respiration), and in the presence of ADP (complex V respiration), or FCCP (maximal respiration) (See methods

section). Sequential treatments with these compounds revealed marked increase in basal, complex V, coupled and uncoupled respiration, and maximal mitochondrial respiration in *Il10–/–* mice compared to controls (Figure 3-2E). Next, we assessed the activity of the respiratory chain complexes I-IV by performing electron flow assays. The activity of all complexes was augmented in *Il10–/–* mitochondria compared to controls (Figure 3-2F). These findings strongly suggest that increased systemic energy utilization and enhanced mitochondrial respiration contribute to the improved metabolic profile of IL-10-deficient mice.

Increased adipose thermogenesis in IL-10 KO mice

Certain WAT depots have the capacity to undergo "beiging", and thereby increase their thermogenic activity (Ohno et al., 2012; Tiraby and Langin, 2003; Ye et al., 2013). To address whether the loss of IL-10 expression affected adipose phenotype at the level of gene expression, we performed RNA-seq on iWAT from chow-fed, ten week-old mice. Remarkably, as depicted by the volcano plot in Figure 3-3A, genes linked to adipocyte thermogenesis were highly upregulated in iWAT from *Il10–/–* mice compared to WT controls. Such genes included *Ucp1*, *Cidea* and *Pm20d1*–a recently identified secreted factor that increases thermogenesis through synthesis of lipidated amino acids (Long et al., 2016). Conversely, genes selectively expressed in WAT and those associated with obesity, including *Mmp12, Trem2, Celec4d, and Atp6v0d2* (Fujimoto et al., 2011; Lee et al., 2014; Park et al., 2015; Xu et al., 2013), were downregulated in *Il10–/–* mice. We further analyzed the correlation between the gene expression signatures of WT or *Il10–/–* iWAT and reference BAT and WAT RNA-seq data from publicly available datasets (Sun et al., 2013). The data plot in Figure 3-3B shows that the gene expression profile of *Il10–/–* iWAT more closely resembled BAT than did WT control iWAT. Quantitative analysis of the number of genes as well as percentage of all adipose subtype-selective genes shown in Figure 3-3B demonstrated a ~4.5-fold increase in brown/beige-selective markers in *Il10–/–* iWAT compared to control (Figure 3-3C and 3-3D). Comparing the

BAT/WAT ratio and *Il10–/–*/WT ratio from the dataset in Figure 3-3D showed that BAT-selective genes such as *Ucp1* and *Cidea* were highly upregulated in the absence of IL-10 (Figure 3-3E). Congruent with these data, gene ontology analysis of our RNA-seq data further showed that markers indicative of brown differentiation were highly enriched in *Il10–/–* iWAT (Figure 3-3F). Finally, we validated these bioinformatic analyses with directed real-time PCR and found that markers of adipose beiging/browning were increased in *Il10–/–* mice, whereas inflammatory markers were reduced or unchanged (Figure 3-3G). Overall, these data suggest that ablation of IL-10 in mice leads to the remodeling of iWAT to a more beige-like phenotype.

Reconstitution of IL-10 rescues the thermogenic phenotype of IL-10-deficient mice

To determine the source of the IL-10 signal leading to these metabolic effects, we performed bone-marrow transplantation (BMT) studies. We reconstituted lethally-irradiated *Il10–/–* mice with either WT bone-marrow (WT→KO) or *Il10–/–* bone-marrow (KO→KO) (Figure 3-10A). Genotyping of blood seven weeks post-BMT showed that the WT allele was fully reconstituted in *Il10–/–* mice (Figure 3-10A). Interestingly, following the transplant WT→KO mice gained more weight and accumulated more fat mass than KO→KO controls (Figure 3-4A and 3-4B). The iWAT depots were larger in WT→KO mice compared to KO→KO controls (Figure 3-4C). WT→KO mice also had higher blood glucose levels and were glucose intolerant relative to KO→KO mice (Figure 3-10B and 3-4D). Importantly, neither group showed apparent signs of colitis (Figure 3-10C).

To investigate whether the thermogenic phenotype of IL-10-deficient mice was rescued by WT bone marrow, we individually housed the transplanted mice in metabolic chambers. This analysis revealed that WT→KO mice had reduced VO2 and EE compared to KO→KO mice (Figure 3-4E). Consistent with this finding, thermogenic gene expression and mitochondrial respiration were repressed in WT→KO mice (Figure 3-4F-H). Together these data point to a hematopoietic origin of the IL-10 signal controlling thermogenesis and energy expenditure.

The IL-10 axis directly represses adipocyte thermogenesis

To explore if bone marrow-derived IL-10 could act on adipose tissue directly, we first characterized IL-10Rα expression in fat. We found that IL-10Rα was highly enriched in the mature adipocyte fraction of mouse adipose tissue and *Il10ra* but not *Il10rb* was upregulated during adipocyte differentiation (Figure 3-5 A-B and 3-11A). Furthermore, *Il10ra* expression was increased in response to cold exposure, high-fat diet, genetic obesity (*db/db* and *ob/ob* mice), and aging (Figure 3-5C and 3-11A). IL-10Rα protein was correspondingly increased during adipocyte differentiation (Figure 3-11B). Furthermore, IL-10 treatment caused activation of STAT3 (pSTAT3), demonstrating that IL-10 signaling is functional in mature adipocytes (Figure 3-11B). We also identified *Il10ra* as a direct PPARγ target gene in adipocytes. *Il10ra* expression was induced in response to PPARγ activation, and publically available ChIP-seq data from analysis of adipose depots revealed robust enrichment of PPARγ at the enhancer region of the *Il10ra* gene locus (Figure 3-11C) (Siersbaek et al., 2012). Collectively, these observations suggested that IL- 10Rα might be upregulated in adipose tissue as a negative feedback mechanism to maintain adiposity and limit thermogenesis.

To test this idea, we knocked down IL-10Rα in iWAT of WT mice using an adenoviral vector expressing an shRNA directed against *Il10ra*. Interestingly, even partial knockdown of IL10Rα protein was sufficient to cause a substantial increase in thermogenic gene expression in iWAT (Figure 3-5D and 3-5E). To further test our hypothesis, we harvested iWAT from mice and treated it *ex vivo* with vehicle or IL-10. We found that IL-10 decreased levels of UCP1 protein and its corresponding mRNA (Figure 3-5F and 3-5G). To establish that these effects were due to direct actions of IL-10 on adipocytes and did not require the presence of other cell types in adipose tissue (such as macrophages), we studied primary beige adipocytes differentiated *in vitro*. Treatment with IL-10 also led to a robust downregulation of thermogenic genes in these cells (Figure 3-11D).

To facilitate our analysis of IL-10 signaling in cultured adipocytes we established an immortalized brown/beige-like preadipocyte cell line that expressed IL-10Rα and was capable of inducing thermogenic genes in response to differentiation cocktail (iBAd cells). We confirmed that IL-10 signaling was operative in these cells by assessing the induction of the canonical IL-10-responsive gene *Socs3* (Figure 3-11E). Reciprocal with the induction of *Socs3*, we observed a dose-dependent decrease in UCP1 levels in response to IL-10. Pretreatment of the cells with an IL-10Rα antibody partially rescued the effect (Figure 3-11F). We also observed decreased mitochondrial respiration upon IL-10 treatment, consistent with the effects observed in iWAT (Figure 3-11G).

After validating our in vitro model, we assessed global gene expression in undifferentiated and differentiated iBAd cells in the presence or absence of IL-10. Genes induced in vehicle-treated cells more than 5-fold on day 5 (D5) of differentiation compared to D0 were plotted as a heat-map (Figure 3-5H). Analysis of the dataset showed that thermogenic genes such as *Ucp1, Cidea, and Pppargc1*α were among the highest induced genes on D5. Moreover, these same genes were also among the most highly inhibited by IL-10 (Figure 3-5H). Interestingly given the repressive effect of IL-10 on thermogenic gene expression, several genes that were highly induced by IL-10 encoded for transcriptional repressors including *Ncor2* and *Hdacs* (Figure 3-5H and 3-11H). Pathway analysis revealed that brown fat cell differentiation and lipid metabolic processes were highly compromised in the presence of IL-10 (Figure 3-5I). Blockade of the browning program by IL-10 was further validated by plotting the RNA-seq vehicle/IL-10 ratio as a function of WAT/BAT ratio. We found that 70% of the genes inhibited by IL-10 were brown-selective genes, suggesting a high degree of specificity for IL-10 action on the browning program (Figure 3-5J).

IL-10 affects transcription of thermogenic genes by altering chromatin architecture

The primary mode of IL-10 action in macrophages is believed to be inhibition of transcription, although the underlying mechanisms are not well understood. We set out to explore the basis for IL-10 effects on adipocyte gene expression. We found that IL-10 inhibited the abundance of primary transcripts of thermogenic genes (Figure 3-12A). Furthermore, IL-10 inhibited the expression of putative enhancer RNAs (eRNAs) from the *Ucp1* gene locus (Figure 3-12A). These findings indicated that IL-10 was acting to block transcription of certain adipocyte genes.

To test whether IL-10-dependent repression of transcription was due to action at the regulatory regions of thermogenic genes, we performed genome-wide ATAC-Seq on differentiated iBAd cells treated with vehicle or IL-10. ATAC-seq peaks correspond to genomic regions that are sensitive to cleavage by transposase due to their open chromatin configuration; these sites serve as markers of regulatory DNA sequences including enhancers and promoters (Buenrostro et al., 2013). Using a parallel approach to the RNA-seq analysis of Figure 3-5H, we identified 3174 ATAC peaks that were enriched more than five-fold on D5 compared to D0 in vehicle treated cells and represented them as a heat map (Figure 3-6A). Peaks indicative of open chromatin appearing at D5 included those at the enhancer/promoter regions of thermogenic genes such as *Ucp1and Cidea*, consistent with the induction of these genes during brown/beige differentiation. In line with the inhibitory effects of IL-10 on thermogenic gene repression, IL- 10 treatment reduced ATAC peak enrichment at the regulatory regions of thermogenic genes (Figure 3-6A).

To qualitatively assess the changes in ATAC-seq peaks, we plotted the data as a bedgraph on the UCSC genome browser. As shown in Figure 3-6B, on D5 of brown differentiation a discreet set of new peaks emerged (Peaks 1 and 3), indicative of newly opened chromatin at putative enhancer regions of the *Ucp1* gene locus. Remarkably, IL-10 treatment caused an almost complete loss of these differentiation-dependent peaks, indicating that the chromatin remained closed in response to IL-10 signaling (Figure 3-6B). These changes in ATAC peaks were consistent with the RNA-seq data showing a decreased *Ucp1* transcript in IL-

10- treated cells and increased transcript in IL-10-deficient mice compared to controls (Figure 3-6B). Importantly, there were a number of prominent ATAC peaks at the *Ucp1* locus that were not affected by IL-10 (e.g. Peak 2), indicating that IL-10 was selectively altering chromatin at specific sites. Specificity was further confirmed by aligning the adipocyte results with ATAC- seq data from *Il10–/–* bone marrow-derived macrophages (Macs) treated with and without IL-10 (Figure 3-6B). Most of the peaks present at the *Ucp1* locus in adipocytes were absent in macrophages.

Directed qPCR analysis of the ATAC seq samples further validated the repressive effects of IL-10 at Peaks 1 and 3 but not 2 (Figure 3-6D). Remarkably, IL-10 treatment also altered chromatin configuration at the regulatory regions of a battery of other thermogenic genes, including *Cidea, Ppargc1α, Elovl3, and Cox8b* (Figure 3-6E and 3-12B). Furthermore, examination of ATAC signals at genes whose expression was not altered by IL-10, including *Fabp4 and Ephx1*, showed that these ATAC peaks were virtually unchanged by IL-10 treatment (Figure 3-6F and 3-12B). The expected increase in ATAC signals at the *Socs3* gene locus served as a positive control for IL-10-mediated transcriptional effects (Figure 3-6F). Finally, we plotted the ratio of vehicle/IL-10 from our ATAC-seq data as a function of gene expression (RNA-Seq) to correlate the changes in chromatin regions with active transcription mediated by IL-10. Changes in ATAC peaks did not necessarily correlate with transcript abundance, suggesting that chromatin configuration does not necessarily translate into transcriptional regulation. However, we found a small cluster of genes in the quadrant 2 of the data plot shown in Figure 3-6G with particularly high ATAC/RNA-seq correlation. This cluster turned out to include thermogenic genes such as *Ucp1* and *Cidea*, further underscoring the specificity of IL-10 action in remodeling chromatin specific for the thermogenic gene program.

IL-10 alters transcription factor occupancy at regulatory regions of thermogenic genes

Thermogenic gene transcription during adipose tissue browning is orchestrated by multiple transcription factors, including CREB/ATF, CEBPs and PPARγ (Harms and Seale, 2013). To investigate if IL-10 signaling in adipocytes selectively affected accessibility at sites of transcription factor binding, we performed *in silico* analysis. We plotted all the ATAC-seq data peaks from D0 and D5 as a function of fold-induction to assess what percentage of peaks showed an increase or decrease in accessibility during differentiation. About 10% of the ATAC peaks show an increase of 5-fold or more on D5 (Figure 3-12C). Furthermore, we found that the distribution of ATAC peaks that were highly enriched during BAT differentiation favored intergenic regions that could possibly contain enhancer elements (Figure 3-7A and 3-7B). We separated the ATAC peaks into 10 equivalently sized bins to assess the peak strength (RPKM) within each category of samples (D0, D5, and D5 + IL-10). Accessibility markedly increased across the bins from D0 to D5, and IL-10 treatment blunted these changes (Figure 3-7C). Next we quantitatively assess transcription factor binding sites in the intergenic/enhancer regions where ATAC peaks were highly enriched. Interestingly, motifs associated with the binding of canonical thermogenic transcription factors such as CREB/ATF, CEBPs, NFIs, and PPARγ were highly enriched on D5 (Figure 3-7D). By contrast AP-1 (Fos/Jun) motifs were highly downregulated during browning. To further investigate the effect of IL-10 on transcription factor enrichment, we analyzed the same regions from panel C-D and divided the motifs into three groups based on the level of IL-10 inhibition. Interestingly, IL-10 caused a substantial loss of enrichment for motifs associated with the same group of thermogenesis-associated transcription factors (Figure 3-7E).

To complement these in silico analyses, we directly tested the functional relevance of the transcription factor motifs identified by the ATAC-seq studies. We performed directed ChIP-qPCR analysis on the regulatory regions of the *Ucp1* gene locus (Figure 3-7F). We confirmed that Peaks 1 and 3 in Figure 3-7F, which changed markedly upon IL-10 treatment, contained sequences known to recruit CREB/ATF and its cofactor PGC-1α (Abe et al., 2015; Cao et al.,

2004; Chen et al., 2009). ATF-2 plays a key role in mediating the thermogenic effects of β-adrenergic signaling. PGC-1α is a master regulator of mitochondrial biogenesis and directly increases thermogenic gene transcription. We found that both ATF-2 and PGC1α showed robust enrichment on chromatin in the region of Peaks 1 and 3, but not Peak 2. IL-10 treatment compromised the recruitment of PGC-1α and ATF-2 to the Ucp1 regulatory regions corresponding to Peaks 1 and 3, but not to the constitutively present Peak 2 region (Figure 3-7F).

Discussion

Although multiple signaling pathways that stimulate adipose tissue "beiging/browning" have been characterized, the physiologic limiters of energy expenditure programs have largely remained unknown. Here we have outlined an unexpected role for IL-10 in altering the adipocyte transcriptional landscape to limit thermogenesis. Loss of IL-10 in mice increased energy expenditure, promoted mitochondrial respiration, improved glucose tolerance and protected against diet-induced obesity, and did so in the absence of overt systemic or adipose tissue inflammation. We further showed that IL-10 acts directly on adipocytes to repress the expression of thermogenic genes by altering the chromatin landscape. These findings challenge conventional assumptions regarding the links between immune and inflammatory signaling and adipocyte metabolism. They further suggest that blockade of IL-10 receptor signaling in fat could represent a tractable approach to de-repress thermogenic gene expression in a therapeutic context.

Immune cells within adipose tissue are increasingly being recognized as important contributors to metabolic homeostasis that can modulate adipocyte function (Osborn and Olefsky, 2012). A number of pro-inflammatory cytokines have been shown to impair lipid storage and to exacerbate insulin resistance in mouse models, leading to the suggestion that inhibition of adipose tissue inflammation might be beneficial in the setting of diabetes (Romeo et

al., 2012; Shoelson et al., 2006). IL-10 is a major anti-inflammatory factor with a well-established role in countering the pro-inflammatory actions of other cytokines (Couper et al., 2008; Hutchins et al., 2013; Saraiva and O'Garra, 2010). Contrary to the expectation that loss of this key anti-inflammatory factor might lead to unrestrained adipose inflammation, IL-10-deficient mice had improved metabolic profiles and were resistant to obesity. It is important to note that this phenotype is distinct from that observed with elevation of pro-inflammatory stimuli such as TNFα and IL-6, which promote weight loss but exacerbate insulin resistance (Petruzzelli et al., 2014; Romeo et al., 2012; Shoelson et al., 2007).

Repressive signals provide critical balance in cellular homeostasis. Such feedback is highly prevalent in immune cell interactions. Anti-inflammatory pathways participate in negative feedback loops during inflammation to block excessive synthesis of pro-inflammatory molecules and promote resolution. It is widely believed that transcriptional repression underlies many of the effects of anti-inflammatory signals such as Type II cytokines, TGF-β, and type I interferons (Murray and Smale, 2012). However, the precise mechanisms whereby these factors silence gene expression have remained poorly understood. Murray and colleagues have postulated that STAT3 is essential for all known IL-10-mediated signaling and that IL-10 elicits its anti-inflammatory action primarily through transcription (Murray, 2005). Several factors induced by the pSTAT3/IL-10 axis, including Bcl-3, Sbno2, Etv3 and NFIL3, have been shown to contribute to IL-10's anti-inflammatory effects (El Kasmi et al., 2007; Hutchins et al., 2013; Kuwata et al., 2003; Murray and Smale, 2012; Smith et al., 2011).

Here we have built upon prior work in macrophages to dissect the actions of IL-10 in a different cell type, where it acts on a largely distinct set of transcriptional target genes. We showed that the IL-10 signaling axis is intact in adipocytes. We found that IL-10 selectively represses the expression of a battery of thermogenic genes in adipocytes, including the critical uncoupling mediator *Ucp1.* To understand the mechanism for this effect we employed a combinatorial approach using ATAC- and RNA-Seq on brown/beige differentiated preadipocytes

treated with and without IL-10. These studies revealed that chromatin at specific sites the regulatory regions of thermogenic genes remained closed during browning in the presence of IL-10. Importantly, this effect was highly selective for the thermogenic program, as the chromatin structure at adipocyte genes not related to beiging/browning was not altered by IL-10. Thus, IL-10 is not a general inhibitor of differentiation-dependent adipocyte transcription, but rather a specific modifier of thermogenesis.

Our analysis further allowed us to identify specific transcription factors whose interactions with regulatory regions of thermogenic genes was dependent on IL-10 signaling. ATAC accessibility at ATFs/CREB and C/EBP motifs was highly enriched during beiging/browning, and IL-10 caused an almost complete reversal of accessibility at these motifs. Consistent with the IL-10-dependent changes in accessibility at ATF motifs, directed ChIP analysis showed reduced occupancy of ATF2 and its cofactor PGC-1α at *Ucp1* regulatory regions in the presence of IL-10. Interestingly, expression of PGC-1α expression itself was also repressed in response to IL-10, potentially contributing to the effects on thermogenic gene expression. Finally, our gene expression analysis identified inhibitory transcriptional regulators that were induced in adipocytes in response to IL-10, including such as HDAC7/5 and NCOR2. It is possible that such factors may also be involved in mediating repressive actions on chromatin accessibility in response to IL-10.

The striking metabolic phenotype of IL-10-deficient mice indicates that basal IL-10 signaling provides a tonic brake that limits thermogenic gene expression. WAT undergoes continuous remodeling under chronic cold exposure. In addition to canonical thermogenic genes (*Ucp1, Elovl3, Cidea, Ppargc1*α), a specific set of thermogenic genes such as *Pm20d1, Otop1, Elovl6, and Slc27a2* are turned on that typically serve as markers for cold exposure. Our RNA-Seq data show that this same set of genes was highly upregulated (~5-fold) in the iWAT of IL-10-deficient mice, even thought they were not exposed to cold (Long et al., 2016; Rosell et al., 2014). Mice housed at ambient temperature are under thermal stress and have more browning

of WAT compared to mice housed at thermoneutrality (Qiu et al., 2014). One consequence of thermal stress is increased local adipose catecholamine signaling, which directly acts on β-adrenergic receptors to activate thermogenic gene expression (Qiu et al., 2014; Rao et al., 2014). Adipose adrenergic signaling initiates a cascade leading to the activation of transcription factors including ATFs, CREBs, PGC1α, and PRDM16 that bind to regulatory regions of thermogenic genes (Harms and Seale, 2013). Our data suggest that IL-10 signaling antagonizes adrenergic signaling and inhibits thermogenesis in adipose tissue.

What is the biological relevance of adipose IL-10 signaling? One important clue to this question is the dynamic expression pattern of IL-10Rα. The expression of IL-10Rα is highly upregulated during adipocyte differentiation, consistent with an important physiological function for IL-10 signaling in adipose tissue. In addition, IL-10Rα expression is further elevated in response to cold exposure, obesity and aging, implying that changes in the activity of the IL-10 axis are also relevant in these contexts. Since the gene encoding IL10Rα is a direct target of PPARγ, it is reasonable to hypothesize that the adipocyte IL-10 axis could serve to facilitate lipid storage and maintain adiposity. Given the central role that IL-10 plays in inflammation and immunity, we further speculate that the IL-10 signaling might function as a mechanism to conserve energy in the setting of increased systemic energy demands such as during infection. Finally, our data suggest that blockade of IL-10 signaling in adipose tissue might have beneficial effects in the context of obesity and insulin resistance. Our finding that adenoviral-mediated knockdown of IL10Rα expression in iWAT was sufficient to induce thermogenic genes supports further research in to the potential therapeutic utility of targeting the adipose IL-10 axis.

Adipose tissue inflammation is widely regarded to be a contributory factor in the development of obesity and metabolic dysfunction. However, paradoxical increases in obesity and insulin resistance in mice depleted of various pro-inflammatory signals, and the development of age-related obesity upon anti-inflammatory ablation suggest a more complex relationship between the immune system, adipose tissue function and systemic metabolism

(Bapat et al., 2015; Mauer et al., 2014; Wallenius et al., 2002; Wernstedt Asterholm et al., 2014). Here we have reported an unexpected role for the anti-inflammatory cytokine IL-10 in regulating adipose thermogenesis, obesity, and insulin sensitivity. Together with prior work, our data support a model in which immune mediators exert complex effects on adipose tissue function, depending upon the physiological signal and microenvironment. They further underscore the view that proper balance of inflammatory and anti-inflammatory molecules is key in maintaining adipose and metabolic homeostasis.

Experimental Procedures

Reagents and Plasmids Dexamethasone (D2915), 3-isobutyl-1-methylxanthine (IBMX, I7018), PPARγ agonist Rosiglitazone (R2408), indomethacin (I7378), T3 (3,3',5-Triiodo-L-thyronine; T2877), and isoproterenol (I6504) were from Sigma-Aldrich. Insulin (12585-014) was from Gibco and recombinant mouse IL-10 (210-10) was purchased from Peprotech. IL10Rα shRNA sequences were designed using BLOCK-iT RNAi designer tool (Invitrogen). Sense and antisense oligos were annealed and cloned into pENTR/U6 plasmid (Invitrogen). Using LR recombination (Invitrogen), shRNA constructs were subcloned into a Gateway-adapted pBLOCK-IT adenovirus vector. The following shRNA oligos were used: lacZ shRNA (control) CACCGGGCCAGCTGTATAGACATCTCGAAAGATGTCTATACAGCTGGCCC; IL10Rα shRNA CACCGCATCTTAGTCATATCTATGCCGAAGCATAGATATGACTAAGATGC. Only sense strands are shown here. Adenovirus was amplified, purified and tittered by Viraquest Inc.

Cell Culture

Murine white and brown preadipocytes were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented 10% fetal bovine serum (FBS). For *in vitro* brown adipocyte differentiation, preadipocytes were grown to confluence in DMEM with 10% FBS plus insulin (5 μg/ml) and T3 (1 nM). Confluent cells were induced to differentiate with dexamethasone (1 μM),

IBMX (0.5 mM), insulin (5 µg/ml), indomethacin (125 nM) and Rosiglitazone (1 µM) for 2 days, followed by insulin, T3 and Rosiglitazone alone. On the fourth day, cells were pretreated for overnight with and without 100 ng/ml IL-10 and next day treated with 10 µM isoproterenol for 5-6 h. White and brown preadipocytes were isolated and immortalized as previously described (Villanueva et al., 2013). Beige/Brown preadipocytes IL10Rα expressing stable cells (iBAd) were generated using the pBabe retroviral system (Hummasti and Tontonoz, 2006). Retrovirus was obtained by transfection of Phoenix E cells with the pBabe vector and harvest of growth media 72 h later. Preadipocytes were transduced with retrovirus overnight and selected with particular antibiotics.

Gene Expression Analysis

Total RNA was isolated using TRIzol reagent (Invitrogen) and reverse transcribed with the iScript cDNA synthesis kit (Biorad). cDNA was quantified by real-time PCR using SYBR Green Master Mix (Diagenode) on an ABI 7900 instrument. Gene expression levels were determined by using a standard curve. Each gene was normalized to the housekeeping gene 36B4 and was analyzed in duplicate. Primers used for real-time PCR are available upon request.

Protein Analysis

Whole cell lysate was extracted using RIPA lysis buffer (Boston Bioproducts) supplemented with complete protease inhibitor cocktail (Roche). Proteins were diluted in Nupage loading dye (Invitrogen), heated at 95°C for 5 min, and run on 4–12% NuPAGE Bis-Tris Gel (Invitrogen). Proteins were transferred to hybond ECL membrane (GE Healthcare) and blotted with IL10Rα (AF-474-SP, R&D Systems), PGC1α (H-300 sc-13067, Santa Cruz Biotechnology), pATF-2 (F1 sc-8398, Santa Cruz Biotechnology), ATF-2 (C-19 sc-187, Santa Cruz Biotechnology), pSTAT3 (9131, Cell Signaling), TLE3 (11372-1-AP, Proteintech Group), and αTubulin (CP06, Calbiochem) antibodies.

Animal Studies

Breeding pairs of *Il10–/–* mice (#002251) and WT (#000664) controls were acquired from Jackson Laboratory and colony maintained in pathogen-free barrier-protected environment (12:12 h light/dark cycle, 22°C-24°C) at UCLA animal facility. Experimental mice were sacrificed at ages mentioned in figure legends for histological and gene expression analysis. For the IL10Rα shRNA adenovirus delivery to fat pads, 2X109 PFU of adenovirus was percutaneously injected into each inguinal fat depot of WT mice at 8-10 weeks of age. In each mouse, Ad-IL10Rα shRNA was injected into iWAT on one side, and Ad-Lacz shRNA (control) was injected into the contralateral side as a control. 4-5 days after the injection, iWAT were resected for gene expression analysis. For ex vivo iWAT IL-10 treatment, 10 weeks old WT mice were housed at cold room (4-6°C) for 6 h and iWATs were isolated and minced, placed in KREB's Ringer Buffer (12 mM HEPES, 121 mM NaCl, 4.9 mM KCl, 1.2 mM MgSO4, 0.33 mM CaCl2) supplemented with 0.1% glucose and incubated with and without IL10 for 30 mins-1h at 37°C. For bone marrow transplantation studies, recipient WT or *Il10–/–* mice (10 weeks of age) were lethally irradiated with 900 rads and transplanted with 3 × 106 bone marrow cells from 8- week-old or older donors (*Il10–/–*) via tail vein injection. At 10 weeks of age, *Il10–/–* and WT mice were fed a 60% high-fat diet (Research Diets) for the indicated times. For glucose tolerance tests, mice were fasted for 6 hr and challenged with an i.p. injection of glucose (2 g/kg). For insulin tolerance tests, mice were fasted for 6 hr and given an i.p. injection of insulin (1 U/kg). Blood glucose levels were monitored using the ACCUCHEK active glucometer (Roche). Body composition was determined by EchoMRI analysis. Indirect calorimetry was performed using a Columbus Instruments Comprehensive Lab Animal Monitoring System (CLAMS, Columbus Instruments). Animals were placed individually in chambers for 3 consecutive days at ambient temperature (26.5°C) with 12 hr light/dark cycles. Animals had free access to food and water. Respiratory measurements were made in 20 min intervals after initial 7-9 hr acclimation period.

Energy expenditure was calculated from VO2 and RER using the Lusk equation, EE in Kcal/hr = (3.815 + 1.232 X RER) X VO2 in ml/min. Statistical significance for EE, VO2, and VCO2 measurements was determined by using two-way ANOVA. Serum triglyceride was measured using Wako L-Type TG M Test kit. Serum cytokines were determined by multiplex immunoassay (Milliplex Kit, Millipore). Animal experiments were conducted in accordance with the UCLA Institutional Animal Care and Research Advisory Committee.

Cellular and mitochondrial respiration

Cell were seeded in a XF24 plate, differentiated, and analyzed in a XF24 analyzer (Seahorse Bioscience/Agilent) as described (Wu et al., 2007). Briefly, oxygen consumption rate (OCR) was measured before and after the sequential injection of 0.75 μM oligomycin, 1 μM FCCP, and 1 μM of rotenone/myxothiazol. Mixing, waiting, and measurement times were 5, 2, and 2 min, respectively. Measures were normalized by total protein. In another set of experiments, mitochondria were isolated from fresh tissues and immediately used in a XF24 analyzer as previously described (Rogers et al., 2011). Briefly, mitochondria were isolated in MSHE+BSA buffer using a 800g/8000g dual centrifugation method and resuspended in MAS buffer. Protein concentration was determined using a Bradford Assay reagent (Bio-Rad) and 20 μg of protein were seeded per well by centrifugation. Coupling and electron flow assays were performed as described (Rogers et al., 2011). For the coupling assay, basal oxygen consumption rate (OCR) was measured in the presence of 10 mM succinate and 2 μM rotenone, and after sequential addition of 4 mM ADP (Complex V substrate), 2.5 μg/ml oligomycin (Complex V inhibitor), 4 μM FCCP (mitochondrial uncoupler) and 4 μM antimycin A (Complex III inhibitor). Coupled respiration was calculated as the difference between basal and response to oligomycin. Uncoupled respiration was the difference between oligomycin and antimycin A injections. For electron flow assays, basal OCR was measured in presence of 10 mM pyruvate (Complex I substrate), 2 mM malate and 4 μM FCCP, and after sequential addition of 2 μM rotenone

(Complex I inhibitor), 10 mM succinate (Complex II substrate), 4 μM antimycin A (Complex III inhibitor) and 1mM TMPD containing 10 mM ascorbate (Complex IV substrate). Complex III respiration corresponds to the antimycin A-sensitive respiration.

Chromatin Immunoprecipitation (ChIP)

ChIP experiments were performed according to standard protocols (Villanueva et al., 2013; Villanueva et al., 2011). Lysed cells were sonicated using a Bioruptor (Diagenode) according to the manufacturer's protocol, and chromatin was immunoprecipitated with antibodies against PGC1α (H-300 sc-13067, Santa Cruz Biotechnology), ATF-2 (C-19 sc-187, Santa Cruz Biotechnology), and IgG (PP64, Millipore) overnight at 4°C in the presence of Protein A beads (GE Healthcare). DNA enrichment was quantified by real-time PCR (ABI 7900, ABI, Carlsbad, CA) using SYBR Green Master Mix (Diagenode or Sigma-Aldrich). Primers used for these studies available upon request. Occupancy was quantified using a standard curve and normalized to input DNA.

RNA-seq

Total RNA was prepared as described (Tong et al., 2016). Strand-specific libraries were generated from 500 ng total RNA using the TruSeq Stranded Total RNA Library Prep Kit (Illumina). cDNA libraries were single-end sequenced (50bp) on an Illumina HiSeq 2000 or 4000. Reads were aligned to the mouse genome (NCBI37/mm9) with TopHat v1.3.3 and allowed one alignment with up to two mismatches per read. mRNA RPKM values were calculated using Seqmonk's mRNA quantitation pipeline.

All RPKMs represent an average from three biological replicates for in-vitro studies, and pooled RNA representation for tissue samples where equal amounts of RNA were pooled from 11 *Il10*-/- animals and 9 WT animals prior to library construction. A gene was included in the analysis if it met all of the following criteria: The maximum RPKM reached 4 at any time point, the gene

length was >200bp, and for in-vitro studies was induced at least 3-fold from Day 0 samples, and the expression was significantly different from the basal (P<0.01) as determined by the DESeq2 package in R Bioconductor. P-values were adjusted using the Benjamini-Hochberg procedure of multiple hypothesis testing (Benjamini and Hochberg, 1995)

ATAC-seq

ATAC-seq libraries were prepared using the Nextera Tn5 Transposase kit (Illumina) as described (Buenrostro et al., 2015) with slight modifications. Libraries were single-end sequenced (50bp) on an Illumina HiSeq 2000. Reads were mapped to the mouse genome (NCBI37/mm9) using Bowtie2. Reads were removed from the subsequent analysis if they were duplicated, mapped to mitochondrial genome, or aligned to unmapped contiguous sequences. Peak calling was performed using MACS2 using parameters callpeak --nomodel -g mm --keep-dup all -q .01 -- llocal 10000. The reads were converted to reads per thousand base pairs peak per million mapped reads (RPKM) by dividing by the total number of reads per sample.

Motif Analysis

MACS2 called ATAC peak regions were used for motif analysis. JASPAR2016 Position Weight Matrices were used to identify binding sites in ATAC peaks using Pscan-ChIP (Zambelli et al., 2009).

Statistics

All data are presented as mean ± SEM and analyzed using Prism (Graphpad). Unpaired two-tailed Student's t test was used for single variable comparison between two groups. One-way ANOVA followed by Dunnett post hoc test was used for multiple comparisons versus the control group. Two-way ANOVA followed by Bonferroni posttests was used to examine interactions

between multiple variables. $p < 0.05$ was considered to be statistically significant and is presented as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, or **** $p < 0.0001$.

Author Contribution

P.R. designed and performed most of the experiments. P.R. and B.J.T. performed animal studies and genome-wide sequencing studies. L.V., and K.R. performed mitochondria respiration analysis. P.R., B.J.T., and C.H. performed bone-marrow transplantation experiment. B.W. performed multiplex immunoassay. P.R. and T.S. performed metabolic chamber studies. S.G.Y., K.R., S.T.S, and P.T. designed experiments. P.R. and P.T. wrote the manuscript.

Acknowledgements

We thank UCLA Broad Stem Cell Research Center Core for sequencing. This work was supported by NIH grants F32DK104484 (P.R.), T32AI007323 and T32GM008042 (B.J.T), HL128822 (T.S.), F32DK109601 (B.W.), HL090533 (K.R., S.G.Y., P.T.), R01GM086372 (S.T.S.), and DK063491 (P.T.).

Conflicts Of Interest

The authors declare no conflicts of interest.

Figure Legends

Figure 3-1. IL-10-deficient mice are protected against obesity. (A) Representative images of 10-week-old chow-fed WT and *Il10–/–* mice showing gross adipose tissue appearance (iWAT and eWAT) and histology (H&E staining). (B) IP glucose tolerance test performed on 10-week-old chow-fed WT and *Il10–/–* mice, n = 7 per group. Comparisons at each time point were made against WT control mice by repeated measure ANOVA. (C) Body weight of 32-week-old chow-fed WT and *Il10–/–* mice, *n*=8,4 per group. (D) Gross appearance of representative 32-week-old chow-fed WT and *Il10–/–* mice and their tissues. (E) Body fat % of 32-week-old chow-fed WT and *Il10–/–* mice determined by Echo MRI, *n*=8,4 per group. (F) Body weight of WT and *Il10–/–* mice fed chow diet for 10 weeks and then 60% high-fat diet (HFD) for 6 weeks. *n* = 16,12 per group. Statistical analysis was performed using Student's t- test. (G) External and gross (adipose tissues and liver) appearance of representative 6 weeks HFD-fed 16 weeks old WT and IL10 KO mice. (H) Body fat and lean mass of mice in F. Statistical analysis was performed using Student's t-test. (I) Representative histology of iWAT and BAT from mice in F. (J) IP glucose tolerance test (GTT) performed on WT and *Il10–/–* mice fed chow diet for 10 weeks and then 60% high-fat diet (HFD) for 6 weeks, n = 7 per group. Comparisons at each time point were made against WT control mice by repeated measure ANOVA. (K) IP insulin tolerance test (ITT) performed on mice in J. n = 7 per group. Comparisons at each time point were made against WT control mice by repeated measure ANOVA. *, P< 0.05; **, P<0.01; n.s, not significant.

Figure 3-2. IL-10 deficiency promotes energy expenditure and enhances mitochondrial respiration.

(A-D) Energy expenditure (EE) rate (kCal/kg/hr), VO2 (ml/kg/hr), VCO2 (ml/kg/hr), and food intake (g) of chow-fed 10-week-old WT and *Il10–/–* mice were analyzed by Columbus Oxymax

metabolic chambers. 12 hr light/dark cycles, 72 hr total duration, each light/dark bar represents 12 h duration. n = 7 per group. Statistical analysis was performed using two-way ANOVA. (E and F) Average oxygen consumption rate (OCR) of mitochondria isolated from iWAT of WT and *Il10–/–* mice in mitochondrial coupling (E) and electron flow assays (F). Samples were treated with different substrates or inhibitors to obtain specific respiration states as indicated (see Methods). Data are averages of six internal replicates and are representative of two independent experiments. Statistical analysis was performed using Student's t-test. (G and H) EE rate (kCal/kg/hr), VO2 (ml/kg/hr), and food intake (g) of 6-week-HFD-fed WT and *Il10–/–* mice, n = 7 per group. Statistical analysis was performed using two-way ANOVA.

Figure 3-3. IL-10-deficient mice have increased browning of white adipose tissue.

(A) Scatter plot of gene expression differences between WT and *Il10–/–* mice as determined by RNA-sequencing of iWAT. Genes with at least 4 RPKM are shown. The log2 ratio of KO/WT expression (x-axis) is shown as a function of max RPKM (y-axis), with select genes indicated with vertical text. Shades of blue correspond to genes downregulated in the *Il10–/–* mice, and shades of red indicate upregulation in *Il10–/–* mice. *n* = 9,11 per group. (B) Gene expression scatter plot of those genes demonstrating >1.5 fold difference in expression between WT and *Il10–/–* mice (y-axis) as a function of Max expression (x-axis). Genes are color coded based on their expression profile in published WAT vs BAT tissue (Sun et al., 2013) with shades of red indicating increasing degrees of BAT-selectivity and shades of blue indicating increasing amounts of WAT-selectivity. Black open circles represent genes demonstrating either no-selectivity, or are lowly expressed in the WAT/BAT dataset (below 4 RPKM). (C) Genes in Panel B are quantitatively shown in a table, both as numbers of genes within each adipose- subtype (WAT or BAT) category, as well as the percent of all strain-selective genes (WT v.s. *Il10–/–*). (D) Genes at least 1.5-fold enriched in either WT (top) or *Il10–/–* mice (bottom) are classified based on their WAT/BAT-selectivity, with darkening shades of red indicating increasing BAT-selectivity

and increasing shades of blue indicating increased WAT-selectivity. There was an ~4.8-fold enrichment for genes showing WAT-selectivity in the iWAT of WT mice, and an ~4.5-fold enrichment of genes showing BAT-selectivity in the iWAT of *Il10–/–* mice mice. (E) Selected genes (y-axis) demonstrating BAT/WAT selectivity as a function of their *Il10–/–*/WT expression ratio (x-axis). (F) Log2 ratio of expression for all the genes with RPKM expression above 4 within the Gene Ontology category for Brown Fat Differentiation (GO:0050873). (G) Directed real-time analysis of gene expression in iWAT from WT and *Il10– /–* mice. *n* = 9,11 per group.

Figure 3-4. Reconstitution of bone marrow IL-10 expression reverses the thermogenic phenotype of IL-10-deficient mice. (A) Body weight of lethally-irradiated chow-fed *Il10–/–* mice reconstituted with WT (WT→KO) or *Il10–/–* (KO→KO) bone marrow 7 weeks post-transplant. *n* = 10,10 per group. Statistical analysis was performed using Student's t-test. (B) Body fat and lean mass of WT→KO and KO→KO mice determined by Echo MRI. *n*=10,10 per group. Statistical analysis was performed using Student's t-test. (C) Gross appearance of iWAT 19 weeks post-BMT. (D) IP GTT performed on WT→KO and KO→KO mice; n = 10 per group. Comparisons at each time point were made by repeated measure ANOVA. (E) EE rate (kCal/kg/hr), VO2 (ml/kg/hr), and food intake (g) analyzed by Columbus Oxymax metabolic chambers. 12 hr light/dark cycles, 72 hr total duration, each light/dark bar represents 12 h duration. n = 10 per group. Statistical analysis was performed using two-way ANOVA. (F) Normalized gene expression in iWAT determined by real-time PCR. *n* = 9 per group. Statistical analysis was performed using Student's t-test. (G-H) Average oxygen consumption rate (OCR) in coupling (E) and electron flow (F) assays of mitochondria isolated from iWAT of WT→KO and KO→KO mice. Data are averages of six internal replicates and are representative of two independent experiments. Statistical analysis was performed using Student's t-test.

Figure 3-5. IL-10 acts directly on adipocyte IL-10Rα to inhibit thermogenesis. (A) Real-time PCR analysis of mRNA encoding IL10Rα during the differentiation of primary stromal vascular fraction (SVF) derived from WT chow-fed 10 week-old mice. Cells were stimulated to differentiate with dexamethasone (1 μM), IBMX (0.5 mM), insulin (5 μg/ml), and GW1929 (20 nM) for 2 days followed by insulin and GW1929 for 5 days. (B) Immunoblot analysis of IL10Rα in SVF and mature adipocyte (adipo) fraction of iWAT from chow-fed 10 week-old mice. (C) Real-time PCR analysis of *Il10ra* mRNA from iWAT of mice at room temperature (RT) and 4° C, 12 weeks chow or HFD-fed mice, and 12 week-old WT and *db/db* animals. Statistical analysis was performed using Student's t-test. *n* = 10-15 per condition. (D) Immunoblot analysis of IL10Rα and UCP1 protein from iWAT of 10 week-old mice injected with adenovirus expressing control shRNA (shCtrl) or shRNA targeting IL10Rα for 72 h. Each lane represents an individual animal. (E) Real-time PCR analysis of gene expression in iWAT transduced with shCtrl or shIL10Rα adenovirus. Data represent averages of 12 mice/group. Statistical analysis was performed using Student's t-test. (F) Immunoblot analysis of protein extracts from iWAT of 10 week-old mice treated ex vivo with vehicle (veh) or 100 ng/ml IL-10. *n* = 4-6 per treatment/cohort. Results are representative of three independent experiments. (G) Real-time PCR of indicated genes from iWAT of 10 weeks old mice ex vivo treated with control (NT) or 100 ng/ml IL-10. *n* = 4-6 per treatment/cohort. (H) Heatmap representation of genes that changed >3-fold (p-value <0.01) by RNA-seq on day 5 (D5) of differentiation of immortalized BAT preadipocytes stably expressing IL10Rα. Each sample is shown in triplicate and compared to expression at Day 0 (D0). Genes are grouped as either induced upon differentiation (top), or repressed during differentiation (bottom). The far right column shows the effect of IL-10 treatment on gene expression at Day 5 of differentiation. Genes are ranked based on IL-10 inhibition, with select genes shown in the text at right. (I) Genes that were induced upon differentiation were divided based on their response to IL-10 treatment, either inhibited >2-fold (red bars) or not effected (black bars) and Gene Ontology analysis was performed with –log10

(p-value) plotted (x-axis) as a function of classifications meeting a p-value of < 0.001. See also Figure 3-12. (J) RNA-Seq data from WAT/BAT ratio plotted as a function of RNA-Seq data from (H) NT/IL10 ratio. The fourth quadrant shows that 70% of the genes repressed by IL-10 are BAT-expressed genes.

Figure 3-6. IL-10 signaling remodels chromatin architecture at thermogenic genes.

(A) Heatmap analysis of ATAC-Seq performed on D0 or D5 of brown preadipocyte differentiation with and without IL-10 treatment for all called peaks demonstrating >5-fold induction (n=3177 sites). Peaks were assigned to the nearest gene, and selected genes are shown in the text. (B) Top: ATAC-Seq bedgraph panels of the *Ucp1* locus showing peak locations relative to the TSS as underlined. Panels compare ATAC signals between immortalized brown preadipocytes (preAd) treated with or without IL-10 to signals from *Il10*–/– bone marrow-derived macrophages (macs) treated with and without IL10 and illustrate the specificity of the adipocyte peaks. Below: RNA-Seq bedgraph of *Ucp1* gene demonstrating expression profile during IL-10 stimulation of iBAd cells and in iWAT tissue from WT and *Il10*–/– mice. (C) ATAC peak strength (y-axis) for selected peaks within the *Ucp1* locus under the indicated conditions. (D) Real-time PCR analysis of *Ucp1* gene expression from ATAC-Seq samples. (E) ATAC-Seq bedgraph panels from thermogenic gene loci (*Cidea, Cox8b, Ppargc1*α) showing peak locations relative to the TSS (black arrow). Green arrows indicate peaks differentially affected by vehicle or IL10 treatment. (F) ATAC-Seq bedgraph panel of non-thermogenic gene loci (*Socs3 and Fabp4*. (G) Correlation plot of ATAC-Seq (y-axis) and RNA-Seq (x-axis) data. Selected genes are shown in red in the 2nd quadrant, providing evidence of correlation between IL-10 inhibitory effects on mRNA expression and chromatin accessibility.

Figure 3-7. IL-10 limits enrichment of thermogenic transcriptional regulators at key enhancer elements. (A-B) Localization of all called ATAC peaks, grouped as either maturation-

induced (D5 vs D0) (B) or all others (A). The data reveal an enrichment of intronic and intergenic localized peaks. (C) ATAC peaks were ranked and binned into 10 groups based on fold accessibility. Average RPKM (y-axis) from D0, D5, D5 + IL-10 samples are plotted as a function of each bin where each sample is indicated by a filled solid color. (D) The 10 bins were ranked by order of increasing fold accessibility and used to perform transcription factor (TF) motif analysis using the ChIP-Pscan JASPAR 2016 database (Zambelli et al., 2009). The –log(p-value) is plotted for each TF identified as indicted by the legend with red values indicating high significance of detection. (E) TF Motif analysis of 10 bin containing 3174 ATAC peaks demonstrating the highest fold induction during maturation was separated into three groups of 1058 peaks based on the degree of IL-10 inhibition. (F) ChIP-qPCR was performed for PGC1α and ATF-2 on iBAd cells treated with and without IL-10. qPCR primers were designed spanning the underlined *Ucp1* ATAC peaks shown in the bedgraph in Figure 3-6B. Statistical analysis was performed using Student's t-test.

Supplemental Figure Legends

Figure 3-8. Protection against age-related obesity and absence of systemic inflammation in *IL-10*-/- mice. (A) Body weight of 10 week old chow-fed WT and *IL10*-/- mice, *n*=20 per group. (B) Representative colon histology from 10 week old chow-fed WT and *IL10*-/- mice. (C) Gross appearance of colon tissue from 10 weeks old chow-fed WT and *IL10*-/- mice. (D) Serum cytokines levels of 10-week-old chow-fed WT and *IL10*-/- mice measured by multiplex immunoassay. (E) Serum triglyceride levels of 10 weeks old chow-fed WT and *IL10*-/- mice. (F) FACS analysis of macrophage population from stromal vascular fraction isolated from iWAT of 10 week old chow-fed WT and *IL10*-/- mice. (G) Weight of iWAT and eWAT from 32-week-old chow-fed WT and *Il10–/–* mice, *n*=4,4 per group.

Figure 3-9. Protection against diet-induced obesity and absence of systemic inflammation in high-fat fed *IL-10*-/- mice. (A) Serum cytokines levels of WT and *Il10–/–* mice fed chow diet for 10 weeks and then 60% high-fat diet (HFD) for 6 weeks measured by multiplex immunoassay. *n* = 16,12 per group (B) Gross appearance of colon tissue from HFD-fed WT and *Il10–/–* mice. (C) Representative histology of liver and colon from HFD-fed WT and *Il10–/–* mice. (D) Weight of iWAT and eWAT from HFD-fed WT and *Il10–/–* mice. *n*=7,7 per group.

Figure 3-10. Decreased blood glucose and absence of ulcerative colitis in bone-marrow transplanted mice

(A) Cartoon representation of bone marrow transplantation experiment and genotyping of WT→KO and KO→KO mice for WT or *IL10–/–* allele. (B) Fasting blood glucose levels of WT→KO and KO→KO mice. *n*=10,10 per group. (C) Gross appearance of colon tissue from WT→KO and KO→KO.

Figure 3-11. IL10Rα expression in various adipogenic models and direct inhibition of thermogenic genes and cellular respiration by recombinant IL-10. (A) Real-time PCR analysis of mRNA encoding IL10Rβ during the differentiation of primary stromal vascular fraction (SVF) derived from WT chow-fed 10 week-old mice and real-time PCR analysis of mRNA encoding IL10Rα in iWAT of chow-fed WT 4weeks and 12 week mice and WT and ob/ob mice and 4 weeks ob/ob and 12 weeks ob/ob mice. (B) Immunoblot showing activation of STAT3 (pSTAT3) in response to IL10 and induction of IL10Rα protein in response to differentiation of brown preadipocytes. (C) Bedgraph showing PPARγ and DNAse hypersensitivity (DHS) ChIP-Seq peaks on the enhancer region of *IL10R*α gene locus (Siersbaek et al., 2012). (D) Real-time PCR analysis of mRNA encoding Ucp1, Cidea, Cox8b, and Pgc1α on D0 and D8 of brown differentiation of primary stromal vascular fraction (SVF) derived from WT chow-fed 10 week-old mice. (E) Real-time PCR analysis of mRNA encoding

Socs3 and Ucp1 on D5 of brown differentiated preBAT IL10Rα treated with and without indicated concentration of recombinant IL10. (F) Real-time PCR analysis of mRNA encoding Ucp1 and Cidea on D5 of brown differentiated preBAT IL10Rα treated with and without IL-10 and IL10Rα neutralizing antibody. (G). Oxygen consumption rate (OCR) in D5 differentiated preBAT IL-10Rα cells treated with and without IL-10. (H) Scatter plot of gene expression differences between NT and IL-10 treatment as determined by RNA-sequencing of iWAT.

Figure 3-12. Inhibition of thermogenic gene primary transcript expression and chromatin accessibilities by IL-10. (A) Real-time PCR analysis of primary mRNA of UCP1, Cidea, and Elovl3 and UCP1 eRNA on D5 of brown differentiated preBAT IL10Rα treated with and without indicated concentration of recombinant IL-10. (B) ATAC-Seq bedgraph panels from indicated genes and treatment conditions. (C) Graph showing 100% ATAC peaks plotted as a function of fold induction in accessibility during brown differentiation. 10% peaks show a 5-fold or more increase. (D) Quantification of the ATAC peak pie chart from Figure 3-7A and 3-7B.

Figure 3-1: IL-10-deficient mice are protected against obesity

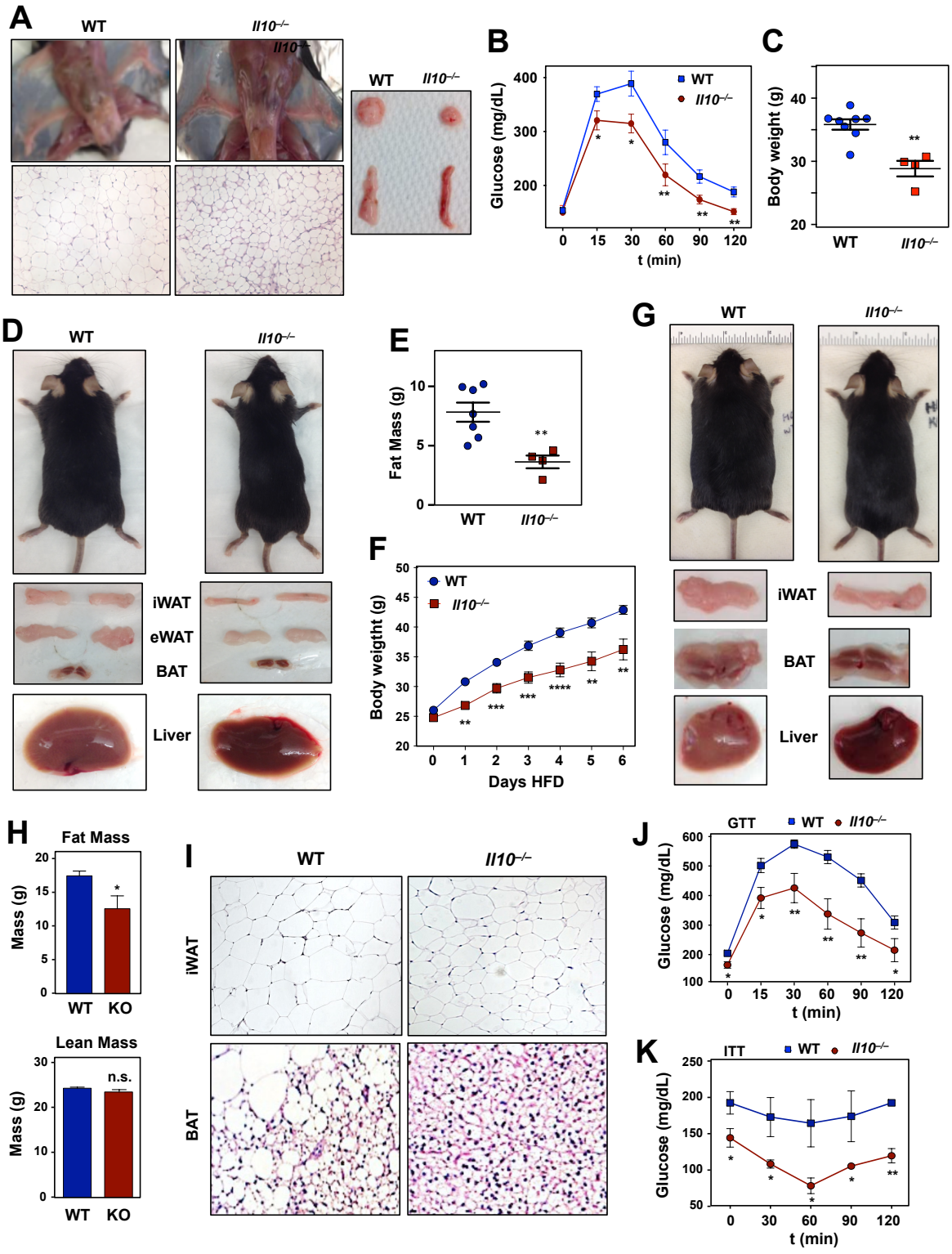


Figure 3-2: IL-10 deficiency promotes energy expenditure and enhances mitochondrial respiration.

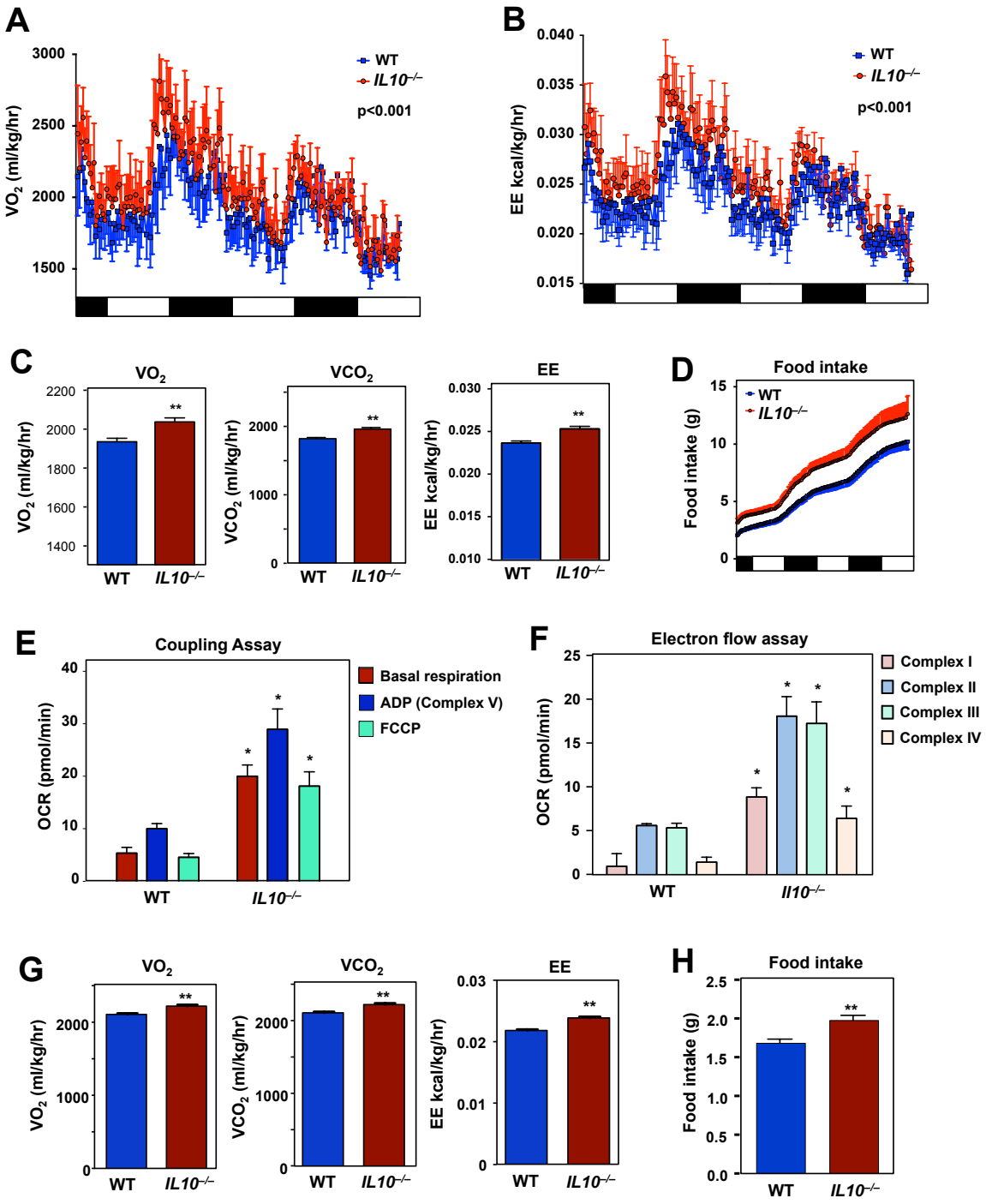


Figure 3-3: IL-10-deficient mice have increased browning of white adipose tissue

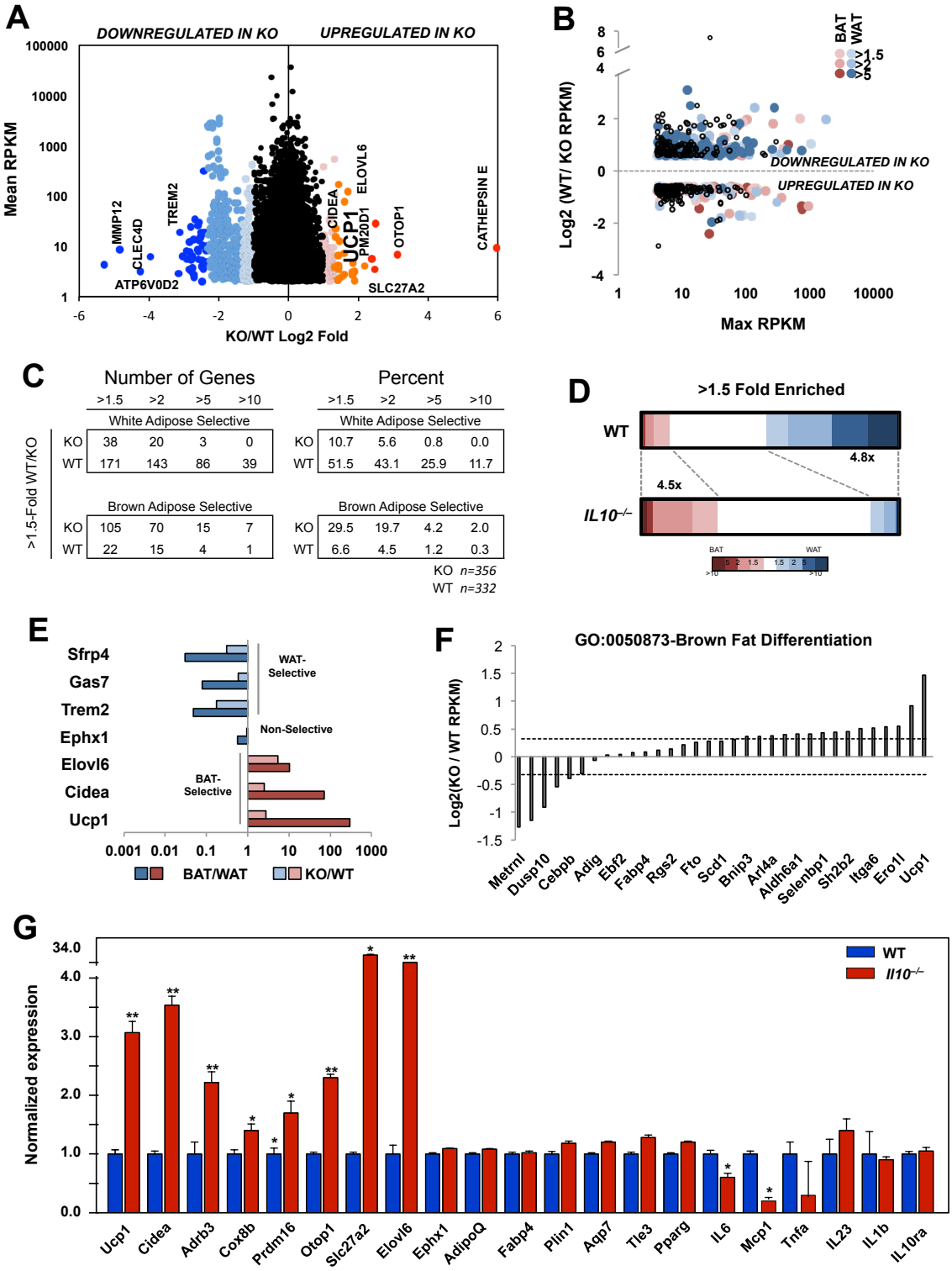


Figure 3-4: Reconstitution of bone marrow IL-10 expression reverses the thermogenic phenotype of IL-10-deficient mice

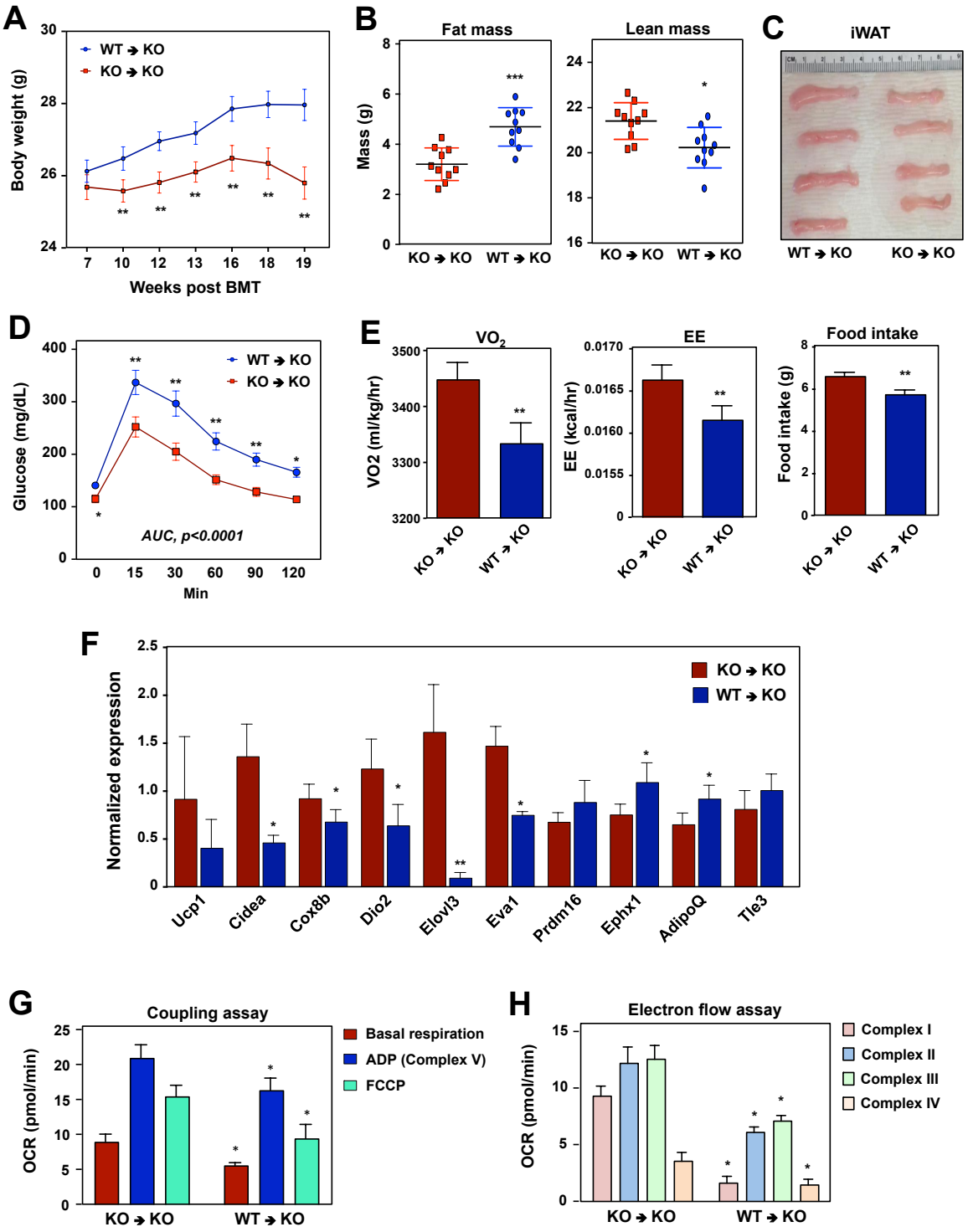


Figure 3-5: IL-10 acts directly on adipocyte IL-10Rα to inhibit thermogenesis

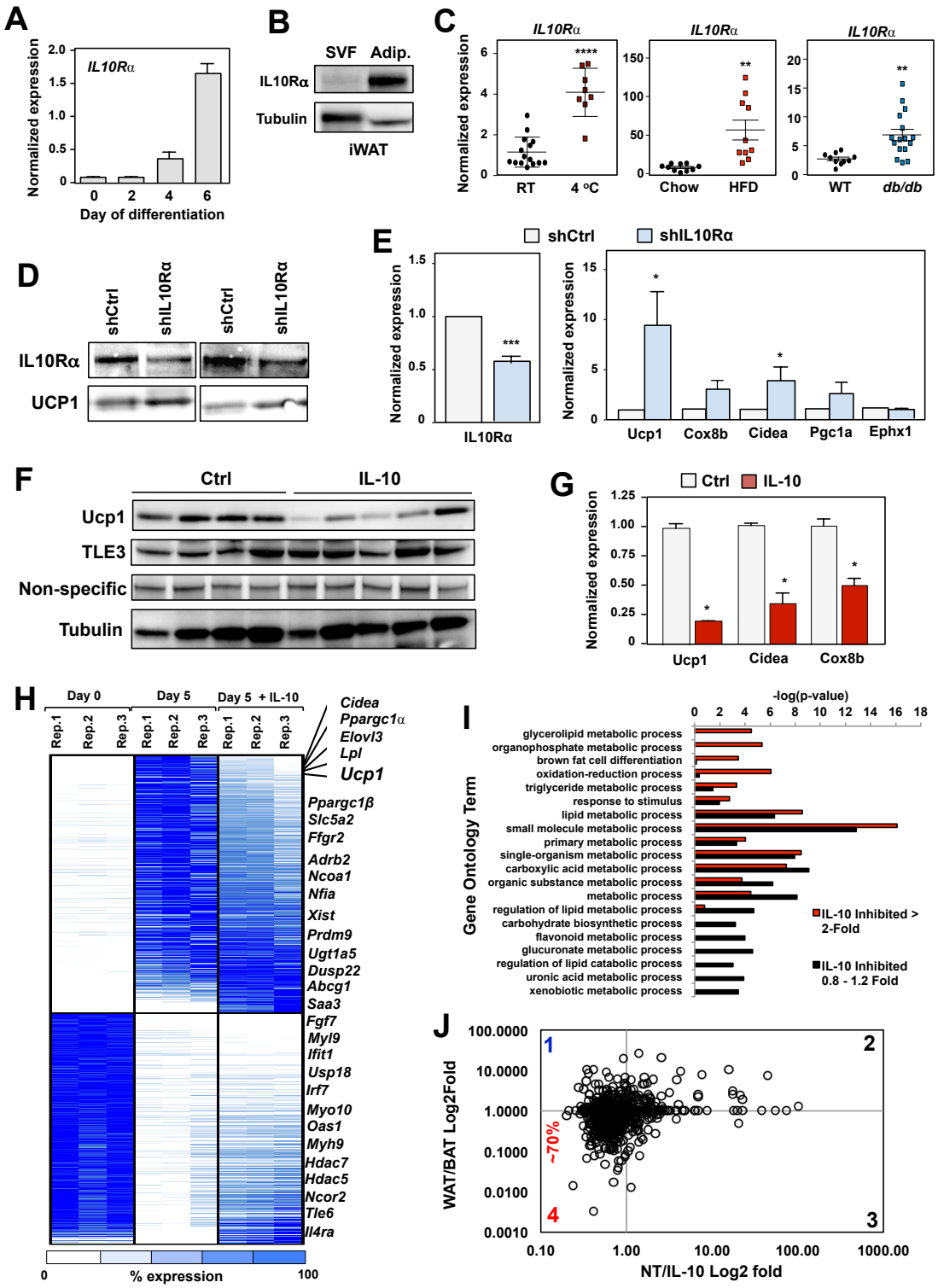


Figure3-6: IL-10 signaling remodels chromatin architecture at thermogenic genes

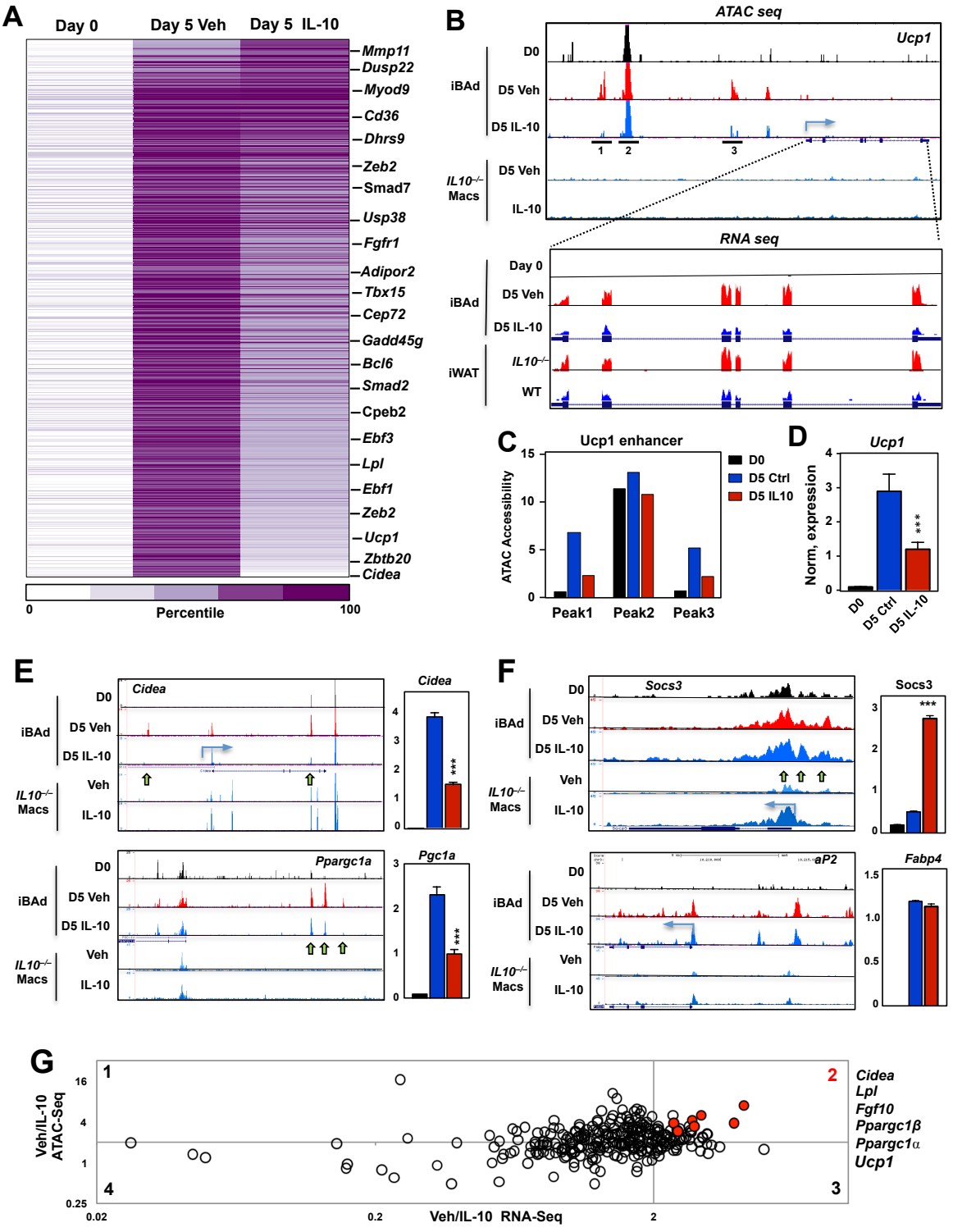


Figure 3-7: IL-10 limits enrichment of thermogenic transcriptional regulators at key enhancer elements

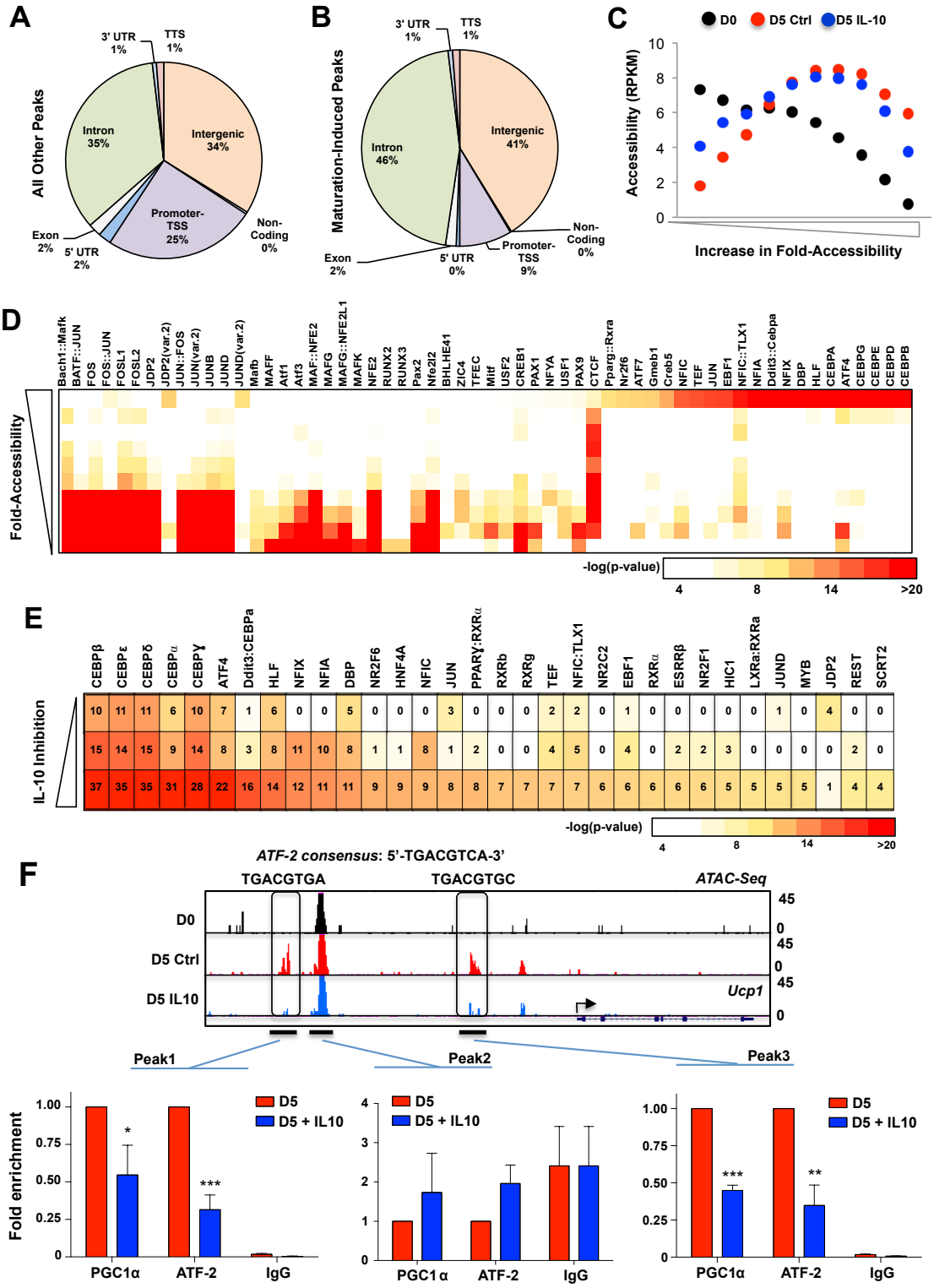


Figure 3-8: Protection against age-related obesity and absence of systemic inflammation in *IL-10-/-* mice

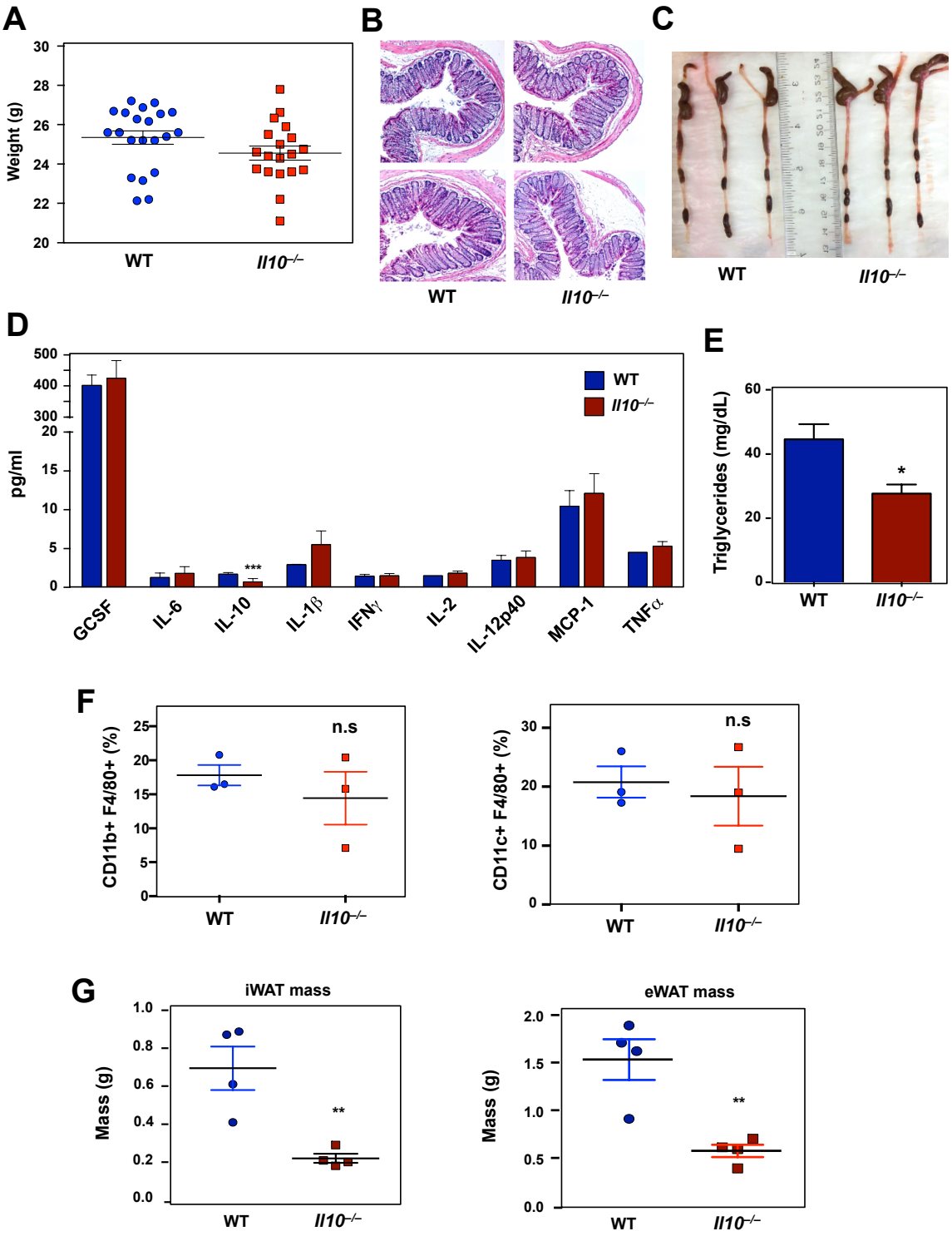


Figure 3-9: Protection against diet-induced obesity and absence of systemic inflammation in high-fat fed *IL-10*-/- mice

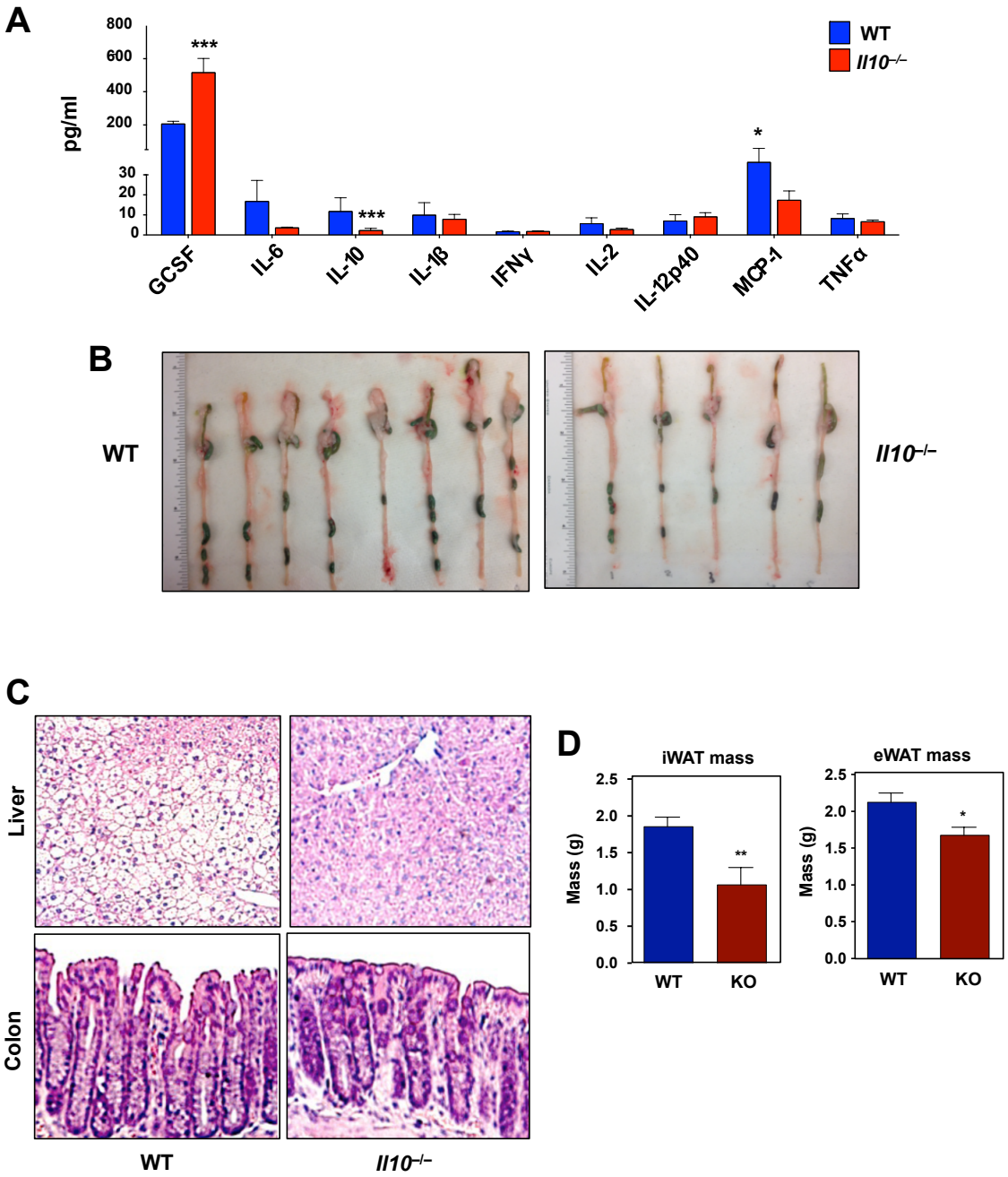


Figure 3-10: Decreased blood glucose and absence of ulcerative colitis in bone-marrow transplanted mice

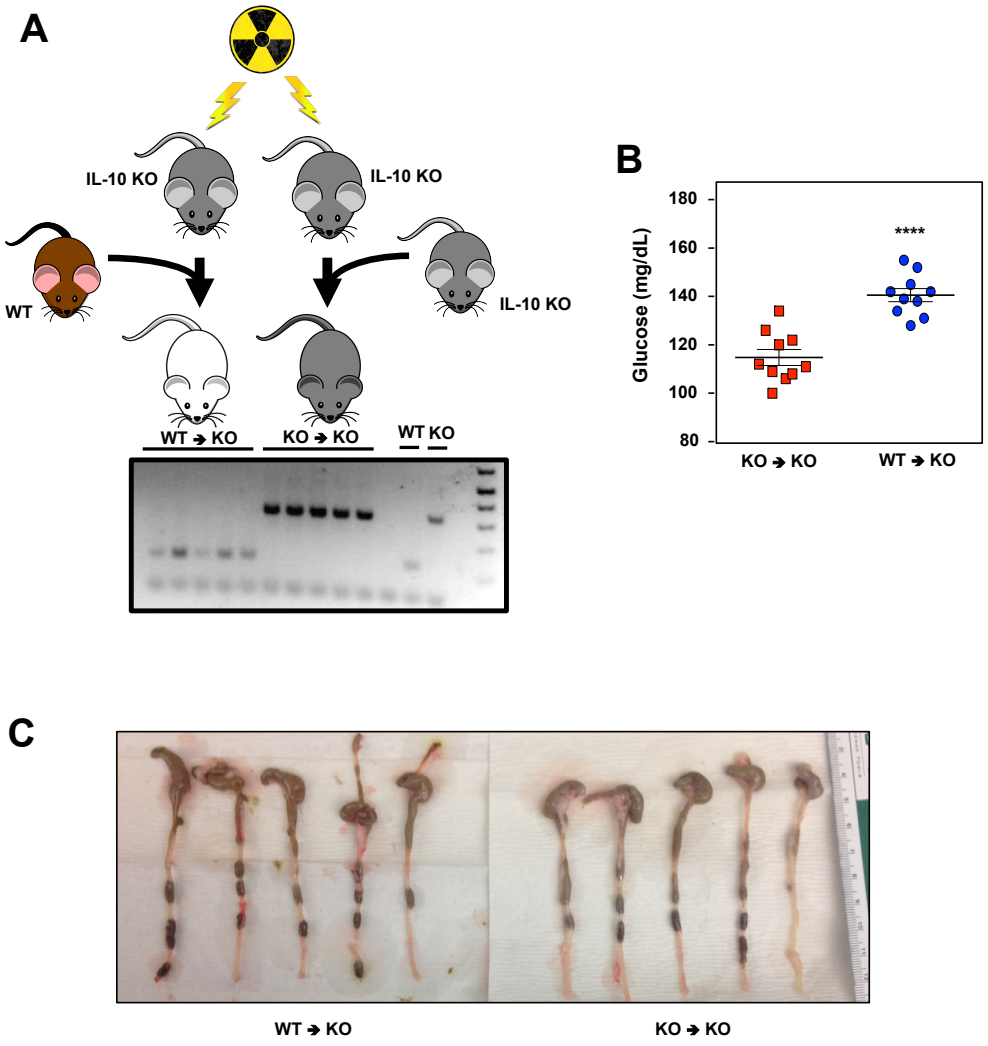


Figure 3-11: IL10Rα expression in various adipogenic models and direct inhibition of thermogenic genes and cellular respiration by recombinant IL-10

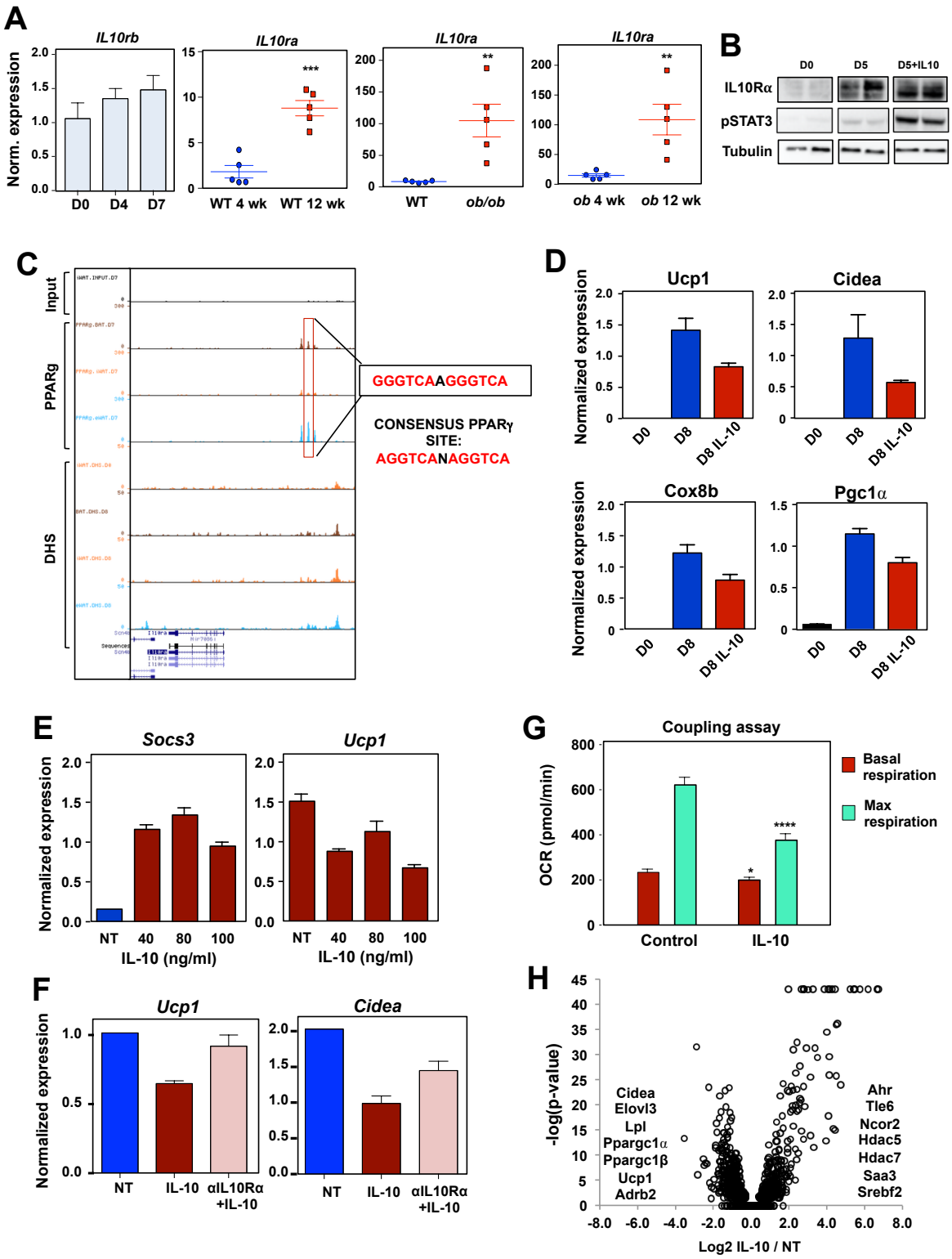
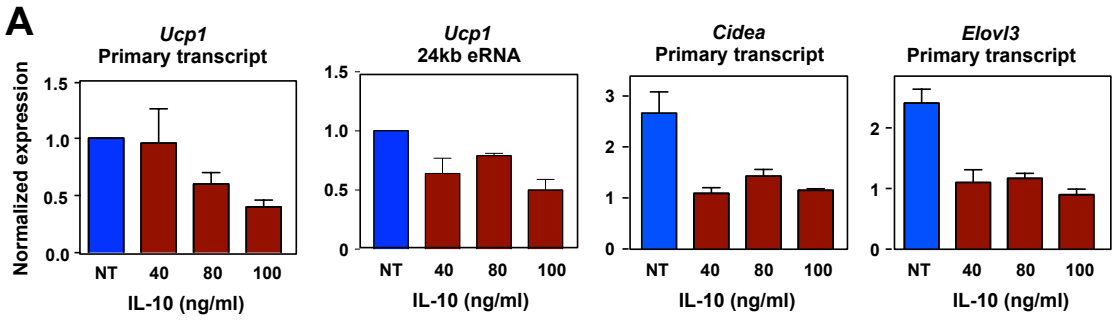
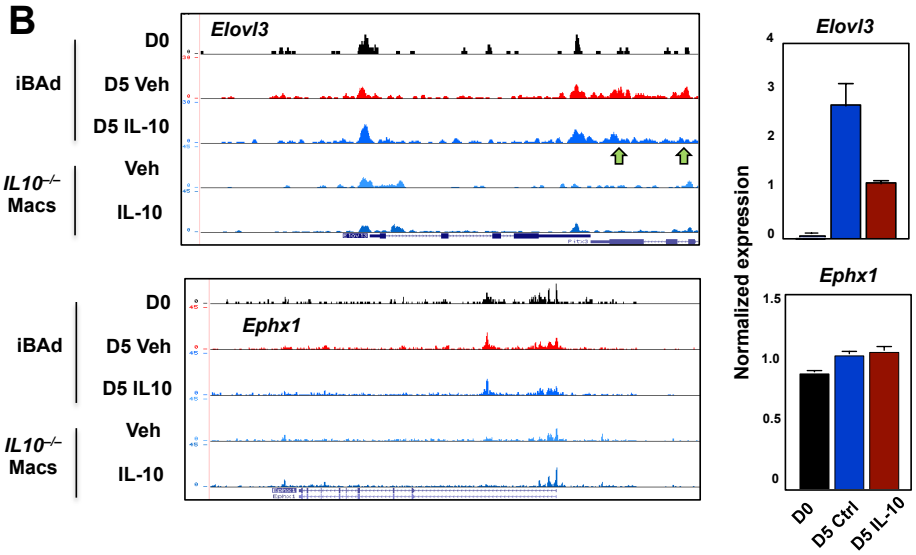
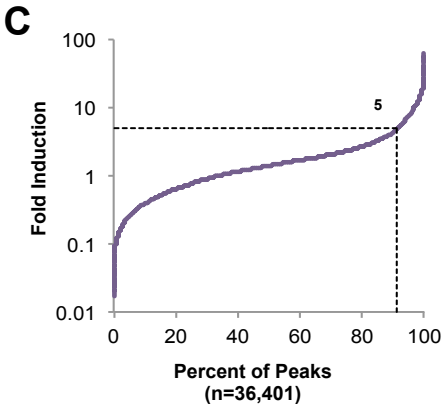


Figure 3-12: Inhibition of thermogenic gene primary transcript expression and chromatin accessibilities by IL-10







D

	All Others	Mature
	32762	3639
Intergenic	34.6	41.2
Non-Coding	0.3	0.2
Promoter-TSS	23.4	8.6
5' UTR	2.0	0.6
Exon	2.1	1.8
Intron	35.7	45.7
3' UTR	0.7	0.7
TTS	1.2	1.2

References

Abe, Y., Rozqie, R., Matsumura, Y., Kawamura, T., Nakaki, R., Tsurutani, Y., Tanimura-Inagaki, K., Shiono, A., Magoori, K., Nakamura, K.*, et al.* (2015). JMJD1A is a signal-sensing scaffold that regulates acute chromatin dynamics via SWI/SNF association for thermogenesis. Nature Communications *6*, 7052.

Attie, A.D., and Scherer, P.E. (2009). Adipocyte metabolism and obesity. J Lipid Res *50 Suppl*, S395-399.

Bapat, S.P., Myoung Suh, J., Fang, S., Liu, S., Zhang, Y., Cheng, A., Zhou, C., Liang, Y., LeBlanc, M., Liddle, C.*, et al.* (2015). Depletion of fat-resident Treg cells prevents age-associated insulin resistance. Nature *528*, 137-141.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) *57*, 289-300.

Brestoff, J.R., Kim, B.S., Saenz, S.A., Stine, R.R., Monticelli, L.A., Sonnenberg, G.F., Thome, J.J., Farber, D.L., Lutfy, K., Seale, P.*, et al.* (2015). Group 2 innate lymphoid cells promote beiging of white adipose tissue and limit obesity. Nature *519*, 242-246.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-29.

Camp, H.S., Ren, D., and Leff, T. (2002). Adipogenesis and fat-cell function in obesity and diabetes. Trends Mol Med *8*, 442-447.

Cao, W., Daniel, K.W., Robidoux, J., Puigserver, P., Medvedev, A.V., Bai, X., Floering, L.M., Spiegelman, B.M., and Collins, S. (2004). p38 mitogen-activated protein kinase is the central regulator of cyclic AMP-dependent transcription of the brown fat uncoupling protein 1 gene. Mol Cell Biol *24*, 3057-3067.

Chen, W., Yang, Q., and Roeder, R.G. (2009). Dynamic interactions and cooperative functions of PGC-1alpha and MED1 in TRalpha-mediated activation of the brown-fat-specific UCP-1 gene. Mol Cell *35*, 755-768.

Couper, K.N., Blount, D.G., and Riley, E.M. (2008). IL-10: the master regulator of immunity to infection. J Immunol *180*, 5771-5777.

den Boer, M.A., Voshol, P.J., Schroder-van der Elst, J.P., Korsheninnikova, E., Ouwens, D.M., Kuipers, F., Havekes, L.M., and Romijn, J.A. (2006). Endogenous interleukin-10 protects against hepatic steatosis but does not improve insulin sensitivity during high-fat feeding in mice. Endocrinology *147*, 4553-4558.

El Kasmi, K.C., Smith, A.M., Williams, L., Neale, G., Panopoulos, A.D., Watowich, S.S., Hacker, H., Foxwell, B.M., and Murray, P.J. (2007). Cutting edge: A transcriptional repressor and

corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway. J Immunol *179*, 7215-7219.

Feldmann, H.M., Golozoubova, V., Cannon, B., and Nedergaard, J. (2009). UCP1 ablation induces obesity and abolishes diet-induced thermogenesis in mice exempt from thermal stress by living at thermoneutrality. Cell Metab *9*, 203-209.

Fujimoto, S., Goda, T., and Mochizuki, K. (2011). In vivo evidence of enhanced di-methylation of histone H3 K4 on upregulated genes in adipose tissue of diabetic db/db mice. Biochem Biophys Res Commun *404*, 223-227.

Fujisaka, S., Usui, I., Bukhari, A., Ikutani, M., Oya, T., Kanatani, Y., Tsuneyama, K., Nagai, Y., Takatsu, K., Urakaze, M.*, et al.* (2009). Regulatory mechanisms for adipose tissue M1 and M2 macrophages in diet-induced obese mice. Diabetes *58*, 2574-2582.

Gao, M., Zhang, C., Ma, Y., Bu, L., Yan, L., and Liu, D. (2013). Hydrodynamic delivery of mIL10 gene protects mice from high-fat diet-induced obesity and glucose intolerance. Mol Ther *21*, 1852-1861.

Guilherme, A., Virbasius, J.V., Puri, V., and Czech, M.P. (2008). Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. Nat Rev Mol Cell Biol *9*, 367-377.

Hammer, M., Mages, J., Dietrich, H., Servatius, A., Howells, N., Cato, A.C., and Lang, R. (2006). Dual specificity phosphatase 1 (DUSP1) regulates a subset of LPS-induced genes and protects mice from lethal endotoxin shock. J Exp Med *203*, 15-20.

Harms, M., and Seale, P. (2013). Brown and beige fat: development, function and therapeutic potential. Nat Med *19*, 1252-1263.

Hong, E.G., Ko, H.J., Cho, Y.R., Kim, H.J., Ma, Z., Yu, T.Y., Friedline, R.H., Kurt-Jones, E., Finberg, R., Fischer, M.A.*, et al.* (2009). Interleukin-10 prevents diet-induced insulin resistance by attenuating macrophage and cytokine response in skeletal muscle. Diabetes *58*, 2525-2535.

Hummasti, S., and Tontonoz, P. (2006). The peroxisome proliferator-activated receptor N-terminal domain controls isotype-selective gene expression and adipogenesis. Molecular endocrinology *20*, 1261-1275.

Hutchins, A.P., Diez, D., and Miranda-Saavedra, D. (2013). The IL-10/STAT3-mediated anti-inflammatory response: recent developments and future challenges. Brief Funct Genomics *12*, 489-498.

Keubler, L.M., Buettner, M., Hager, C., and Bleich, A. (2015). A Multihit Model: Colitis Lessons from the Interleukin-10-deficient Mouse. Inflamm Bowel Dis *21*, 1967-1975.

Kowalski, G.M., Nicholls, H.T., Risis, S., Watson, N.K., Kanellakis, P., Bruce, C.R., Bobik, A., Lancaster, G.I., and Febbraio, M.A. (2011). Deficiency of haematopoietic-cell-derived IL-10 does not exacerbate high-fat-diet-induced inflammation or insulin resistance in mice. Diabetologia *54*, 888-899.

Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K., and Muller, W. (1993). Interleukin-10-deficient mice develop chronic enterocolitis. Cell *75*, 263-274.

Kuwata, H., Watanabe, Y., Miyoshi, H., Yamamoto, M., Kaisho, T., Takeda, K., and Akira, S. (2003). IL-10-inducible Bcl-3 negatively regulates LPS-induced TNF-alpha production in macrophages. Blood *102*, 4123-4129.

Lang, R., Patel, D., Morris, J.J., Rutschman, R.L., and Murray, P.J. (2002). Shaping gene expression in activated and resting primary macrophages by IL-10. J Immunol *169*, 2253-2263.

Lee, J.T., Pamir, N., Liu, N.C., Kirk, E.A., Averill, M.M., Becker, L., Larson, I., Hagman, D.K., Foster-Schubert, K.E., van Yserloo, B.*, et al.* (2014). Macrophage metalloelastase (MMP12) regulates adipose tissue expansion, insulin sensitivity, and expression of inducible nitric oxide synthase. Endocrinology *155*, 3409-3420.

Lee, M.W., Odegaard, J.I., Mukundan, L., Qiu, Y., Molofsky, A.B., Nussbaum, J.C., Yun, K., Locksley, R.M., and Chawla, A. (2015). Activated type 2 innate lymphoid cells regulate beige fat biogenesis. Cell *160*, 74-87.

Lee, Y.H., Petkova, A.P., Mottillo, E.P., and Granneman, J.G. (2012). In vivo identification of bipotential adipocyte progenitors recruited by beta3-adrenoceptor activation and high-fat feeding. Cell Metab *15*, 480-491.

Long, J.Z., Svensson, K.J., Bateman, L.A., Lin, H., Kamenecka, T., Lokurkar, I.A., Lou, J., Rao, R.R., Chang, M.R., Jedrychowski, M.P.*, et al.* (2016). The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. Cell *166*, 424-435.

Lumeng, C.N., Bodzin, J.L., and Saltiel, A.R. (2007). Obesity induces a phenotypic switch in adipose tissue macrophage polarization. J Clin Invest *117*, 175-184.

Mauer, J., Chaurasia, B., Goldau, J., Vogt, M.C., Ruud, J., Nguyen, K.D., Theurich, S., Hausen, A.C., Schmitz, J., Bronneke, H.S.*, et al.* (2014). Signaling by IL-6 promotes alternative activation of macrophages to limit endotoxemia and obesity-associated resistance to insulin. Nat Immunol *15*, 423-430.

Miller, A.M., Wang, H., Bertola, A., Park, O., Horiguchi, N., Ki, S.H., Yin, S., Lafdil, F., and Gao, B. (2011). Inflammation-associated interleukin-6/signal transducer and activator of transcription 3 activation ameliorates alcoholic and nonalcoholic fatty liver diseases in interleukin-10-deficient mice. Hepatology *54*, 846-856.

Moore, K.W., de Waal Malefyt, R., Coffman, R.L., and O'Garra, A. (2001). Interleukin-10 and the interleukin-10 receptor. Annu Rev Immunol *19*, 683-765.

Murray, P.J. (2005). The primary mechanism of the IL-10-regulated antiinflammatory response is to selectively inhibit transcription. Proc Natl Acad Sci U S A *102*, 8686-8691.

Murray, P.J., and Smale, S.T. (2012). Restraint of inflammatory signaling by interdependent strata of negative regulatory pathways. Nat Immunol *13*, 916-924.

Nakata, M., Yamamoto, S., Okada, T., Gantulga, D., Okano, H., Ozawa, K., and Yada, T. (2016). IL-10 gene transfer upregulates arcuate POMC and ameliorates hyperphagia, obesity and diabetes by substituting for leptin. Int J Obes (Lond) *40*, 425-433.

Odegaard, J.I., Lee, M.W., Sogawa, Y., Bertholet, A.M., Locksley, R.M., Weinberg, D.E., Kirichok, Y., Deo, R.C., and Chawla, A. (2016). Perinatal Licensing of Thermogenesis by IL-33 and ST2. Cell *166*, 841-854.

Ohno, H., Shinoda, K., Spiegelman, B.M., and Kajimura, S. (2012). PPARgamma agonists induce a white-to-brown fat conversion through stabilization of PRDM16 protein. Cell Metab *15*, 395-404. Osborn, O., and Olefsky, J.M. (2012). The cellular and signaling networks linking the immune system and metabolism in disease. Nat Med *18*, 363-374.

Park, M., Yi, J.W., Kim, E.M., Yoon, I.J., Lee, E.H., Lee, H.Y., Ji, K.Y., Lee, K.H., Jang, J.H., Oh, S.S.*, et al.* (2015). Triggering receptor expressed on myeloid cells 2 (TREM2) promotes adipogenesis and diet-induced obesity. Diabetes *64*, 117-127.

Petruzzelli, M., Schweiger, M., Schreiber, R., Campos-Olivas, R., Tsoli, M., Allen, J., Swarbrick, M., Rose-John, S., Rincon, M., Robertson, G.*, et al.* (2014). A switch from white to brown fat increases energy expenditure in cancer-associated cachexia. Cell Metab *20*, 433-447.

Qiu, Y., Nguyen, K.D., Odegaard, J.I., Cui, X., Tian, X., Locksley, R.M., Palmiter, R.D., and Chawla, A. (2014). Eosinophils and type 2 cytokine signaling in macrophages orchestrate development of functional beige fat. Cell *157*, 1292-1308.

Rao, R.R., Long, J.Z., White, J.P., Svensson, K.J., Lou, J., Lokurkar, I., Jedrychowski, M.P., Ruas, J.L., Wrann, C.D., Lo, J.C.*, et al.* (2014). Meteorin-like is a hormone that regulates immune-adipose interactions to increase beige fat thermogenesis. Cell *157*, 1279-1291.

Rogers, G.W., Brand, M.D., Petrosyan, S., Ashok, D., Elorza, A.A., Ferrick, D.A., and Murphy, A.N. (2011). High throughput microplate respiratory measurements using minimal quantities of isolated mitochondria. PLoS One *6*, e21746.

Romeo, G.R., Lee, J., and Shoelson, S.E. (2012). Metabolic syndrome, insulin resistance, and roles of inflammation--mechanisms and therapeutic targets. Arterioscler Thromb Vasc Biol *32*, 1771-1776.

Rosell, M., Kaforou, M., Frontini, A., Okolo, A., Chan, Y.W., Nikolopoulou, E., Millership, S., Fenech, M.E., MacIntyre, D., Turner, J.O.*, et al.* (2014). Brown and white adipose tissues: intrinsic differences in gene expression and response to cold exposure in mice. Am J Physiol Endocrinol Metab *306*, E945-964.

Rosen, E.D., and Spiegelman, B.M. (2014). What we talk about when we talk about fat. Cell *156*, 20-44.

Rosenwald, M., Perdikari, A., Rulicke, T., and Wolfrum, C. (2013). Bi-directional interconversion of brite and white adipocytes. Nat Cell Biol *15*, 659-667.

Rossato, M., Curtale, G., Tamassia, N., Castellucci, M., Mori, L., Gasperini, S., Mariotti, B., De Luca, M., Mirolo, M., Cassatella, M.A.*, et al.* (2012). IL-10-induced microRNA-187 negatively regulates TNF-alpha, IL-6, and IL-12p40 production in TLR4-stimulated monocytes. Proc Natl Acad Sci U S A *109*, E3101-3110.

Saraiva, M., and O'Garra, A. (2010). The regulation of IL-10 production by immune cells. Nat Rev Immunol *10*, 170-181.

Schaljo, B., Kratochvill, F., Gratz, N., Sadzak, I., Sauer, I., Hammer, M., Vogl, C., Strobl, B., Muller, M., Blackshear, P.J.*, et al.* (2009). Tristetraprolin is required for full anti-inflammatory response of murine macrophages to IL-10. J Immunol *183*, 1197-1206.

Seale, P., Kajimura, S., and Spiegelman, B.M. (2009). Transcriptional control of brown adipocyte development and physiological function--of mice and men. Genes Dev *23*, 788-797. Shoelson, S.E., Herrero, L., and Naaz, A. (2007). Obesity, inflammation, and insulin resistance. Gastroenterology *132*, 2169-2180.

Shoelson, S.E., Lee, J., and Goldfine, A.B. (2006). Inflammation and insulin resistance. J Clin Invest *116*, 1793-1801.

Siersbaek, M.S., Loft, A., Aagaard, M.M., Nielsen, R., Schmidt, S.F., Petrovic, N., Nedergaard, J., and Mandrup, S. (2012). Genome-wide profiling of peroxisome proliferator-activated receptor gamma in primary epididymal, inguinal, and brown adipocytes reveals depot-selective binding correlated with gene expression. Mol Cell Biol *32*, 3452-3463.

Smith, A.M., Qualls, J.E., O'Brien, K., Balouzian, L., Johnson, P.F., Schultz-Cherry, S., Smale, S.T., and Murray, P.J. (2011). A distal enhancer in Il12b is the target of transcriptional repression by the STAT3 pathway and requires the basic leucine zipper (B-ZIP) protein NFIL3. J Biol Chem *286*, 23582-23590.

Sun, L., Goff, L.A., Trapnell, C., Alexander, R., Lo, K.A., Hacisuleyman, E., Sauvageau, M., Tazon-Vega, B., Kelley, D.R., Hendrickson, D.G.*, et al.* (2013). Long noncoding RNAs regulate adipogenesis. Proc Natl Acad Sci U S A *110*, 3387-3392.

Tiraby, C., and Langin, D. (2003). Conversion from white to brown adipocytes: a strategy for the control of fat mass? Trends Endocrinol Metab *14*, 439-441.

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

Townsend, K.L., and Tseng, Y.H. (2014). Brown fat fuel utilization and thermogenesis. Trends Endocrinol Metab.

van Marken Lichtenbelt, W.D., Vanhommerig, J.W., Smulders, N.M., Drossaerts, J.M., Kemerink, G.J., Bouvy, N.D., Schrauwen, P., and Teule, G.J. (2009). Cold-activated brown adipose tissue in healthy men. N Engl J Med *360*, 1500-1508.

Villanueva, C.J., Vergnes, L., Wang, J., Drew, B.G., Hong, C., Tu, Y., Hu, Y., Peng, X., Xu, F., Saez, E.*, et al.* (2013). Adipose subtype-selective recruitment of TLE3 or Prdm16 by PPARgamma specifies lipid storage versus thermogenic gene programs. Cell Metab *17*, 423-435.

Villanueva, C.J., Waki, H., Godio, C., Nielsen, R., Chou, W.L., Vargas, L., Wroblewski, K., Schmedt, C., Chao, L.C., Boyadjian, R.*, et al.* (2011). TLE3 is a dual-function transcriptional coregulator of adipogenesis. Cell Metab *13*, 413-427.

Virtanen, K.A., Lidell, M.E., Orava, J., Heglind, M., Westergren, R., Niemi, T., Taittonen, M., Laine, J., Savisto, N.J., Enerback, S.*, et al.* (2009). Functional brown adipose tissue in healthy adults. N Engl J Med *360*, 1518-1525.

Wallenius, V., Wallenius, K., Ahren, B., Rudling, M., Carlsten, H., Dickson, S.L., Ohlsson, C., and Jansson, J.O. (2002). Interleukin-6-deficient mice develop mature-onset obesity. Nat Med *8*, 75-79.

Wernstedt Asterholm, I., Tao, C., Morley, T.S., Wang, Q.A., Delgado-Lopez, F., Wang, Z.V., and Scherer, P.E. (2014). Adipocyte inflammation is essential for healthy adipose tissue expansion and remodeling. Cell Metab *20*, 103-118.

Wu, J., Bostrom, P., Sparks, L.M., Ye, L., Choi, J.H., Giang, A.H., Khandekar, M., Virtanen, K.A., Nuutila, P., Schaart, G.*, et al.* (2012). Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. Cell *150*, 366-376.

Wu, M., Neilson, A., Swift, A.L., Moran, R., Tamagnine, J., Parslow, D., Armistead, S., Lemire, K., Orrell, J., Teich, J.*, et al.* (2007). Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells. Am J Physiol Cell Physiol *292*, C125-136.

Xie, L., Fu, Q., Ortega, T.M., Zhou, L., Rasmussen, D., O'Keefe, J., Zhang, K.K., and Chapes, S.K. (2014). Overexpression of IL-10 in C2D macrophages promotes a macrophage phenotypic switch in adipose tissue environments. PLoS One *9*, e86541.

Xu, X., Grijalva, A., Skowronski, A., van Eijk, M., Serlie, M.J., and Ferrante, A.W., Jr. (2013). Obesity activates a program of lysosomal-dependent lipid metabolism in adipose tissue macrophages independently of classic activation. Cell Metab *18*, 816-830.

Ye, L., Wu, J., Cohen, P., Kazak, L., Khandekar, M.J., Jedrychowski, M.P., Zeng, X., Gygi, S.P., and Spiegelman, B.M. (2013). Fat cells directly sense temperature to activate thermogenesis. Proc Natl Acad Sci U S A *110*, 12480-12485.

Zambelli, F., Pesole, G., and Pavesi, G. (2009). Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. Nucleic Acids Res *37*, W247-252.

CHAPTER 4

Highly Selective Changes in Chromatin Accessibility in Response to Pro-Inflammatory and Anti-Inflammatory Stimuli

Summary

Chromatin remodeling events that alter the accessibility of DNA to transcriptional regulators are known play important roles in the transcriptional response of mammalian cells to extracellular stimuli. However, much remains to be learned about the extent to which chromatin remodeling contributes to the activation of promoters and enhancers for inducible genes, as well as the determinants of chromatin remodeling. Here, we used Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) to characterize chromatin changes that occur in mouse macrophages following stimulation with both pro- and anti-inflammatory agents. Although small changes in transposase access occurred at most inducible control regions, large changes that appear to represent robust nucleosome remodeling events were observed at only a few dozen inducible promoters, but at several thousand intergenic sites. Nucleosome remodeling at promoters was associated with highly specific transcriptional regulatory events. Notably, whereas NF-kB largely binds inducible promoters with constitutively accessible chromatin, NF-kB binding motifs were greatly enriched at distant enhancers that exhibited robust remodeling, suggesting that NF-kB either contributes to inducible remodeling or that inaccessible nucleosomes play a common role in restricting access of NF-κB to intergenic enhancers. An analysis of the impact of the potent anti-inflammatory cytokine, IL-10, revealed pronounced effects on chromatin accessibility at a surprisingly small number of enhancers in close proximity to the target genes that are most strongly suppressed.

Introduction

Eukaryotic control of inducible gene expression involves a complex array of regulators including stimulus induced transcription factors (TFs), the pre-configured distribution of TF consensus sites scattered throughout the genome, and the accessibility of those consensus sites to TF binding. Chromatin accessibility is a property that arises from dynamic DNA:histone associations that are both cell type- and signal-dependent (Weintraub and Groudine, 1976). Accessible regions of DNA form in regions of high transcriptional activity, and result from the association of DNA binding factors, histone variants, and permissive histone modifications thought to either directly influence nucleosome dynamics or establish a favorable environment for TF binding and accessibility via indirect mechanisms (Calo and Wysocka, 2013). Some TFs contain "pioneering" abilities and exhibit a partial or complete affinity for target motifs embedded within stable nucleosomes. Pioneer factor binding is thought to alter the DNA-histone structure to allow for subsequent TF binding from occurring and further establish regions of open chromatin. When signal-dependent non-pioneer factors enter the nucleus upon activation, they rapidly associate with target motifs previously made available for binding and are restricted in their ability bind motifs in a closed chromatin environment. Nucleosome remodeling can be achieved through a number of mechanisms, but the functional consequences converge on the same goal of allowing TF recognition of target consensus sites and subsequent participation in regulating gene expression.

Bone marrow-derived macrophages have been extensively studied as a model system for understanding mechanisms of transcriptional control. Microarrays and RNA-Sequencing (RNA-Seq) has been used to monitor global gene expression changes in response to stimulation, and a number of important insights relating to transcriptional control have emerged from these studies. When macrophages become active, they potently induced hundreds of genes demonstrating early transient, sustain, and late induction kinetics. Early-induced Lipopolysacchardie (LPS) or lipid A response genes are enriched for CpG island promoters and contain feature suggestive of open chromatin including basal DNAse I sensitivity, RNA Pol II

association, and permissive histone marks (Hargreaves et al., 2009; Ramirez-Carrozzi et al., 2009). Late induced genes contain a greater proportion of the cyclohexamide (CHX)-sensitive secondary response genes (SRGs) that are more likely to contain promoters depleted of CpG dinucleotides and require interferon signaling or yet to be discovered inducible protein activators for their expression. The low CpG content at these promoters favor stable nucleosome assembly, which adds an additional barrier for transcription at these sites. Previously, we have used model gene studies on the *Il12b* gene promoter to understand how chromatin dynamics occur in response to stimulation. Rapid remodeling of a well-positioned nucleosome over the transcriptional start site (TSS) of *Il12b* precedes or coincides with transcriptional activation (Weinmann et al., 1999). The ability to classify genes with the additional requirement for nucleosome remodeling will be key to better understanding the stepwise progression of gene activation, and provide a framework for carrying out future mechanistic studies.

Methods of profiling accessible chromatin genome-wide became possible with the advent of DNAse Hypersensitive Site sequencing (Crawford et al., 2005), and later with the Assay for Transposase Accessible Chromatin (ATAC-Seq) (Buenrostro et al., 2013). Others have utilized this approach to study determinants of enhancer selection in macrophages from distinct developmental lineages and in response to environmental signals (Gosselin et al., 2014; Lavin et al., 2014). What has not been extensively studied is how chromatin accessibility changes occur in response to an acute stimulation. We therefore have utilized ATAC-Seq to monitor chromatin changes in macrophages following a short time course of stimulation and analyze the data in context with the large body of work that has already been performed. This framework is then used to address how an important physiologic inhibitor of inflammation, IL-10, influences the induction of open chromatin. This analysis has reveled the selectivity in potent sensitivity to IL-10 treatment at both the level of transcriptional inhibition and repression of inducible ATAC sites, and will be useful for further mechanistic studies aimed at identifying the players responsible for achieving the observed selectivity.

Results

Mapping Inducible Chromatin Changes During an Acute Stimulation

To gain insights into the properties, regulation, and function of inducible chromatin during the macrophage response to stimulation, we first performed ATAC-seq in BMDMs treated with lipid A for 0, 30, 60, and 120 minutes. From an analysis of at least two biological replicates per time point, we identified 101,448 called peaks (FDR <0.01) of accessible chromatin during the time course. 11,456 (11.3%) peaks were significantly induced and 4,292 peaks were identified as significantly repressed ($p<0.05$, Figure 1A). The vast majority of the remaining peaks (81.9%) were constitutive and changed less than 2 fold of their basal accessibility. The induced ATAC peaks ranged in fold accessibility as low as 1.4 to as high as 68.3, but 64.5% of all induced peaks were less than 5 fold induced. This continuum of fold accessibility likely indicates that either some induced peaks occur in regions of pre-existing open chromatin that undergo either small structural changes of the nucleosome or these small changes are not reflective of chromatin modification and instead represent increased accessibility due to TF binding. Consistent with this notion, 77.5% of peaks induced < 3 fold had a detectable ATAC peak in the basal state as compared to 5.3% of peaks induced > 5 fold (data not shown). When scanning the DNA of the called peaks we observed a bias in CpG di-nucleotide content across peaks classified as stable or dynamic (induced and repressed). While the mean CpG density demonstrate an aggregate trend towards lower CpG content at inducible sites, there is a marked reduction of high CpG content sites in the induced regions, and to a lesser extent, the repressed sites compared to constitutive sites that change <2 fold (Figure 1E).

We next addressed where the accessible regions reside in the genome, with a specific interest in understanding how inducible peaks differ from constitutive sites. Approximately 10% of all peaks overlap promoters, and an additional 44% and 40% occur within introns and intergenic sites respectively. Inducible peaks less frequently associate with promoters (3%), with the majority of sites contained within introns or at intergenic sites (93%). Repressed peaks demonstrate an intermediate association with promoters as compared to all peaks and inducible

peaks (Figure 1B), and represent a small fraction (2-5%) of the type of ATAC peak found across all genomic locations (Figure 1D). Of all classified regions, promoters most frequently associated with stable chromatin, with 94% called constitutive, 4% induced, and 2.5% repressed. Upon visual inspection of the promoter-associated peaks we observed that some inducible peaks overlap the transcriptional unit of known lipid A induced genes. These peaks often extend beyond the length of the annotated gene and likely reflect broad accessibility changes reflective of activate transcription rather than TF binding or nucleosome remodeling (Figure 1C), highlighting the need to additional scrutiny of peaks classified as inducible.

Initial Analysis of Inducible Chromatin and Transcription Factor Binding Sites

To complement the statistical approach used to initially characterize inducible peaks, we ranked all called peaks based on their maximum fold increase in accessibility and generated 35 bins of approximately 2,900 peaks / bin. We then used these bins to scan for an over representation of TF motifs and to understand how these motifs are distributed across peaks demonstrating different levels of fold accessibility. Two-hundred motifs showed an enrichment above 5 (-log(p-value)), with multiple motifs within the same TF family demonstrating similar trends of enrichment (Figure 2A). Of interest, while a broad range of ETS family members (Spi1, ETV6, ELF5, and others) were enriched across many bins consistent with their role in macrophage development, we observed enrichments for lipid A-inducible TF in bins corresponding to inducible peaks. AP1 family member motifs began showing enrichments in the weakly induced bins and continued to have enrichment in the stronger induced bins, although the strongest enrichments were found in bins corresponding to 2.5 - 3.5 fold induction. NF-κB family members only demonstrated enrichments in the most strongly induced ATAC sites, with the last bin showing the highest enrichment. This finding is unexpected, as many of the strong RelA:p50 binding to promoters of induced genes occurs in regions of open chromatin (Tong et al., 2016). CTCF motifs displayed a preference for peaks that remain stable during lipid A stimulation, and

IRF1 and STAT1 motifs were found to be enriched in peaks that are remain unchanged or are slightly repressed (Figure 2C).

When we analyze the genomic distribution of peaks binned according to fold induction we again observe a bias for promoter-associated peaks to be found in bins demonstrating low response to stimulation and intergenic/intronic sites preferentially associated with sites that are either induced or repressed (Figure 2C). Bin 35 has the lowest percentage of peaks that overlap promoters (1.6%), which prompted us to investigate in detail which promoters exhibit this unique characteristic of inducible accessibility.

Inducible Promoter Accessibility

We have previously shown that strong promoter changes in chromatin accessibility rarely occur during the induction of the most potently induced primary response genes in BMDM. In the broader analysis shown here, we found that only 52 of 21,168 annotated promoters within the mouse genome exhibit an ATAC-seq peak that is induced more than 5-fold during the lipid A time-course (i.e. an inducible peak that overlaps the region from -500 to +150 relative to the TSS). Upon visual inspection of the 52 peaks, we observed that some genes had a mis-annotated TSS, were due to transcriptional effects, or were not called statistically induced. Removing these sites left 36 candidate promoters that have a large fold increase in accessibility.

Eleven of these promoters were previously implicated in demonstrating nucleosome remodeling during the lipid A response (Figure 3A), while the remaining promoters served as an opportunity to learn about the regulation and implications of strong increases in promoter accessibility on target gene expression. To compare ATAC changes with transcriptional changes we utilized previously published chromatin-associated RNA-Seq data. 16 of the remaining ATAC peaks were found to be associated with genes that were potently induced above 10 fold, but overall were weakly expressed (13/16 with RPKM<3) and the remaining 12 peaks were associated with genes that were weakly induced.

The heatmap analysis suggested that inducible promoters correlate with inducible transcription, but to evaluate the magnitude of induction among the two datasets we compared the promoter ATAC fold accessibility to caRNA-Seq fold induction for all PRG and SRG (Figure 3B). Both PRG and SRGs with strong promoter ATAC fold induction exhibited high levels of transcriptional induction. When comparing ATAC induction to promoter CpG content, we observe a similar trend with low CpG density correlating with strong ATAC induction. *Ccl5*, as an example, demonstrates the highest transcriptional induction, largest ATAC fold change, and low CpG content. This trend can be readily seen among SRGs as well, but with a less strict association with RNA-Seq fold induction.

To investigate the regulation of these promoter changes, we performed ATAC-Seq in cells that were pretreated with cyclohexamide (CHX) prior to stimulation. None of the primary response promoters were inhibited during CHX treatment, while 2 of the secondary response gene promoters had diminished accessibility (*Il12b* and *Ifi47,* <30%). The other 4 SRGs maintained induction of accessibility in the presence of CHX suggesting that the secondary response component required for activation of these 4 SRGs might proceed after promoter accessibility has been established. There was no trend in CHX-sensitivity or resistance between SRG that were previously characterized as interferon signaling dependent (IFNAR-D) or those that are interferon independent SRGs (IFNAR-I, Figure 3C).

IRF3 was previously shown to be required for the nucleosome remodeling and subsequent RelA association at the promoter of *Ccl5* (Tong et al., 2016). This strict requirement was unique to only *Ccl5*, as no other PRG with strongly inducible promoter accessibility demonstrated reduced association of RelA in the absence of IRF3. To expand upon this finding and to address whether we can detect IRF3 dependent promoter induction across the additional 35 promoters we performed ATAC-Seq in *Irf3-/-* BMDM stimulated for 0 and 120m with lipid A in the presence or absence of CHX. The CHX pretreatment limits the ability for other IRF family members to partially compensate for the loss of IRF3 in this system. When a direct comparison between WT and *Irf3-/-* BMDM stimulated under the presence of CHX were compared, *Ccl5*

stands out as the only strongly induced promoter with IRF3-dependence. Consistent with this, it is one of the few promoters with detectable association of IRF3 in ChIP-Seq (Figure 3D). Visual inspection of the region surrounding *Ccl5* demonstrates that other peaks in the region upstream of the TSS demonstrate induction in response to stimulation, with some peaks showing diminished accessibility in the *Irf3-/-* cells (Figure 3E).

Inducible Chromatin Changes Upstream of Inducible Transcription

The majority of inducible changes during an acute stimulation occur at intronic and intergenic regions, but because we have observed transcriptional effects that can lead to mis-interpretation of DNA accessibility using ATAC-Seq, we focused our attention to understanding the contribution of intergenic induced ATAC sites to inducible transcription. Prior binning of all peaks revealed TF motifs that preferentially associate with stable peaks (CTCF, PU.1) as well as motifs that are found at induced (AP1, NFκB), and repressed (IRF1, STAT1) sites. A potentially confounding variable is that promoters associate most frequently with stable peaks, therefore motif enrichments could be due to either genomic location or ATAC fold induction. When all intergenic peaks were grouped into 20 equivalently sized bins a similar motif enrichment pattern emerged. Overall CpG content at intergenic sites were lower than those reported for promoter regions, but the trend for stable ATAC sites associating with higher CpG density remained, and inducible peaks showed a similar preference for occurring over low CpG density regions (Figure 4A, 4D).

We next sought to address how frequently inducible intergenic sites are found in close proximity to an induced gene. To perform this analysis, all expressed genes (RPKM>1) were separated into a gene expression category of either demonstrating <2 fold, 2-10 fold, or >10 fold induction. Using the TSS as reference, we counted the number of genes having at least one >5 fold induced ATAC peak at different distance bins located upstream up to 200kb (Figure 4B). The counts were normalized for the different numbers of genes within each category and represented as the percent of genes with an induced ATAC signal within the distance groups.

Strongly induced genes exhibit a much greater likelihood of containing chromatin upstream of their TSS that is strongly induced in response to stimulation. This is made clearer by plotting the fold enrichment of genes in close proximity to an inducible ATAC peak as compared to genes within the <2 fold induced category. Both 2-10 fold and >10 fold induced gene groups demonstrate an enrichment for induced ATAC sties, with the greatest differences observed within the first 10Kb of a gene's TSS. The above-mentioned analysis did not consider the presence of multiple induced ATAC sites within each distance bin. To account for these differences, genes were counted if they contained at least 2 inducible sites within each bin and a percentage was calculated relative to the original number of genes identified to have an induced peak in the above table (Figure 5B, left). The enrichment of genes that have multiple induced ATAC sites within 10Kb of a TSS was even more pronounced for the >10 fold genes (15.7%) compared to 2-10 fold (3.6%).

While we could observe an enrichment of >10 fold genes containing inducible ATAC sites at distances of 100Kb upstream of a TSS, we chose to study the association of peaks within 50Kb of a TSS because this distance represents a >3 fold enrichment over <2 fold induced genes. When scanning the entirety of a 50kb window across the TSSs of genes within each expression category we observe that all classes of genes associate with ATAC sites that remain unchanged during stimulation at a frequency of 25-30% without strong differences among the gene classes (Figure 4E). Differences between the gene classes begin to emerge when comparing the frequency of ATAC sites that are induced between 2-3 fold, and these differences get pronounced when focusing on those peaks that are induced >5 fold. A possible explanation for these findings is that >10 fold induced genes are more frequently found clustered together in the genome, and therefore have the potential to utilize shared regulatory regions upstream of their TSS. To address this possibility, we counted the number of TSS occurrences 50Kb upstream of each gene and separated them according to their expression category. Indeed, 12% of >10 fold induced genes contain an additional TSS of another >10 fold induced gene, as compared to 4% containing a TSS of a 2-10 fold induced gene and 1%

containing a TSS from a <2 fold induced gene. Similarly, a higher proportion of <2 fold induced genes are within 50kb of additional <2 fold gene TSSs (40%), as opposed to a >10 fold induced gene's TSS (19%).

To explore the relationship between numerous inducible ATAC peaks and transcription, we counted the number of ATAC peaks that are within 50Kb upstream of 10 fold induced genes, separating peaks that fall into categories of either 0.9-1.1 fold, 2-3 fold, or >5 fold. While there are far fewer >5 fold induced peaks compared to peaks in the lower fold induction categorized, 35 genes are found to have at least 2 peaks with >5 fold induction, compared to 38 genes with 2-3 fold induced peaks and 12 genes with 0.9-1.1 fold peaks. This correlation becomes more evident as we begin to look at those genes with at least 3 similarly induced peaks upstream of their TSS. Only the >5 fold induced peaks are found in multiples of 5 or more within this distance window, and many of these induced genes show stronger induction magnitudes compared to the genes with multiple peaks belonging to different induction classes (Figure 4F, 4G).

Analysis of NF-κB's Contribution to Inducible Chromatin

One surprising finding from the motif analysis in Figure 4A was that NF-κB motif enrichment was only observed at sites that demonstrated the strong induction following stimulation, while previous analysis of NF-κB to promoters occurred most frequently at sites exhibiting open chromatin properties before stimulation which increased in accessibility by only a few fold. When we looked for evidence of RelA binding to all intergenic ATAC peaks, we observed small enrichment of RelA ChIP-Seq signal across many weakly induced sites, with approximately 2-3 fold enrichment for sites induced >5 fold (data not shown). To investigate the relationship of NF-κB binding and motif strength across peaks demonstrating varying degrees of fold induction, we compared RelA:p50 protein binding microarray motif scores to RelA ChIP-Seq peak scores for ATAC sites that were non-induced (0.9-1.1 fold), weakly induced (2-3 fold) or strongly induced (>5 fold). A motif threshold of >6.4 (strong motif) and peak score of >19 (strong peak) were

used to classify ATAC peaks based on previous analysis (Tong et al., 2016). The PBM threshold demonstrated unique enrichment in the ATAC peaks exhibiting >5 fold induction as compared to RelA binding that occurred in non-induced ATAC sites (Figure 5C). When motif strength and peak strength were directly compared, an enrichment of strong motifs could be clearly seen in the ATAC sites with strong induction (Figure 5A). This enrichment could also be seen to a lesser extent in ATAC sites with weaker induction (Figure 5B).

Interestingly, a strong enrichment was also observed for sites containing strong motifs, but weak or absence of reproducible RelA ChiP-Seq peaks. This enrichment was partially reduced when a lower stringency was used to characterize RelA peaks and only a single replicate had a called RelA peak (Figure 5B, right panel), which could reflect partial detection of RelA interactions due to only a fraction of cells within the population exposing accessible DNA, and/or a fraction of the cells that have strong RelA associations after the DNA is made available for binding.

Kinetic Analysis of NF-κB Binding at Induced Sites

One hypothesis that emerges from these analyses is to question the role of RelA binding to inducible chromatin. Others have proposed a role for NF-κB in binding to nucleosomal DNA (Angelov et al., 2004), but structural studies and other in-vitro experiments have demonstrated a low affinity of NF-κB to nucleosomal DNA (Chen et al., 1998; Saccani et al., 2001). When we compared the kinetics of ATAC induction and RelA association in the 189 ATAC sites with strong motif and strong peak, we observe that the majority of them share strong induction and binding at the 30 minutes time point, with sustained accessibility and increased RelA binding over the 2-hour stimulation (Figure 5D). Consistent with the notion that DNA must first be made accessible for subsequent RelA binding, there were four induced greater than 5 fold by 30m in the ATAC data but had an average peak score of less than 5 at the 30m timepoint. All 4 of these peaks however, had detectable, yet weak, peaks at the 30m timepoint. An additional 5 peaks

were < 2.5 fold induced by 30m and instead exhibited the strongest induction between the 60/30m or the 120/60m time points yet showed average RelA peaks scores >10 at 30m.

Rapid Activation of ATAC Sites and Classification of Primary Response Induction

Of the 498 induced sites that contain a strong RelA motif, 90% show the strongest inducible ATAC changes by 30m of stimulation. This is similar to what is observed across all inducible ATAC sites where 80.8% of peaks demonstrate the largest change after 30 minutes of stimulation (Figure 5E). Accessibility continues to rise for the majority of these, but rapid changes are evident across most of the inducible peaks. To further characterize the regulation of these peaks, we calculated the % of maximum accessibility with a pretreatment of CHX. 37.9% of the inducible peaks were reduced to less than 30% of their maximum induction in WT cells, indicating a role for new protein synthesis in regulating the activation of these sites. Interestingly, 106 of these sites contain strong RelA motifs, while the majority of all other strong motif sites remained CHX resistant (Figure 5F). This finding is consistent with most RelA biding contributing to the primary response, although our data does not rule out the possibility of additional waves of RelA activation leading to binding at CHX sensitive sites after two hours of stimulation.

Inducible peaks were separated based on their requirements for new protein synthesis, strong motifs for RelA, as well as their requirement for IRF3 either under normal conditions and with pretreatment of CHX. Peaks were further separated based on the time point at which they exhibited the most potent induction of accessibility. This classification resulted in the grouping of 9 clusters of peaks and allowed for a qualitative assessment of overall chromatin dynamics. Putative IRF3 dependent peak clusters reached max accessibility later than those with strong RelA motifs (Figure 5G), consistent with what is known about delayed IRF3 activation through the TRIF pathway. The IRF3 dependent activation may be more complex, however, as *Ifnb1* is an IRF3 dependent gene and thus *Irf3-/-* BMDM may phenocopy interferon deficient cells. This would similarly result in the late kinetics of activation we observed in the *Irf3-/-* affected sites.

A motif analysis performed on the 9 clusters of ATAC sites reveal strong enrichment of NF-κB in clusters 1, 3, 4, 5, and 7, while we observed strong κB PBM scores only in clusters 4 and 5. We additionally found AP1 enrichment in clusters 1, 2, and 7. When comparing the CHX-R and CHX-S sites, we observed ETS family member motifs in cluster 1 (CHX-R), and POU family motifs in cluster 7 (CHX-S) (Figure 5H).

Interleukin-10 Inhibits Formation of Accessible DNA Upstream of *Il12b*

The entirety of the previous analysis has been based on trying to understand the global properties, regulation, and function of inducible chromatin changes in BMDM during an acute stimulation. Our results have highlighted the extend to which RelA binding to strong motifs can contribute to inducible chromatin, as well as separate those sites which require additional protein synthesis for their activation. While important to document and understand the general principles governing inducible chromatin changes during the lipid A response through the use of chemical inhibitors and genetic deletions, we wanted to supplement this approach with the use of a biologically active cytokine with inhibitory properties. We chose to focus on Interleukin-10 as it has been shown to provide potent inhibition of target gene transcription, but little is known about precisely how specific genes are targeted for inhibition.

To perform this analysis we performed caRNA-Seq in *Il10-/-* BMDM treated with lipid A or lipid A & exogenous Il-10 for 0, 15, 30, 60, and 120 minutes. This method allows for evaluation of nascent transcription and prevents substantial bias in any effects of IL10 treatment on mRNA stability. After identifying significantly induced genes demonstrating over 10 fold induction (RPKM>1), we determined the effect of simultaneous addition of IL-10. The majority of genes were largely unaffected by IL-10 treatment (Figure 6A), but both Il10 super-induced and inhibited genes could be seen. When separating genes based on the time point of RNA-Seq collection, we could identify immediate IL-10 effects on the super-induction of lipid A induced genes, such as *Socs3* and *Niacr1*. As time of stimulation went on, however, an increasing number of genes displayed inhibition to IL10- treatment ($p<0.01$, Figure 6B). In order to gain

insights into the properties of genes that demonstrate sensitivity to IL10 treatment, we utilized a previous classification of potently induced genes in macrophages, where specific TFs and signaling pathways have been implicated in regulating target gene expression. We plotted maximum effect of IL10 treatment on genes within 95% of their maximum expression during any time point and could observe weak inhibition across many clusters of genes consistent with IL-10 not having a role in selectively suppressing a single TF family or signaling cascade. We did observe potent inhibition that was reserved for select genes previously known to be sensitive to IL-10 signaling. Importantly, *Il12b* stood out as demonstrating the strongest inhibition by IL-10, showing the max inhibition at the 120m time point.

While strict mechanisms of IL-10 mediated gene repression have remained elusive, the best evidence supports a model by which IL-10 utilizes STAT3 to either directly or indirectly repress transcriptional responses (Murray, 2005; Takeda et al., 1999). *Il12b* was previously identified to undergo nucleosome remodeling of the promoter and upstream 9Kb enhancer, and prior experiments were directed at determining whether IL-10 inhibited the remodeling step that precedes transcription. Restriction enzyme accessibility assays were unable to detect significant reduction in chromatin accessibility at the *Il12b* promoter or 9Kb enhancer with IL-10 treatment, although a severe reduction in RNA pol II was observed at the promoter (Zhou et al., 2004). Our ATAC analysis identified strong chromatin changes at the *Il12b* promoter as well as the 9Kb hypersensitive site (HSS) and an additional 17Kb, 27Kb, and 30Kb HSS that were previously uncharacterized. We therefore re-assessed IL-10's potential role in limiting DNA accessibility at putative enhancer sites upstream of inhibited genes, with an immediate interest in assaying regions upstream of *Il12b*.

To perform this analysis we performed ATAC-Seq in *Il10-/-* BMDM treated for 0m, or 120m with IL10, lipid A, or lipid A & IL-10 co-treatment. Because prior studies have demonstrated IL-10's ability to inhibit transcription of the *Il12b* gene with modest effects on promoter remodeling, we chose to initially investigate whether upstream-induced ATAC sites are sensitive to IL10 treatment. Visual inspection of the *Il12b* locus revealed slight repression of

accessible DNA formation at the *Il12b* promoter and 9Kb enhancer consistence with previous reports, as well as more pronounced inhibition of the 17kb, 27Kb, and 30Kb HSS indicating a potential role for these sites acting as enhancer elements in coordinating inducible transcription. Consistent with this, we observe inducible association of H3K27Ac across most sites, with the highest enrichment observed at the 27Kb enhancer (Figure 6F). Considering one aspect of active enhancers is their ability to recruit active polymerase, we asked whether IL10 treatment had an effect on the production of enhancer-associated RNA (eRNA). BMDM were stimulated for 0, 30, 60, 120, and 360m with or without IL-10 and eRNA production was analyzed by qPCR using primers flanking ATAC sites. We chose to focus on the 9Kb and 27Kb enhancer, as these peaks were among the strongest signals in the ATAC data and dmeosntrated K27Ac association. We could detect low levels of eRNA production using conventional qRT-PCR, with maximal eRNA production occurring at 60m, consistent with the time point we observe maximal *Il12b* transcription in the caRNA-Seq. IL-10 co-treatment resulted in reduced production of eRNA (Figure 6G), indicating the repressed ATAC sites correspond to decreased enhancer activity. Notably, these HSSs had diminished eRNA production in cells freshly isolated from in-vivo compartments following lipid A +/- IL10 pretreatment intraperitoneal injections, indicating they reflect regulatory sites utilized in cells differentiated in an endogenous environment and in response to acute stimulation (Figure 7D).

Selectivity of IL10's Effects on Chromatin and Analysis of ATAC Regions Upstream of Inducible Genes

The analysis above indicate that IL10 repressed ATAC sites may represent a mechanism by which IL10 represses gene transcription. To address how frequently we can observe IL10 repressing ATAC changes, we focused again on intergenic peaks induced >5-fold. To enrich for sites likely correlated with gene transcription we further focused on the peaks within 50Kb upstream of an induced gene. This resulted in 149 induced ATAC sites, which were further separated, based on their sensitivity to CHX, IRF3, and RelA motif score. Only 5 of the 149

peaks were sensitive (<30%) to IL10 treatment, and three of these sites were contained within the *Il12b* locus (Figure 7A). One site demonstrating IL10 mediated repression was found upstream of *Il27* and the other repressed peak was located between two inducible genes, *Rsad2* and *Cmpk2.* These 3 genes near an inhibited peak demonstrate modest and more variable effects by IL10 treatment during the first 2 hours, but our results leave open the possibility that repressed ATAC sites could further influence the sustained expression of these genes after the initial two hours of stimulation. In addition to sites that were repressed, we could identify ATAC sites that increased >3-fold with the addition of IL10. These sites were frequently found upstream of IL10 super-induced genes, including *Rnd1* and *Il1rn* (Figure 7B, 7C), which demonstrated strong transcriptional induction.

Discussion

Accessible chromatin is a requisite for many inducible TF binding events during the macrophage response, although many strong consensus sites have the potential to be embedded within closed regions unavailable for TF binding. Dynamic chromatin changes during stimulation likely reflect an additional mechanism of controlling inducible gene expression, and have the potential to provide lasting immunologic memory of previous environmental signals. Here, we observe rapid induction of accessible regions upstream of many inducible genes. While the majority of sites exhibit strong changes occurring within 30m of stimulation, most sites continue to increase in accessibility throughout the two-hour stimulation. The sites that demonstrate dynamic properties coincide with low CpG content at both promoters as has been previously reported, as well as at distal intergenic sites. This preference for induced sites containing low CpG content likely reflects the ability for stable nucleosome formation at low CpG content that can be subsequently modified by stimulus induced TFs and chromatin remodelers. Intergenic sites demonstrate a similar profile albeit at overall much lower CpG content.

The number of artifacts we could detect with visualization of the data that went unnoticed with bioinformatics analysis initially surprised us. A number of ATAC regions classified as

induced appear to have resulted from overall effects of nearby transcription, either throughout the gene body, or extending far past the end of annotated genes and into intergenic peaks. We were able to manually inspect all induced ATAC peaks overlapping promoters, but additional methods of ATAC analysis or incorporation of additional datasets will likely be needed to address some of these deficiencies genome-wide in the future. By focusing on the small numbers of promoter peaks with strong ATAC induction, we were able to determine that a majority of them correlate with strong inducible transcription for genes previously implicated in demonstrating nucleosome remodeling. Furthermore, when we directly assayed for the contribution of IRF3 to all promoters with strong ATAC induction, we observed only a single gene, *Ccl5*, that demonstrates strict dependence consistent with previous work which focused on potently induced genes (Tong et al., 2016). These results indicate that through the careful examination of potently induced genes using stringent criteria we were able to gain insights without missing additional information from genes with lower magnitudes of expression or induction.

We have previously shown that NF-κB binds most frequently to promoters with open chromatin, but our motif analysis of inducible sites revealed that only the strongest induced sites were enriched for strong consensus sites. This finding was in conflict with the observation from ChIP-Seq data where we observed a modest enrichment of RelA binding to strongly induced ATAC sites (data not shown). However, once we considered the strength of consensus sites we were able to observe an enrichment of strong motifs under peaks with strong induction. It is interesting to consider the possibility that the strong motifs occurring in strongly induced ATAC sites are functional targets of NF-κB in instances when stimuli provide the additional signal needed to remodel a positioned nucleosome, but stable nucleosome prevent NF-κB binding to these sites without the additional stimulus needed for nucleosome remodeling. A potential limitation of this work is that we are unable to determine the proportion of cells under study in which the inducible chromatin changes occur, and with what frequency TF binding occurs in the cells that exhibit induction. This knowledge would have allowed for a deeper understanding of

why we observe an enrichment of strong RelA:p50 consensus sites at peaks with strong induction, but cannot reliably detect strong RelA ChIP-Seq peaks. Should large chromatin changes only occur in a small fraction of the cells, overall readout of TF binding would be reduced compared to sites at which near 100% of the cells are available for binding. When we relaxed our criteria for calling RelA binding, we did indeed observe a greater enrichment of strong peaks occurring at strong motifs, indicating fractional occupancy may explain some of the enrichment at strong motifs with weak peaks, but some sites remained which had no detectable binding.

An unexpected finding of our studies conducted in the presence of CHX revealed that while both *Il12b* and *Il27* were dependent upon new protein synthesis for their expression, CHX only prevents induction of open chromatin at the *Il12b* promoter; the events leading to induction of the promoter at *Il27* appear to be in direct response to cell stimulation, yet productive transcription does not occur with CHX treatment. This indicates the mechanisms responsible for remodeling can be distinct from mechanisms of transcription, which has been previously shown for individual model genes (Weinmann et al., 2001).

While our main goal was to provide a comprehensive analysis of chromatin changes in context with inducible transcription, we wanted to provide insights into a biologically important inhibitor to complement our studies using chemical inhibitors and genetic deletions. The cytokine IL-10 was an obvious choice for both its important clinical potential as well as its poorly understood mechanism of repression. Our analysis of IL-10's ability to potently inhibit a limited number of genes provides a model from which we are now able to focus our attention towards individual genes in order to inform potential mechanisms. While we are unable to provide a mechanistic link between IL-10 signaling and the repression of induced ATAC sites upstream of inhibited genes, a potential explanation for our results is that IL-10 may either be inhibiting a process distal to the promoter which limits transcriptional activation, or could be directly targeting a form of regulation at the promoter and the ATAC changes upstream are merely consequence of less activity at the promoter.

EXPERIMENTAL PROCEDURES

Animal Studies

C57BL/6 and *IL10*$^{-/-}$mice (#002251) were purchased from Jackson Laboratories (Bar Harbor, ME) and *Irf3-/-* mice were a gift from Genhong Cheng. Experiments were performed under the written approval of the UCLA Animal Research Committee (ARC) in accordance to all federal, state, and local guidelines. In-vivo isolation of Peritoneal Exudate Cells (PEC) and Spleen was performed following 60m intraperitoneal injection of 25ug IL10 (or isovolemic PBS), followed by 1ug lipid A for 90m. PECs were harvested with PBS lavage of peritoneal cavity following euthanasia.

Cell Culture

BMDMs for ATAC-Seq were prepared from 6-10 week-old C57BL/6, *IL10*$^{-/-}$, or *Irf3*$^{-/-}$ male mice. Macrophages were scraped and re-seeded on day 4 of differentiation using 20ng/mL MCSF (Peprotech) and activated on day 6 with 100 ng/ml lipid A (Sigma). When indicated, cells were preincubated for 15 min with 10 mg/ml CHX (Sigma) or treated simultaneously with 10ng/mL IL-10 (eBioscience). BMDM for caRNA-Seq were prepared as above and differentiated with CMG conditioned media containing MCSF (Takeshita et al., 2000).

RNA-seq

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

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

locus. All RPKMs represent an average from two biological replicates. A gene was included in the analysis if it met all of the following criteria: The maximum RPKM reached 1 at any time point, the gene was induced at least 10-fold, and the induced expression was significantly different from the basal (P<0.01) as determined by the DESeq2 package in R Bioconductor (Robinson et al., 2010). P-values were adjusted using the Benjamini-Hochberg procedure of multiple hypothesis testing (Benjamini and Hochberg, 1995).

To determine the impact of IL10 on expression, the basal RPKM in lipid A treated samples was set at 0% and the maximum RPKM at 100%. The maximum RPKMs in the IL10 treated samples were converted to percent expression using this scale. Under circumstances were the gene reaches >95% of its max at multiple time points, the greatest effect of IL10 is shown.

ATAC-seq

ATAC-seq libraries were prepared using the Nextera Tn5 Transposase kit (Illumina) as described (Buenrostro et al., 2015) with slight modifications. Libraries were single-end sequenced (50bp) on an Illumina HiSeq 2000. Reads were mapped to the mouse genome (NCBI37/mm9) using Bowtie2. Reads were removed from the subsequent analysis if they were duplicated, mapped to mitochondrial genome, or aligned to unmapped contiguous sequences. All samples include at least 2 biological replicates, and replicates were merged prior to peak calling with MACS2. Overlapping peaks were merged together and used as probes for quantifying reads. The reads were converted to reads per million (RPKM) by dividing by the total number of reads within a peak divided by the peak length per million mapped reads.

Motif Analysis

The promoters of genes (-500 to +150 bp) were used for promoter analysis. The strongest p50:RelA binding site within each called peak was identified using a PBM dataset (Siggers et al., 2012). JASPAR2016 PWMs were used to identify the best matching motifs using

PscanChIP (Zambelli et al., 2013). If multiple reference motifs were found for a given TF the average score was used for analysis and duplicates were removed.

qRT-PCR

RNA was extracted using TRI-reagent (Molecular Research Center), and treated with DNaseI (Qiagen), and purified using an RNeasy kit (Qiagen). 2ug of RNA was reverse-transcribed using random hexamers and primers flanking the left and right of ATAC peaks were used to amplify products using SYBR-green detection. The following primers were used: *Il12b* 27Kb eRNA-R: F-GGAAGATCATTGCCCAGCTA, R- TTCTTGTCTCCCACCTTGCT, *Il12b* 9Kb eRNA-R: F-CTGTCTTCCCATGCTGTGTG, R- GCCAGTGTGTGTGAGCAGAT, *Il12b* 9Kb eRNA-L: F-CATTGCTCCAGACACTGAGG, R- TTTCTGTTCGCCAGAATGAA, *Il12b* 27Kb eRNA-L TGCAACAGCTCTACCAAACG, R- CCATTTGAGACATGCATTGG, *Il12b* mRNA: F-TGACATGTGGAATGGCGTCTCT, R- GGCGGGTCTGGTTTGATGAT, *Actb* mRNA: F-AGAGGGAAATCGTGCGTGAC, R-CAATAGTGATGACCTGGCCGT.

Figure Legends

Figure 4-1: Accessible DNA Dynamics During the Lipid A Response

(A) BMDMs were stimulated with Lipid A over a time course and ATAC-Seq was performed to analyze the properties of accessible regions. Left: The distribution of maximum fold accessibility values over the two hour stimulation period is shown for the 101,448 called peaks in WT BMDM. Middle: The distribution of maximum fold accessibility for all 11,456 significantly induced peaks ($p<0.05$ with multiple hypothesis testing, RPKM >1). Dashed gray lines indicate 2.5- and 5-fold cutoffs. Right: The distribution of minimum fold accessibility for the 4,292 significantly repressed peaks ($p<0.05$ with multiple hypothesis testing, Basal RPKM>4).

(B) All peaks (left), significantly induced (middle), and significantly repressed peaks (right) were annotated according to genomic location using HOMER (Heinz et al., 2010).

(C) Bedgraph visualization of select promoter peaks from repressed, constitutive, induced, and transcription-associated ATAC-peak categories. Time of stimulation is indicated at left of each panel.

(D) Peaks falling within each genomic location are separated based on their responsiveness to lipid A treatment: constitutive (blue), induced (red), and repressed (green). Y-axis scale set to 75% to allow comparisons within ATAC peak classes.

(E) CpG Obs/Exp ratios were calculated for peaks and plotted with respect to peak classification in shades of blue. The violin plots feature small diamonds indicating average CpG Obs/Exp ratios for each peak category. X-axis scale is a semi-log plot, with no CpG detection shown to the left of the dashed gray line.

Figure 4-2: Transcription Factor Motif Analysis of Repressed, Constitutive, and Inducible ATAC Peaks

(A) All ATAC peaks were grouped into 35 equivalently sized bins based on maximum fold induction and subsequent motif enrichments were identified using JASPAR2016 motif matrices in Pscan-ChIP (http://159.149.160.88/pscan_chip_dev). TF family annotations are indicated

above select TF motif groups and were obtained from AnimalTFDB (http://www.bioguo.org/AnimalTFDB). Each color indicates a scale of TF enrichment based on the –log(p-value).

(B) Genomic distributions of ATAC peaks are shown as a function of ATAC bins.

(C) Individual TF motif enrichments are plotted as a function of ATAC bins.

Figure 4-3: Properties of Inducible Promoter Sites

(A) Heatmap representation of the 39 induced peaks overlapping promoters. The CpG Observed/Expected ratio is shown with shades of orange indicting increasing levels of CpG density. Log2-Normalized RPKM values are presented in both the ATAC-Seq data as well as the corresponding chromatin-associated RNA-Seq data (GSE67357). The last three columns indicate the percent accessibility under conditions of CHX pretreatment, *Irf3-/-* BMDM (IRF3) and *Irf3-/-* BMDM pretreated with CHX (IRF3-CHX) compared to the 120m WT stimulation.

(B) Previously characterized primary response (PRG) and secondary response genes (SRG) were used for promoter ATAC analysis. The max ATAC fold (x-axis) is shown as a function of Max RNA-Seq fold (y-axis, left) or CpG Observed/Expected values of the ATAC peak overlapping the promoter (y-axis, right). Individual genes demonstrating large ATAC fold changes are indicated with text.

(C) Bedgraph visualization of SRGs previously characterized as dependent (-D) or Independent (-I) on interferon receptor signaling (IFNAR) for their activation. Within each classification, a selected gene is shown as either CHX resistant (CHX-R) or sensitive (CHX-S). The time course of sample collection is indicated at left, with +CHX indicating time points with CHX pretreatment.

(D) The 39 inducible promoters from the heatmap in (A) are shown in the scatter plot with the RPKM from WT (x-axis) or *Irf3-/-* (y-axis) under CHX & lipid A treatment for 120m. The color of each circle indicates the presence of an IRF3 ChIP-Seq peak (peak score<19 in light blue, peak >19 in dark blue). The absence of color indicates no called peak in the promoter.

(E) Bedgraph visualization of peaks centered on the *Ccl5* promoter.

Figure 4-4: Relationship of Inducible Intergenic ATAC Sites to Inducible Gene Expression

(A) Average properties and motif enrichments for all intergenic ATAC peaks grouped into 20 equivalently sized bins (2078 peaks/bin) according to their fold induction values.

(B) Tables (left) are shown for counting the occurrence of >5 fold induced ATAC sites at regions upstream of expressed genes (>1 RPKM) that display <2 fold induction, 2-10 induction, or >10 fold induction. Percentages based on the size of each gene group are presented in the table below the raw counts. Enrichment values of ATAC peak occurrences for weakly induced and strongly educed genes are shown relative to the occurrence in <2 fold induced genes as separate by distance bins on the x-axis.

(C) Tables (left) are shown for the number of genes with multiple induced ATAC peaks at distances upstream of the TSS. Percentages are calculated by counting the number of genes with multiple peaks in the distance bin divided by the total number of genes with at least one induced peak in the same distance group. Enrichment values of multiple ATAC peak occurrences for weakly induced and strongly educed genes are shown relative to the occurrence in <2 fold induced genes as separate by distance bins on the x-axis.

(D) Scatter plot showing the relationship between CpG content (x-axis) and ATAC max fold induction (y-axis) of the 2404 >5 fold induced ATAC peaks.

(E) (Left) The percent of genes with an ATAC peak belonging to either 0.9-1.1 fold, 2-3 fold, or >5 fold induction within 50Kb of the TSS is shown. Genes are separated based on fold induction with <2 fold, 2-10 fold, and >10 fold genes shown with increasing intensities of blue. (Right) The percent of genes with an additional gene TSS located within 50kb upstream.

(F) Scatter plot display of RNA-Seq fold induction (y-axis) compared to the numbers of ATAC peaks located within 50kb upstream (x-axis) of all >10 fold induced genes. The scatter plots are separated based on the type of peak detected, >5 fold induced sites are colored in red and

placed in the top panel. 2-3 fold induced ATAC peaks are colored orange and located in the middle panel. 0.9-1.1 fold ATAC peaks are colored black and located at bottom.

(G) Bedgraph visualization of genes with multiple induced ATAC peaks upstream of their TSS.

Figure 4-5: Analysis of NF-κB Binding to Induced Intergenic Peaks

(A) PBM Z scores of p50:RelA (y-axis) and RelA ChIP-seq peak scores (x-axis) in the 0.9-1.1 fold induced (left) and >5 fold induced (right) ATAC peaks were plotted. The horizontal dashed line indicates the PBM Z score threshold (6.4), and the vertical dashed line indicates the ChIP-seq peak score threshold (19).

(B) Tables are shown for counting the occurrence of RelA binding to intergenic sites upstream of genes from different classes of expression. RelA ChiP-Seq signal is separated into three groups, no binding, peak score <19, and peak score >19. PBM motif scores of the best motif match under the RelA peak is shown as >6.4 and <6.4. The frequency of RelA binding to regions upstream of the gene TSS is indicated in distance bins. Percentages based on the size of each gene group are presented in the table below the raw counts.

(C) A line graph is shown indicating the p50:RelA motif Z score enrichment in the >5 fold induced ATAC peaks relative to the called binding events in the 0.9-1.1 fold induced ATAC peaks.

(D) Kinetics of ATAC and RelA Peak strength at sites with strong binding and strong motifs, separated based on peaks which show coincident kinetics (left), early RelA association (middle) or late RelA associating (right).

(E) Heatmap representation of the 2404 >5 fold induced intergenic peaks. RPKMs from ATAC-Seq are converted to percent of maximum accessibility and plotted with shades of purple in a lipid A treated time course adjacent to a time course with a CHX pretreatment. The fold change per time point is shown in shades of red and blue (Δ/ Time Point). The CHX sensitivity is shown as % of max accessibility with shades of orange, RelA PBM >6.4 motifs are shown in. The last

three columns indicate the percent accessibility compared to 120m WT under conditions of CHX pretreatment (CHX), *Irf3-/-* BMDM (IRF3), and in *Irf3-/-* BMDM pretreated with CHX (IRF3-CHX).

(F) ATAC peaks were separated based on their RelA motif score (x-axis) and the percent of peaks are shown corresponding to their sensitivity to CHX (as indicated with black, yellow, orange).

(G) Average ATAC RPKM are shown for clusters with strong RelA motifs (clusters 4 & 5) alongside clusters with evidence of IRF3 dependence (clusters 2 & 6).

(H) Heatmap visualization of motif enrichments across clusters of induced intergenic ATAC sites with select TF families shown.

Figure 4-6: Investigation of IL-10 Mediated Gene Repression and Inhibition of ATAC Peaks

(A) Fold induction values in IL-10-/- cells treated with lipid A (green) as compared to IL-10-/- cells treated with lipid A & IL10 (blue).

(B) Gene expression plots separated by time point are shown, and RPKMs with or without IL-10 are shown on the x- and y-axis respectively. Selected genes are highlighted in text, with red circles indicating IL-10 responsive genes with $p<0.01$.

(C) Genes are separated based on previous classifications of PRG and SRG. The *Il12b* gene is highlighted with text. Red dashed lines indicate 30% and 300% expression with IL-10 co-treatment compared to lipid A alone. The maximum IL10 inhibition is shown for genes that have reached >95% of their maximum expression.

(D) Bedgraph visualization of ATAC sites in the *Il12b* locus in IL10-/- BMDM treated with IL10, Lipid A, or co-treatment of both Lipid A and IL-10 for 120m. Peaks with IL-10 inhibition are highlighted with red dashed rectangles.

(E) Left: Fold Accessibility of ATAC peaks upstream of *Il12b* in WT or *Il10-/-* cells with lipid A or lipid A & IL10. Right: RPKM of H3K27AC ChIP-Seq at ATAC sites in either basal or 120m stimulated BMDM.

(F) Detection of enhancer-transcribed RNA using primers flanking ATAC peaks and measured with qRT-PCR. BMDM were stimulated for 0, 30, 60, 120, and 360 minutes with either lipid A (black) or lipid A & IL10 (red), error bars indicated standard error of the mean (S.E.M.). Two replicates of chromatin-associated RNA-Seq of nascent *Il12b* transcripts are shown at right.

Figure 4-7: Selectivity in IL10-mediated ATAC changes at Sites Upstream of Inhibited Genes

(A) Properties of the 149 peaks with >5 fold induction and are within 50kb upstream of a >10 fold induced gene. IL-10 effect is shown at far right column in shades of green according to scale at bottom.

(B) Bedgraph visualization of ATAC data in *Il10-/-* BMDM at 0m or 120m stimulation as shown at left of panels. Two genes are highlighted showing >300% accessibility with the addition of IL10 (*Rnd1*, *Il1rn*) and three genes with an ATAC peak <30% accessible with the addition of IL10 (*Cmpk2*, *Rsad2*, *Il27*).

(C) Selected caRNA-Seq data with two replicates shown.

(D) In-vivo stimulated expression data of *Il12b* inhibition by IL10 as well as reduction of enhancer transcribed RNA at the 9Kb and 27Kb enhancers upstream of *Il12b* in both peritoneal exudate cells (PEC) as well as whole splenic tissue (mouse n=3 PBS, 5 lipid A, 5 lipid A & IL10). Error bars indicate standard error of the mean, *P<0.05, **P<0.01 using unpaired two-tailed Student's t-test.

Figure 4-1: Accessible DNA Dynamics During the Lipid A Response

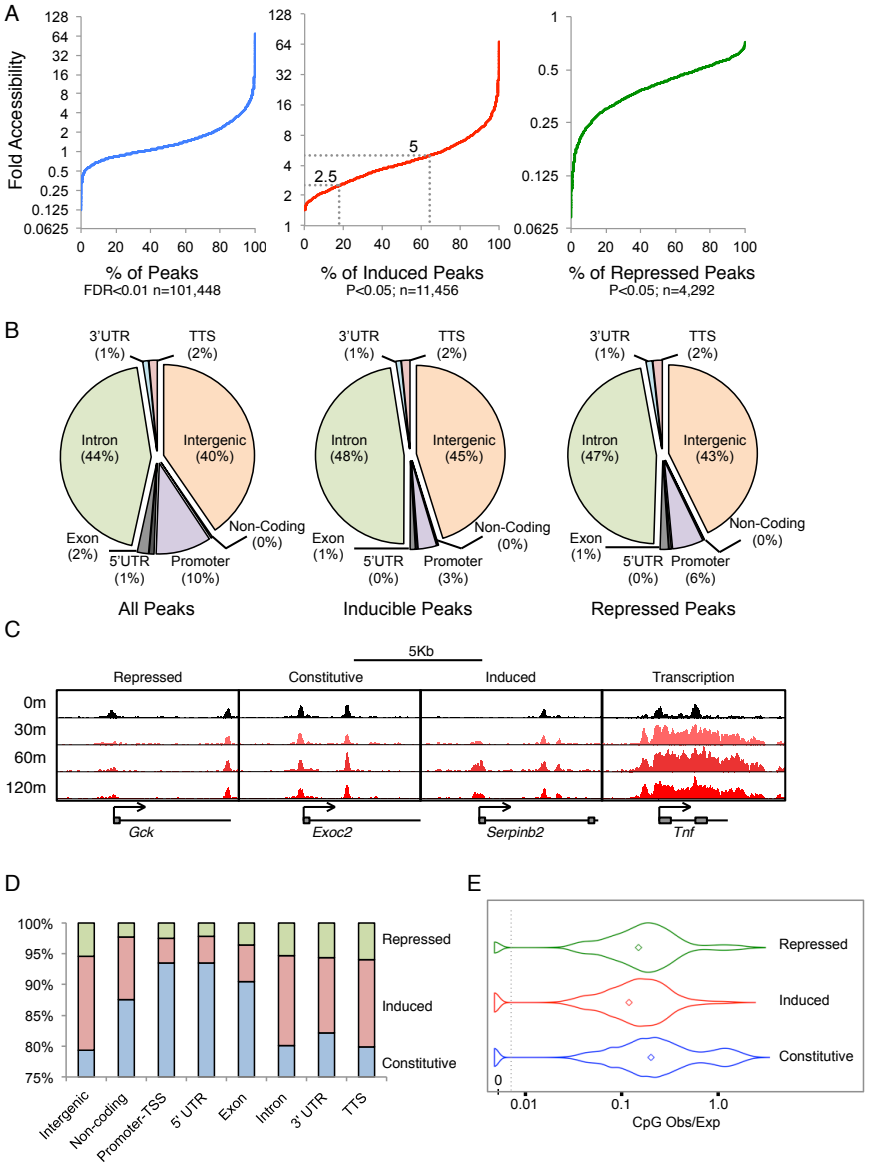


Figure 4-2: Transcription Factor Motif Analysis of Repressed, Constitutive, and Inducible ATAC Peaks

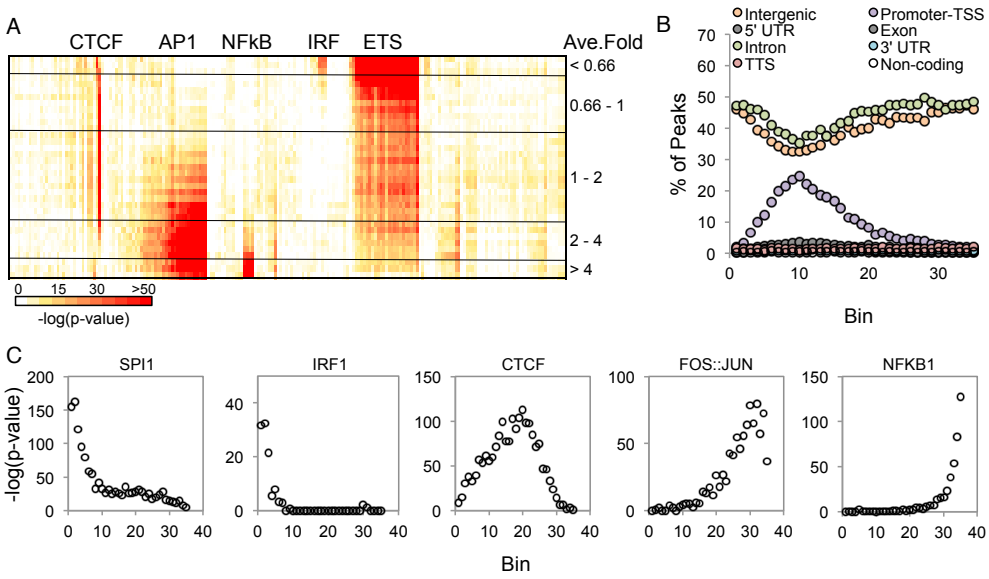


Figure 4-3: Properties of Inducible Promoter Sites

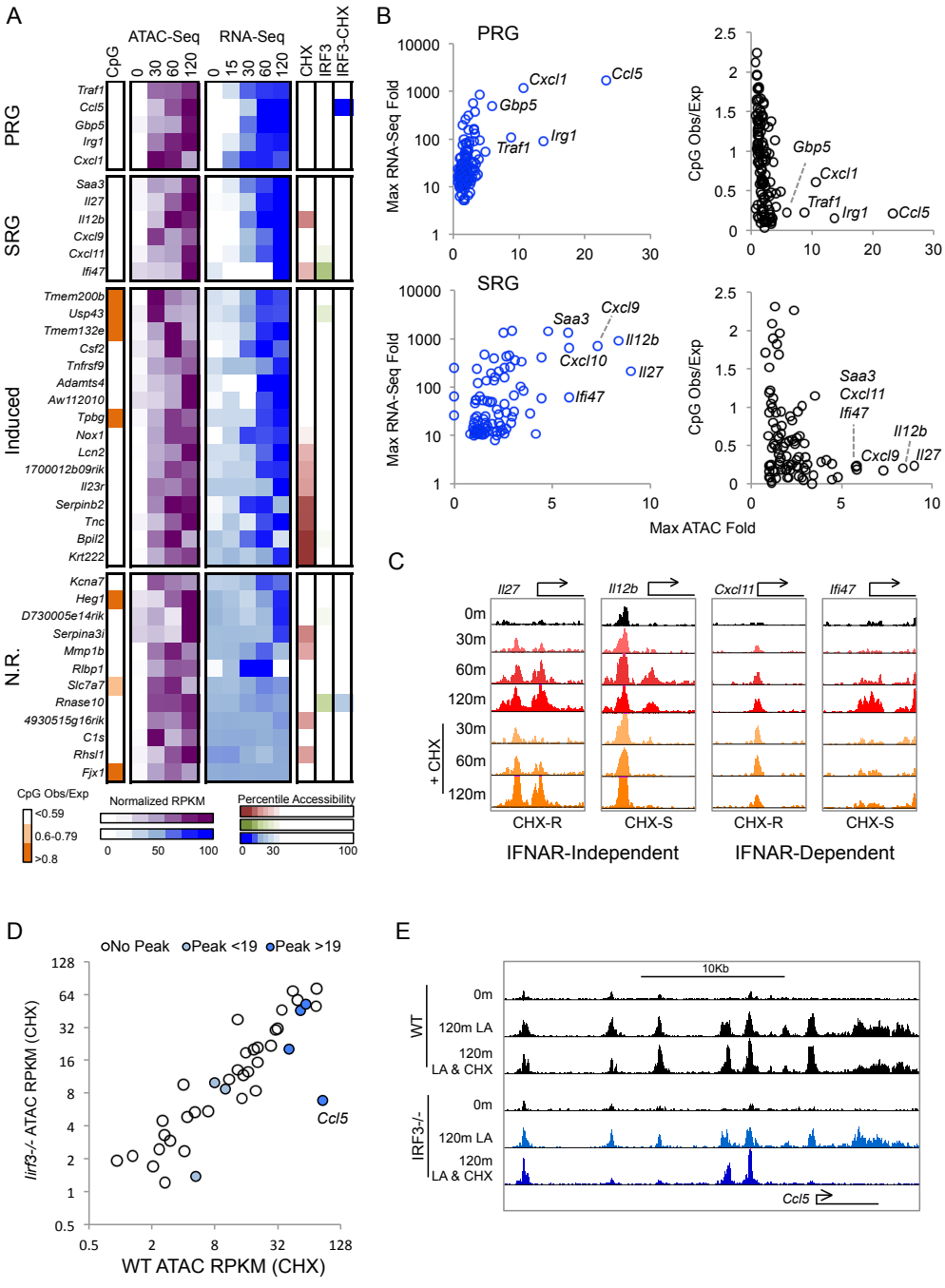


Figure 4-4: Relationship of Inducible Intergenic ATAC Sites to Inducible Gene Expression

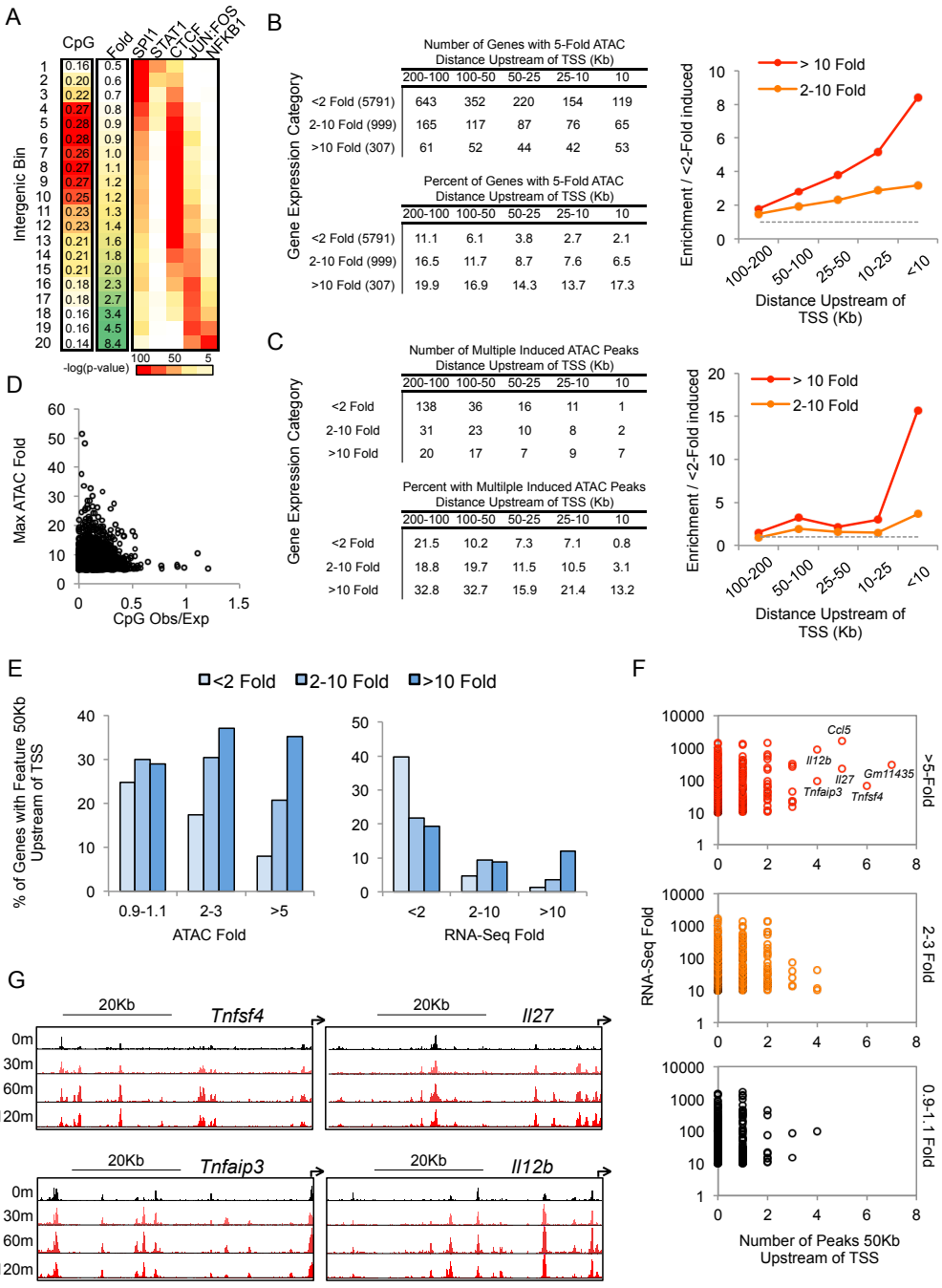


Figure 4-5: Analysis of NF-κB Binding to Induced Intergenic Peaks

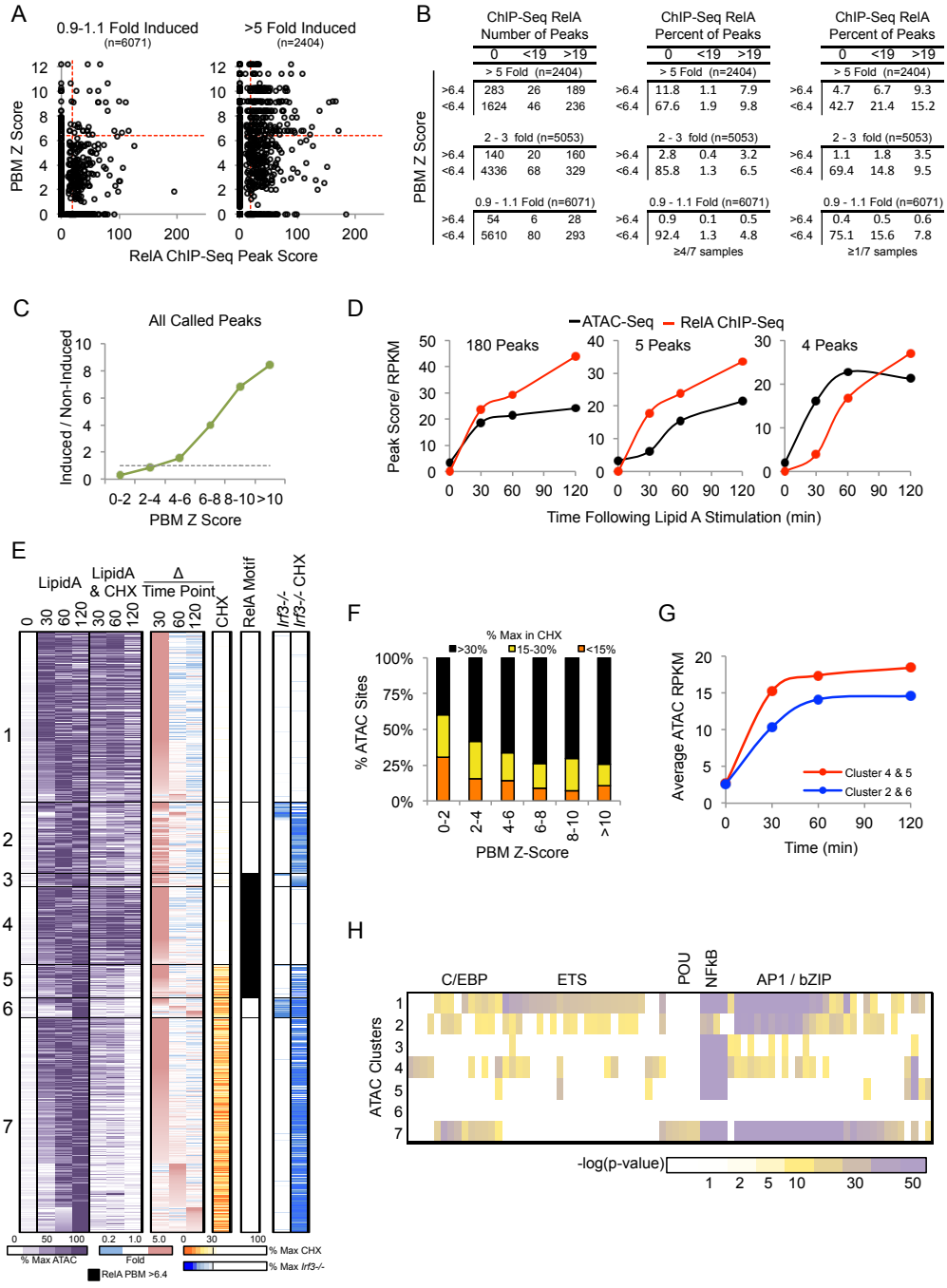


Figure 4-6: Investigation of IL-10 Mediated Gene Repression and Inhibition of ATAC Peaks

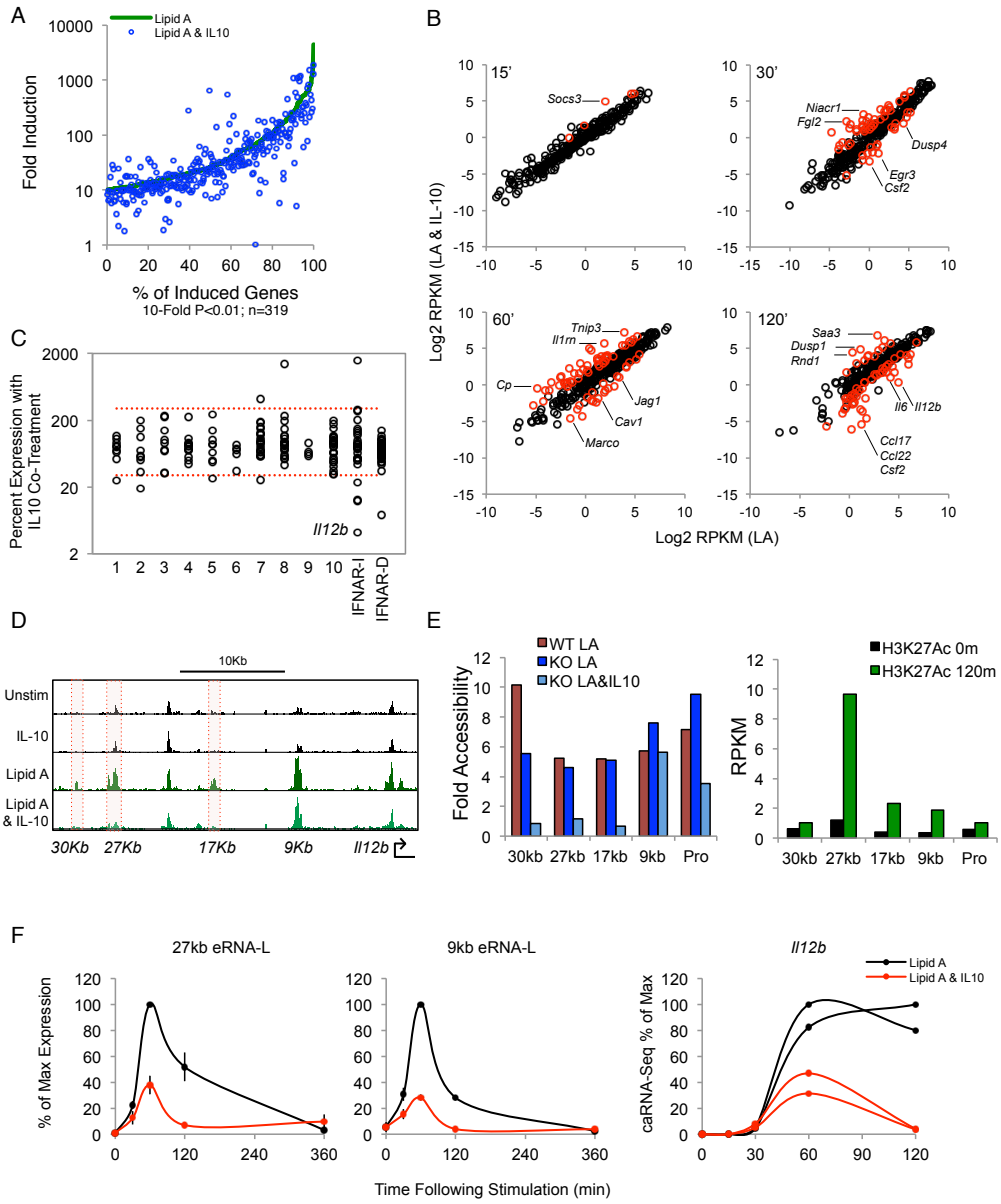
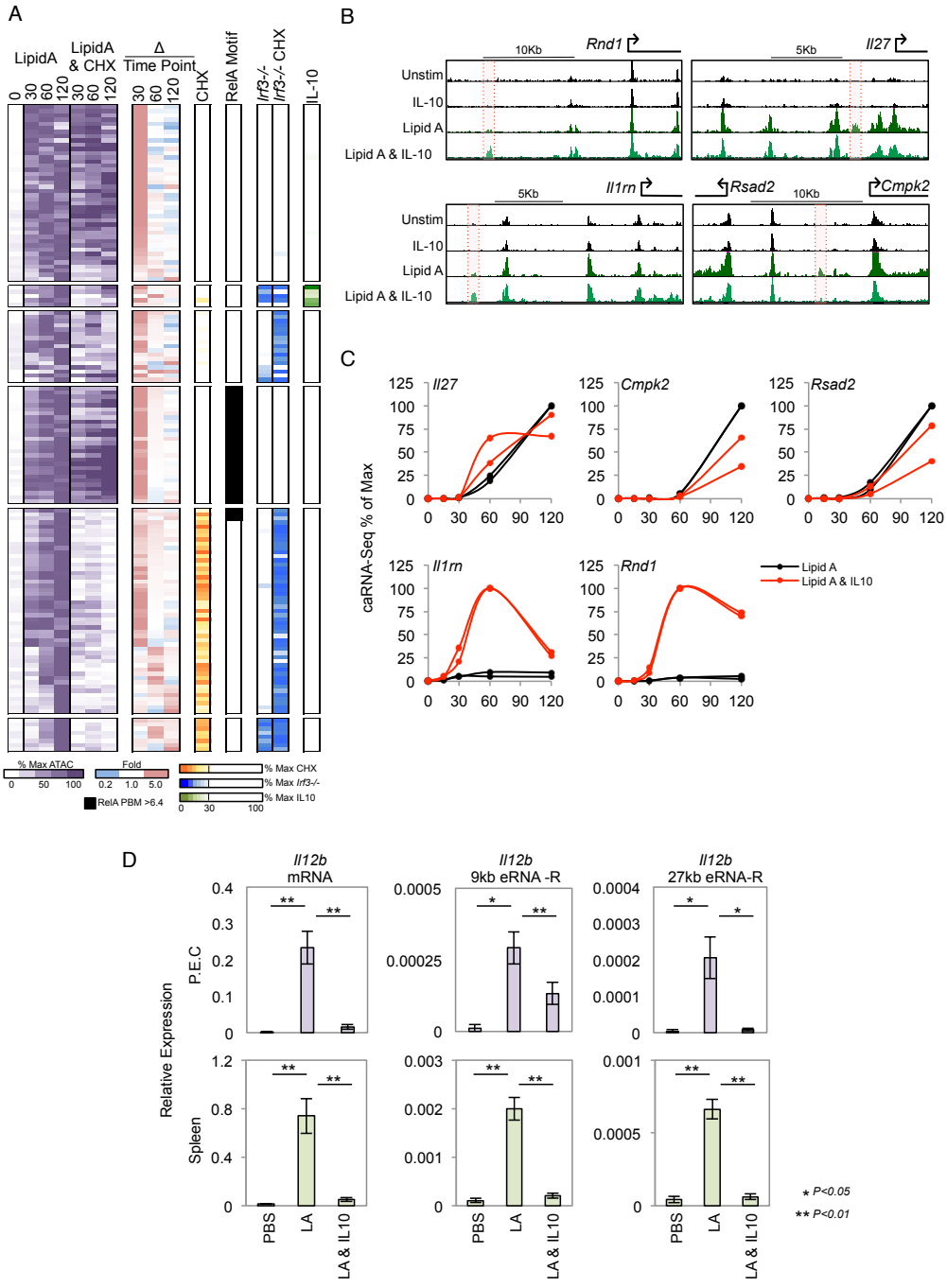


Figure 4-7: Selectivity in IL10-mediated ATAC changes at Sites Upstream of Inhibited Genes



References

Angelov, D., Lenouvel, F., Hans, F., Müller, C.W., Bouvet, P., Bednar, J., Moudrianakis, E.N., Cadet, J., and Dimitrov, S. (2004). The histone octamer is invisible when NF-??B binds to the nucleosome. J. Biol. Chem. *279*, 42374–42382.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. *57*, 289–300.

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. In Current Protocols in Molecular Biology, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p. 21.29.1-21.29.9.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Chen, F.E., Huang, D.B., Chen, Y.Q., and Ghosh, G. (1998). Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. Nature *391*, 410–413.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Elliott, H., Chen, Y., Bernat, J.A., Ginsburg, D., et al. (2005). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS) Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). 123–131.

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

Groudine, M., and Weintraub, H. (1982). Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. Cell *30*, 131–139.

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

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

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

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.

Murray, P.J. (2005). The primary mechanism of the IL-10-regulated antiinflammatory response is to selectively inhibit transcription. Proc. Natl. Acad. Sci. U. S. A. *102*, 8686–8691.

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.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Saccani, S., Pantano, S., and Natoli, G. (2001). Two waves of nuclear factor kappaB recruitment to target promoters. J. Exp. Med. *193*, 1351–1359.

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

Takeda, K., Clausen, B.E., Kaisho, T., Tsujimura, T., Terada, N., Förster, I., and Akira, S. (1999). Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of stat3 in macrophages and neutrophils. Immunity *10*, 39–49.

Takeshita, S., Kaji, K., and Kudo, a (2000). Identification and characterization of the new osteoclast progenitor with macrophage phenotypes being able to differentiate into mature osteoclasts. J. Bone Miner. Res. *15*, 1477–1488.

Tong, A.-J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation. Cell 1–15.

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

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

Weintraub, H., and Groudine, M. (1976). Chromosomal Subunits in Active Genes Have an Altered Conformation. Science (80-.). *193*, 848–856.

Zambelli, F., Pesole, G., and Pavesi, G. (2013). PscanChIP: Finding over-represented transcription factor-binding site motifs and their correlations in sequences from ChIP-Seq experiments. Nucleic Acids Res. *41*, 535–543.

Zhou, L., Nazarian, A.A., and Smale, S.T. (2004). Interleukin-10 Inhibits Interleukin-12 p40 Gene Transcription by Targeting a Late Event in the Activation Pathway. Mol. Cell. Biol. *24*, 2385–2396.

APPENDIX A

A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation

Article

Cell

A Stringent Systems Approach Uncovers Gene-Specific Mechanisms Regulating Inflammation

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

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

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

[3]Co-first author

*Correspondence: smale@mednet.ucla.edu

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

SUMMARY

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

INTRODUCTION

The molecular biology revolution of the 1970s was followed by a 20-year period during which gene regulation was studied at the level of individual model genes. Near the turn of the century, the emergence of DNA microarrays and whole-genome sequences opened avenues toward the study of gene regulation at a global scale, making it possible to identify genes and networks that characterize a cell type, environmental response, or disease state. More recently, RNA sequencing (RNA-seq) has emerged as a method that allows global transcript levels to be evaluated with greater accuracy (Marioni et al., 2008). RNA-seq also provides an opportunity to monitor nascent transcripts in addition to mRNA (e.g., Bhatt et al., 2012; Core et al., 2008; Rabani et al., 2011). For studies of stimulus-induced transcription, nascent transcript levels provide more accurate information about the kinetics with which gene transcription is activated, and they allow transcription to be studied independently of mRNA stability.

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

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

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

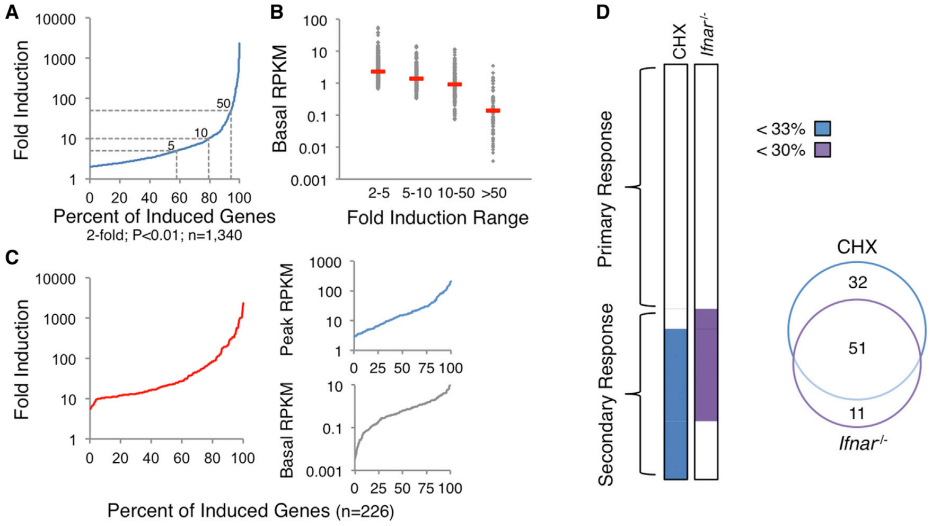


Figure 1. Properties of the Lipid A-Induced Transcriptional Cascade

Chromatin-associated transcripts from BMDMs stimulated with lipid A were analyzed by RNA-seq.

(A) The distribution of maximum fold induction values over the 2 hr stimulation period is shown for the 1,340 significantly induced (2-fold, $p < 0.01$) and expressed (three RPKM) genes. With multiple hypothesis testing, two weakly induced genes exhibited q values >0.01. The dashed gray lines represent 5-, 10-, and 50-fold induction thresholds.

(B) The 1,340 induced genes were grouped into bins, with basal RPKMs shown for each bin and red dashes indicating median RPKMs.

(C) The distributions of maximum fold inductions (left), peak RPKMs (top right), and basal RPKMs (bottom right) are shown for the 226 genes selected for analysis.

(D) The 226 genes were separated into PRG and SRG groups on the basis of their expression in CHX-treated and *Ifnar*$^{-/-}$ BMDMs. Genes were classified as SRGs if they were expressed <33% in CHX or <30% in *Ifnar*$^{-/-}$ samples. The Venn diagram indicates the number of genes affected by CHX treatment, the absence of IFNAR, or both.

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

RESULTS

Basic Properties of the Transcriptional Cascade

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

Of the 3,863 expressed genes, 1,340 (34.7%) were induced by at least 2-fold ($p < 0.01$) (Figure 1A). Importantly, however, 79.5% of these genes were induced less than 10-fold (Figure 1A). If all genes induced by 2-fold or greater were evaluated together, the analysis would be dominated by weakly induced genes. Notably, most induced genes encoding key cytokines, chemokines, and transcription factors were induced by >10-fold (data not shown). We therefore focused on the potently induced genes, with the resulting insights then examined in the context of the weakly induced genes (see below). Notably, the basal transcript levels of the weakly induced genes were generally higher than those of the strongly induced genes (Figure 1B).

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

Separation of Primary and Secondary Response Genes

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

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

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

Separation of IFNAR-Dependent and -Independent SRGs

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

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

Interestingly, although these 29 SRGs were strongly induced in the absence of IFNAR signaling, a subset, including the critical T cell polarizing cytokines *Il12b*, *Il6*, *Lif*, and *Il27* (Metcalfe, 2011; Shih et al., 2014), were induced much less potently by PAM than by lipid A (Figure 2D, bottom). In fact, *Il12b*, *Il6*, *Lif*, and *Il27* exhibited greater differential induction by TLR4 versus TLR2 ligands than any other PRG or SRG (Figure 2E). This finding suggests that the TRIF pathway activated by lipid A but not by PAM may be important for the activation of these genes, but not due to its role in activating IFNAR signaling. Consistent with this possibility, a direct comparison of WT to *Trif*$^{-/-}$ BMDMs revealed strong TRIF dependence of these genes (Figure 2D, bottom). Together, the data suggest that lipid A induces the expression of key T cell polarizing cytokines (*Il12b*, *Il6*, *Lif*, and *Il27*) much more potently than does PAM because the TRIF pathway strongly promotes the expression of these genes in an IFNAR-independent manner.

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

Initial Analysis of PRGs

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

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

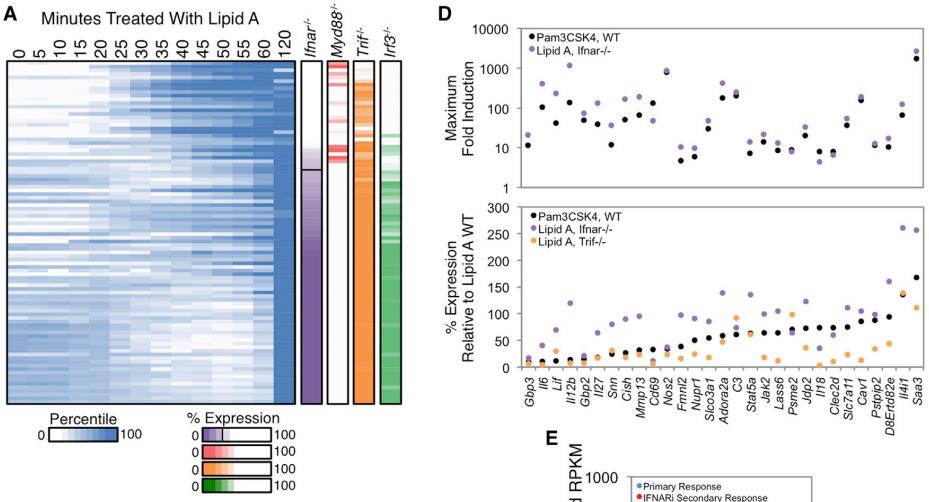


Figure 2. Analysis of IFNAR-Independent and -Dependent SRGs

(A) Activation kinetics are shown for SRGs from BMDMs stimulated at 5-min intervals from 0–60 min, with an additional 120-min time point. Shades of blue indicate percentile values. Genes were sorted on their maximum percent expression in *Ifnar*$^{-/-}$ BMDMs relative to WT BMDMs (purple column). The maximum percent expressions in *Myd88*$^{-/-}$, *Trif*$^{-/-}$, *and Irf3*$^{-/-}$ BMDMs are shown to the right. See also Figure S1.

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

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

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

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

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

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

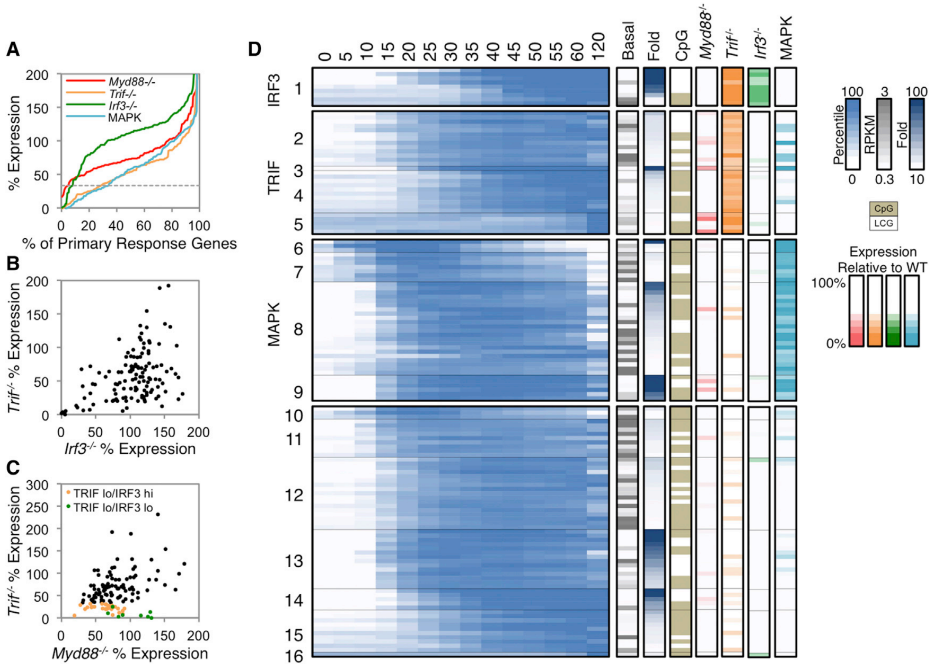


Figure 3. Properties of PRGs

(A) The distribution of the maximum percent expressions in *Myd88*$^{-/-}$ (red), *Trif*$^{-/-}$ (orange), *Irf3*$^{-/-}$ (green), and MAP kinase inhibitor-treated (light blue) BMDMs stimulated with lipid A are shown for the 132 PRGs. The horizontal dashed gray line indicates the 33% expression threshold.

(B and C) The percent expression of each PRG is shown in *Trif*$^{-/-}$ versus *Irf3*$^{-/-}$ cells (B) or in *Trif*$^{-/-}$ versus *Myd88*$^{-/-}$ cells (C). TRIF lo (<33% relative to WT) IRF3 hi (>33% relative to WT) genes are in orange, and TRIF lo (<33% relative to WT) IRF3 lo (<33% relative to WT) genes are in green.

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

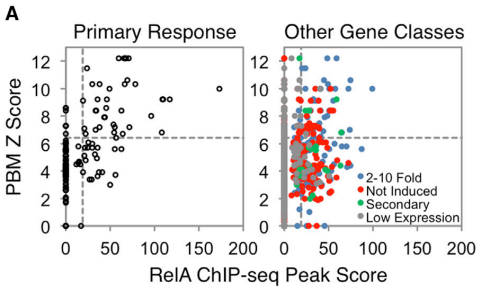
See also Figures S2 and S3.

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

In addition to the degree of dependence of each PRG on MyD88, TRIF, IRF3, and MAPKs, Figure 3D indicates basal transcript and fold-induction values. Furthermore, Figure 3D indicates which genes contain CpG-island or low CpG (LCG) promoters. As shown previously (Bhatt et al., 2012), all early transiently induced genes (e.g., clusters 6 and 10) contain CpG-island promoters and a high percentage of the most potently induced genes contain LCG promoters (e.g., clusters 1 and 14), whereas the two promoter types are distributed fairly randomly among the other clusters.

Initial Transcription Factor Binding Motif and ChIP-Seq Analyses

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



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	**NF-κB inducing 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	**NF-κB signaling**
	Cd40	7.2	-83	55.2	**NF-κB inducing receptor**
TGGAAATTCC	*Icam1*	7.2	-174	38.0	Transmembrane receptor
AGGGGATTTT	*Traf1*	7.1	-106	20.9	**NF-κB signaling**
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

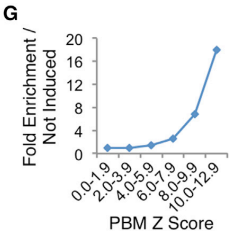


Figure 4. NF-κB Interactions at the Promoters of Defined Gene Classes

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

(legend continued on next page)

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

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

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

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

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

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

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

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

The substantial but lesser enrichment of promoters with strong NF-κB peaks (score >19) and weak binding motifs (*Z* score <6.4) among the strongly induced PRGs is also of interest. In these promoters, NF-κB may bind directly to weak motifs. Alternatively, NF-κB may be recruited 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. Although the significance of these interactions remains unknown, our ability to classify these promoters and distinguish them from the more

(B) Tables are shown indicating the distribution of genes from (A) for both numbers (left) and percentages (right) of genes.
(C–F) Tables are shown indicating the best matching κB motif in each promoter (column 1), the gene name (column 2), the PBM p50:RelA *Z* score (column 3), the position of the motif relative to the TSS (column 4), the RelA ChIP-seq peak score (column 5), and either the function or fold induction (column 6). This information is included for the PRGs with (C) strong κB motifs and strong RelA binding, (D) strong κB motifs that do not support RelA binding, (E) weak κB motifs and strong RelA binding, and (F) other NF-κB and IκB family members.
(G) A line graph is shown indicating the p50:RelA motif *Z* score enrichment in the promoters of the PRGs relative to the promoters of uninduced genes.
See also Figure S5.

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

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

Examination of NF-κB-Regulated Genes

A major goal of this study was to elucidate the logic through which the lipid A-induced transcriptional cascade is regulated. The identities of the 37 strongly induced PRGs that combine strong ChIP-seq peaks and strong motifs provide compelling evidence of an underlying logic; specifically, more than a third (13 of 37; Figure 4C) encode NF-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 signaling (*Tnfaip3*, *Tnip3*, and *Traf1*). Strikingly, these 13 genes include all of the NF-κB/IκB family members and direct regulators of NF-κB signaling found among the 132 PRGs. Notably, the promoters of genes encoding the two NF-κB family members and one IκB family member missing from this list also combine a strong RelA ChIP-seq peak with a strong NF-κB motif (Figure 4F); these genes were not among the 132 PRGs because they were only weakly induced.

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

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

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

Kinetic and Functional Analysis of Putative NF-κB Targets

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

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

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

Gene-Specific Regulation of IRF3-Dependent Genes

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

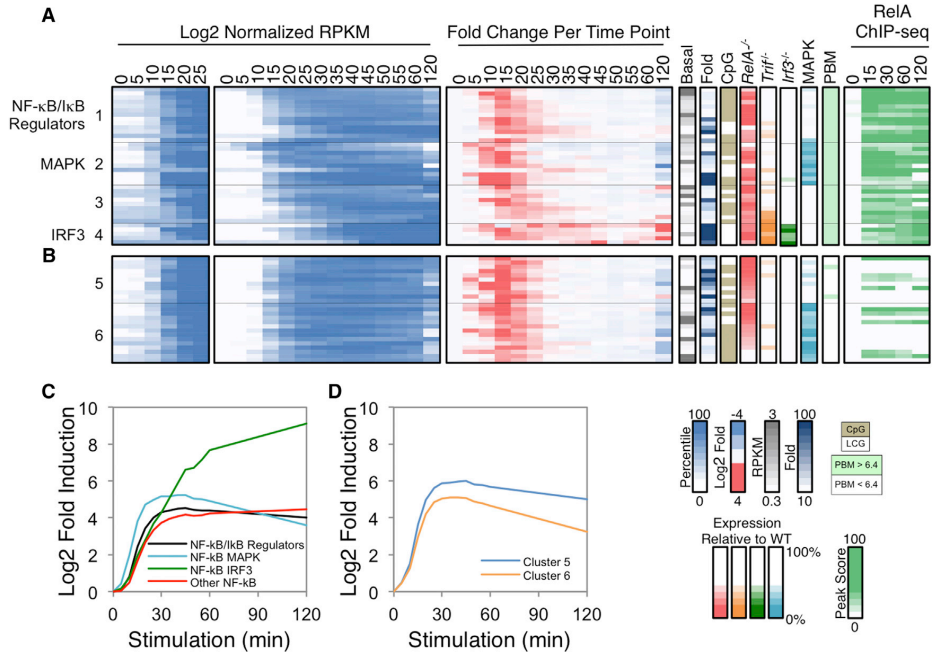


Figure 5. Kinetic and Functional Analysis of Putative NF-κB Target Genes

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

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

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

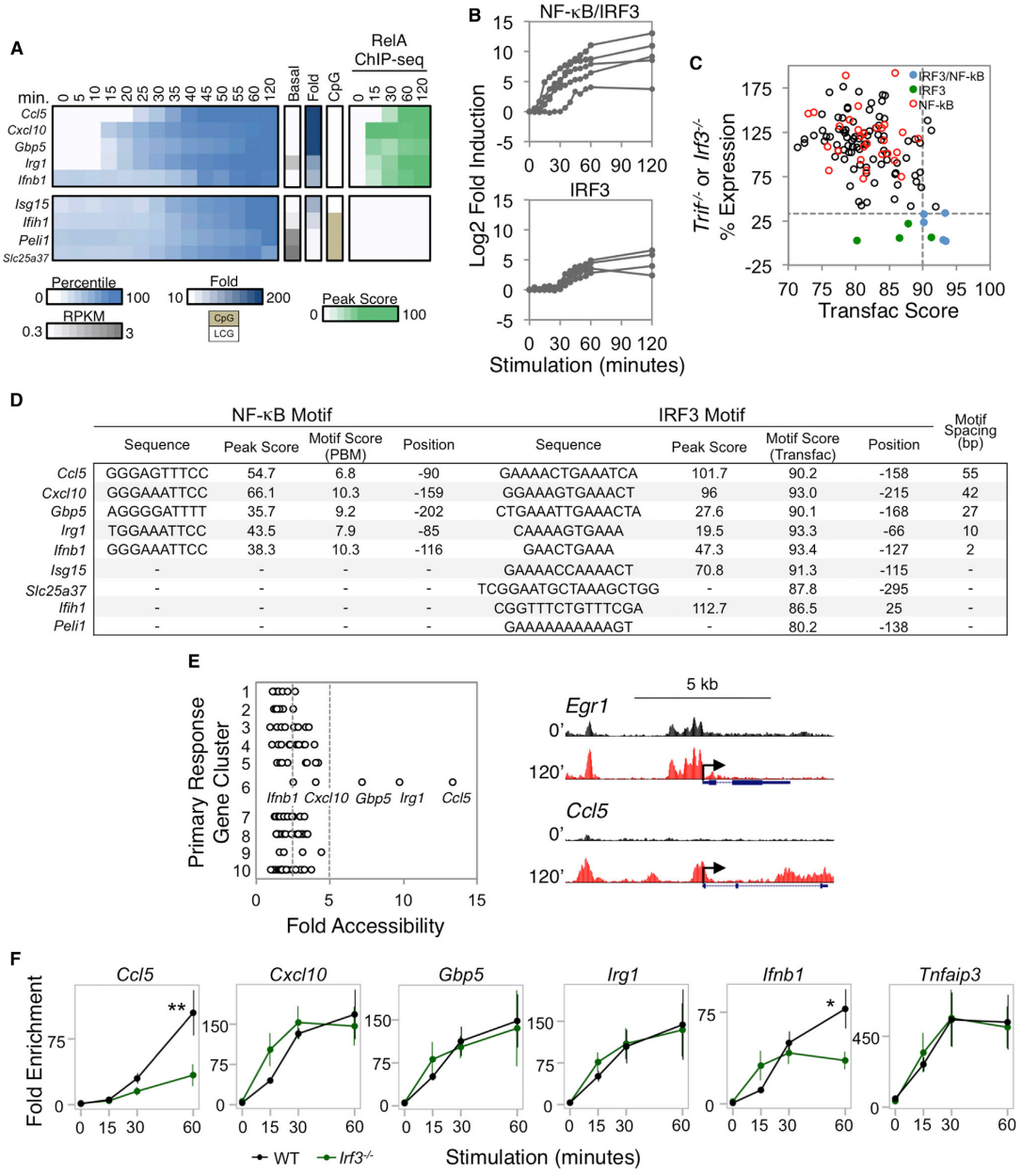
(D) The average activation kinetics of the two additional clusters from Figure 5B (clusters 5 and 6) are shown.

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

An examination of the five genes exhibiting both NF-κB promoter binding and IRF3 dependence reveals the extent to which genes have evolved unique regulatory strategies. Within this group, the expression kinetics of *Ccl5* and *Ifnb1* are each unique, whereas *Cxcl10*, *Gbp5*, and *Irg1* are similar (Figure 6A). These latter three genes were initially induced 10–15 min post-stimulation along with most NF-κB-dependent genes. Consistent with the hypothesis that NF-κB contributes to this early induction, RelA ChIP-seq peaks were observed at these genes by 15 min post-stimulation (Figure 6A, right), and their induction at early time points was unaltered in *Irf3*$^{-/-}$ macrophages (data not shown). IRF3 dependence was observed only at later times, consistent with prior knowledge that IRF3 activation is relatively slow (Kagan et al., 2008).

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

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



	NF-κB Motif				IRF3 Motif				Motif Spacing (bp)
	Sequence	Peak Score	Motif Score (PBM)	Position	Sequence	Peak Score	Motif Score (Transfac)	Position	
Ccl5	GGGAGTTTCC	54.7	6.8	-90	GAAAACTGAAATCA	101.7	90.2	-158	55
Cxcl10	GGGAAATTCC	66.1	10.3	-159	GGAAAGTGAAACT	96	93.0	-215	42
Gbp5	AGGGGATTTT	35.7	9.2	-202	CTGAAATTGAAACTA	27.6	90.1	-168	27
Irg1	TGGAAATTCC	43.5	7.9	-85	CAAAAGTGAAA	19.5	93.3	-66	10
Ifnb1	GGGAAATTCC	38.3	10.3	-116	GAACTGAAA	47.3	93.4	-127	2
Isg15	-	-	-	-	GAAAACCAAAACT	70.8	91.3	-115	-
Slc25a37	-	-	-	-	TCGGAATGCTAAAGCTGG	-	87.8	-295	-
Ifih1	-	-	-	-	CGGTTTCTGTTTCGA	112.7	86.5	25	-
Peli1	-	-	-	-	GAAAAAAAAAAGT	-	80.2	-138	-

Figure 6. Analysis of IRF3 Target Genes

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

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

(legend continued on next page)

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

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

The above results support a hypothesis in which multiple distinct mechanisms regulate the nine IRF3-dependent genes. To examine this hypothesis further, three additional experiments were performed. First, IRF3 ChIP-seq experiments revealed that strong IRF3 peaks (>19) coincide with strong IRF motifs (>90) at the promoters of only six of the 132 primary response genes, including the five NF-κB/IRF3 genes described above and the IRF3-dependent *Isg15* gene that lacks NF-κB binding (see peak scores in Figure 6D; a detailed analysis of the IRF3 ChIP-seq data will be presented elsewhere). An IRF3 ChIP-seq peak was also observed in the promoter of one of the IRF3-dependent genes that lacks a strong IRF motif (*Ifih1*; Figure 6D).

Second, ATAC-seq experiments revealed weak increases in chromatin accessibility upon lipid A stimulation at PRGs in many different classes (Figure 6E). However, the largest increase was observed at the *Ccl5* promoter, with large increases also observed at the *Irg1* and *Gbp5* promoters (Figure 6E). The large increase at the *Ccl5* promoter is consistent with the hypothesis that a nucleosome remodeling requirement is responsible for the delayed binding of RelA. Furthermore, the absence of an inducible ATAC-seq signal at the *Ifnb1* promoter is consistent with prior evidence that the promoter is nucleosome-free prior to stimulation. However, the strong increases in ATAC-seq signal at the *Irg1* and *Gbp5* promoters were surprising, given the rapid RelA binding and induction of these genes.

The third experiment performed was ChIP-qPCR examining RelA binding in *Irf3*$^{-/-}$ macrophages. This experiment revealed strong IRF3 dependence of RelA binding to the *Ccl5* promoter (Figure 6F), consistent with our evidence from nuclease accessibility experiments that IRF3 is important for nucleosome remodeling at this promoter (Ramirez-Carrozzi et al., 2009). At the *Ifnb1* promoter, the initial binding of RelA was not dependent on IRF3 (Figure 6F); however, the increase in RelA binding at later time points exhibited IRF3 dependence, consistent with the notion that IRF3 stabilizes RelA binding while promoting synergistic transcriptional activation (Agalioti et al., 2000; Panne et al., 2007). Finally, although potently induced ATAC-seq signals were observed at the *Irg1* and *Gbp5* promoters, RelA binding to these promoters was not IRF3 dependent (Figure 6F). Thus, the nucleosome remodeling observed at these promoters by ATAC-seq is likely to be dictated by NF-κB itself or by other rapidly induced factors.

Together, these results support a model in which the mechanisms by which NF-κB and IRF3 regulate the *Ccl5* and *Ifnb1* genes are unique, with these two transcription factors contributing to *Irg1* and *Gbp5* activation (and possibly *Cxcl10* activation) by a third distinct mechanism. To determine whether these mechanisms appear to be unique only because we focused on a stringently defined group of PRGs, we asked whether any additional annotated promoters throughout the genome could be identified that possess the basic DNA properties of the five NF-κB/IRF3 genes (i.e., a strong RelA ChIP-seq peak [>19], a strong NF-κB motif [*Z* score >6.4], a strong IRF3 motif [Transfac score $\geq$ 90], and a distance between the NF-κB and IRF3 motifs of less than 100 bp [see Figure 6D]). Strikingly, only six additional promoters from among the 21,168 annotated promoters share these properties (data not shown).

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

(C) For each PRG, the higher maximum percent expression from either *Trif*$^{-/-}$ or *Irf3*$^{-/-}$ BMDMs (y axis) was assessed against the best scoring IRF3 motif (x axis) within the promoter based on the IRF Transfac PWM. The five IRF3/NF-κB genes are shown in blue, and the four IRF3 genes are shown in green. The PRGs containing strong NF-κB promoter motifs and RelA ChIP-seq peaks are shown in red. The horizontal dashed line indicates the expression threshold (33%), and the vertical dashed line indicates the Transfac threshold (90).

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

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

(F) RelA ChIP-qPCR was performed using WT and *Irf3*$^{-/-}$ BMDMs stimulated with lipid A. The relative enrichment of RelA binding was normalized to a negative control region. The RelA binding kinetics at the promoters of the five NF-κB/IRF3 genes were compared to the *Tnfaip3* promoter as a control (far right). The data shown represent an average of three biological replicates. Error bars indicate the SE. $^{**}p < 0.01$; $^{*}p < 0.05$.

Regulation of Transiently Transcribed Genes by Serum Response Factor

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

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

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

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

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

Last, an analysis of the 132 PRGs led to the identification of only two additional genes that exhibit similarly rapid induction kinetics as the six genes discussed above: *Btg2* and *Ier2* (Figure 7D). These two genes lack SRF ChIP-seq peaks and motifs in their promoters but instead contained strong promoter NF-κB ChIP-seq peaks. This finding raises the question of how these two genes achieve induction kinetics similar to those of the genes whose promoters are directly bound by SRF. Interestingly, both of these genes contain strong SRF ChIP-seq peaks at upstream regions (Figure 7E); in both instances, the SRF peaks coincide with CpG islands and are conserved through evolution (data not shown). The SRF peaks at the *Btg2* and *Ier2* loci are 10 and 1 kb upstream of their TSSs, respectively. Remarkably, only three other PRGs contained SRF ChIP-seq peaks within the region 10 kb upstream of their TSS, indicating that this property is rare. These results suggest that SRF contributes to the early transient induction of these genes by cooperating with NF-κB bound to the promoters.

DISCUSSION

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

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

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

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

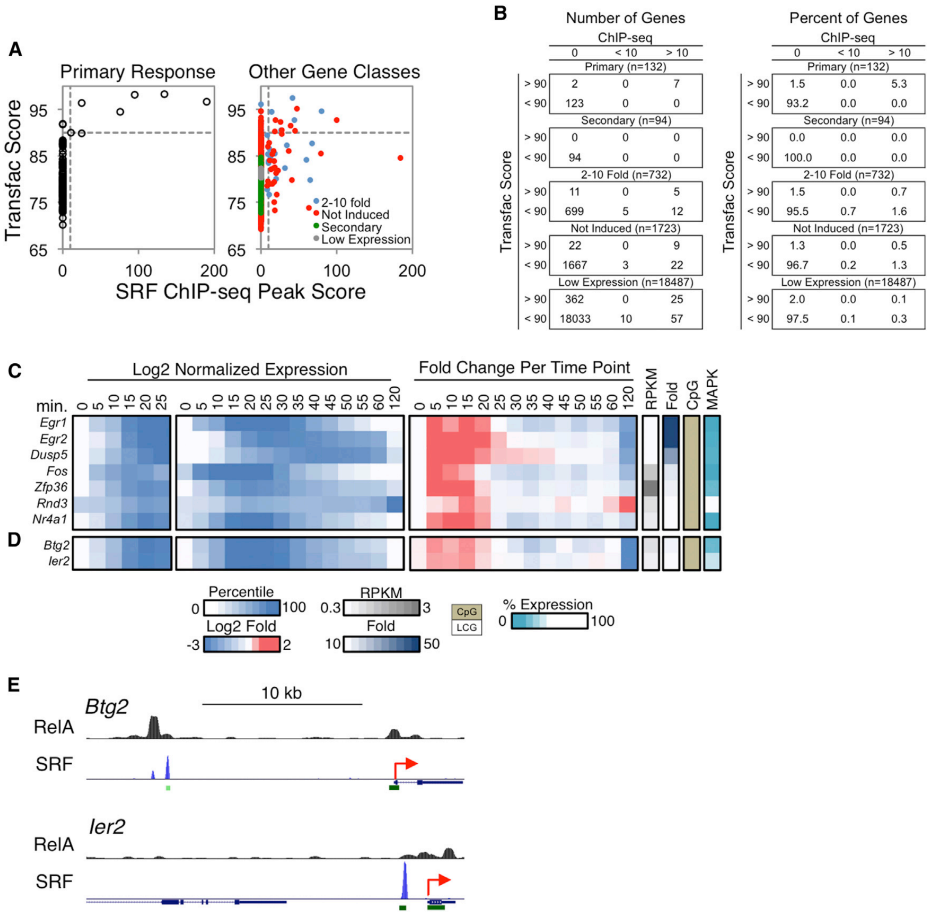


Figure 7. Analysis of SRF Target Genes

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

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

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

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

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

See also Figures S6 and S7.

these same factors play more subtle roles in the regulation of hundreds of additional genes. For example, although only a small fraction of genes induced by 2- to 10-fold contain promoter binding sites for NF-κB, this factor may contribute to the induction of a large fraction of weakly induced genes by binding to distant enhancers. Nevertheless, the current results document a clear distinction between strongly induced and weakly induced genes with respect to the prevalence of NF-κB promoter binding and promoter motifs, providing a framework for studies to elucidate the diverse mechanisms by which NF-κB contributes to the lipid A response. Similarly, the ability to identify consistent properties of genes that appear to be regulated by SRF, IRF3, and IRF3/NF-κB provides a step toward a precise understanding of the broad mechanisms regulating the transcriptional cascade. In addition

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

EXPERIMENTAL PROCEDURES

Cell Culture

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

RNA-Seq

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

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

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

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

ChIP-Seq

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

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

ATAC-Seq

ATAC-seq libraries were prepared using the Nextera Tn5 Transposase kit (Illumina) as described (Buenrostro et al., 2015) with slight modifications. Libraries were single-end sequenced (50 bp) on an Illumina HiSeq 2000. Reads were mapped to the mouse genome (NCBI37/mm9) using Bowtie2. Reads were removed from the subsequent analysis if they were duplicated, mapped to mitochondrial genome, or aligned to unmapped contiguous sequences. Promoter accessibility was calculated by totaling all reads within −500 to +150 bp relative to the TSS. The reads were converted to reads per million (RPM) by dividing by the total number of reads per sample. The average RPM from four replicates was used to quantify the fold increase in promoter accessibility.

Motif Analysis

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

ACCESSION NUMBERS

All data are accessible through GEO Series accession number GSE67357.

SUPPLEMENTAL INFORMATION

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

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

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

Received: July 17, 2015
Revised: November 1, 2015
Accepted: January 13, 2016
Published: February 25, 2016

REFERENCES

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-β promoter. Cell *103*, 667–678.

Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science *326*, 257–263.

Barish, G.D., Yu, R.T., Karunasiri, M., Ocampo, C.B., Dixon, J., Benner, C., Dent, A.L., Tangirala, R.K., and Evans, R.M. (2010). Bcl-6 and NF-kappaB cistromes mediate opposing regulation of the innate immune response. Genes Dev. *24*, 2760–2765.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J.R. Stat. Soc. *57*, 289–300.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1–21.29.9.

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

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

Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol. Cell *47*, 810–822.

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

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

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

Kagan, J.C., Su, T., Horng, T., Chow, A., Akira, S., and Medzhitov, R. (2008). TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. Nat. Immunol. *9*, 361–368.

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

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods *7*, 709–715.

Lim, C.-A., Yao, F., Wong, J.J.-Y., George, J., Xu, H., Chiu, K.P., Sung, W.-K., Lipovich, L., Vega, V.B., Chen, J., et al. (2007). Genome-wide mapping of RELA(p65) binding identifies E2F1 as a transcriptional activator recruited by NF-kappaB upon TLR4 activation. Mol. Cell *27*, 622–635.

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

Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., et al. (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. Proc. Natl. Acad. Sci. USA *100*, 12247–12252.

Metcalfe, S.M. (2011). LIF in the regulation of T-cell fate and as a potential therapeutic. Genes Immun. *12*, 157–168.

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

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

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

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol. *29*, 436–442.

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

Ramsey, S.A., Klemm, S.L., Zak, D.E., Kennedy, K.A., Thorsson, V., Li, B., Gilchrist, M., Gold, E.S., Johnson, C.D., Litvak, V., et al. (2008). Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. PLoS Comput. Biol. *4*, e1000021.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Shih, H.Y., Sciumè, G., Poholek, A.C., Vahedi, G., Hirahara, K., Villarino, A.V., Bonelli, M., Bosselut, R., Kanno, Y., Muljo, S.A., and O'Shea, J.J. (2014). Transcriptional and epigenetic networks of helper T and innate lymphoid cells. Immunol. Rev. *261*, 23–49.

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

Toshchakov, V., Jones, B.W., Perera, P.-Y., Thomas, K., Cody, M.J., Zhang, S., Williams, B.R.G., Major, J., Hamilton, T.A., Fenton, M.J., and Vogel, S.N. (2002). TLR4, but not TLR2, mediates IFN-beta-induced STAT1alpha/beta-dependent gene expression in macrophages. Nat. Immunol. *3*, 392–398.

Treisman, R. (1994). Ternary complex factors: growth factor regulated transcriptional activators. Curr. Opin. Genet. Dev. *4*, 96–101.

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

Zhao, B., Barrera, L.A., Ersing, I., Willox, B., Schmidt, S.C.S., Greenfeld, H., Zhou, H., Mollo, S.B., Shi, T.T., Takasaki, K., et al. (2014). The NF-κB genomic landscape in lymphoblastoid B cells. Cell Rep. *8*, 1595–1606.

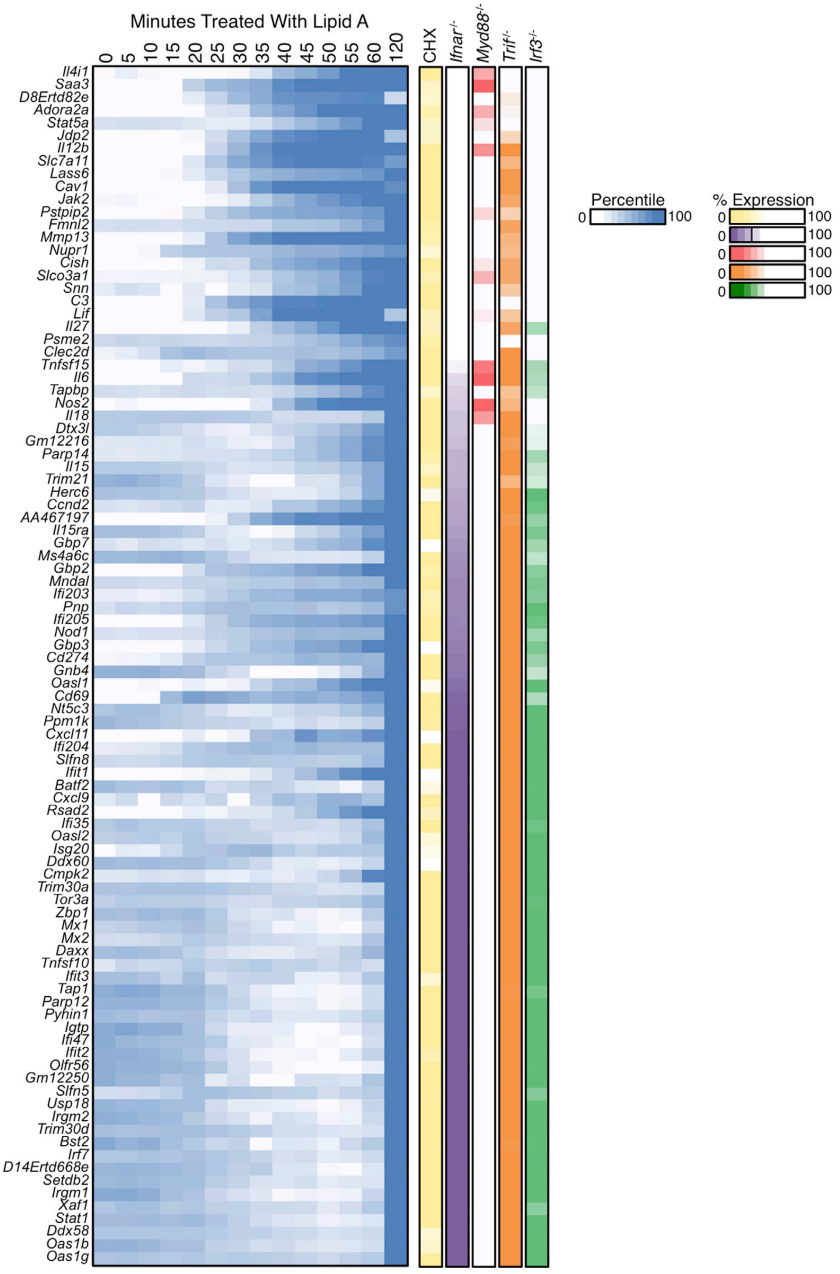


Figure S1. IFNAR-Independent and IFNAR-Dependent Secondary Response Genes, Related to Figure 2

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

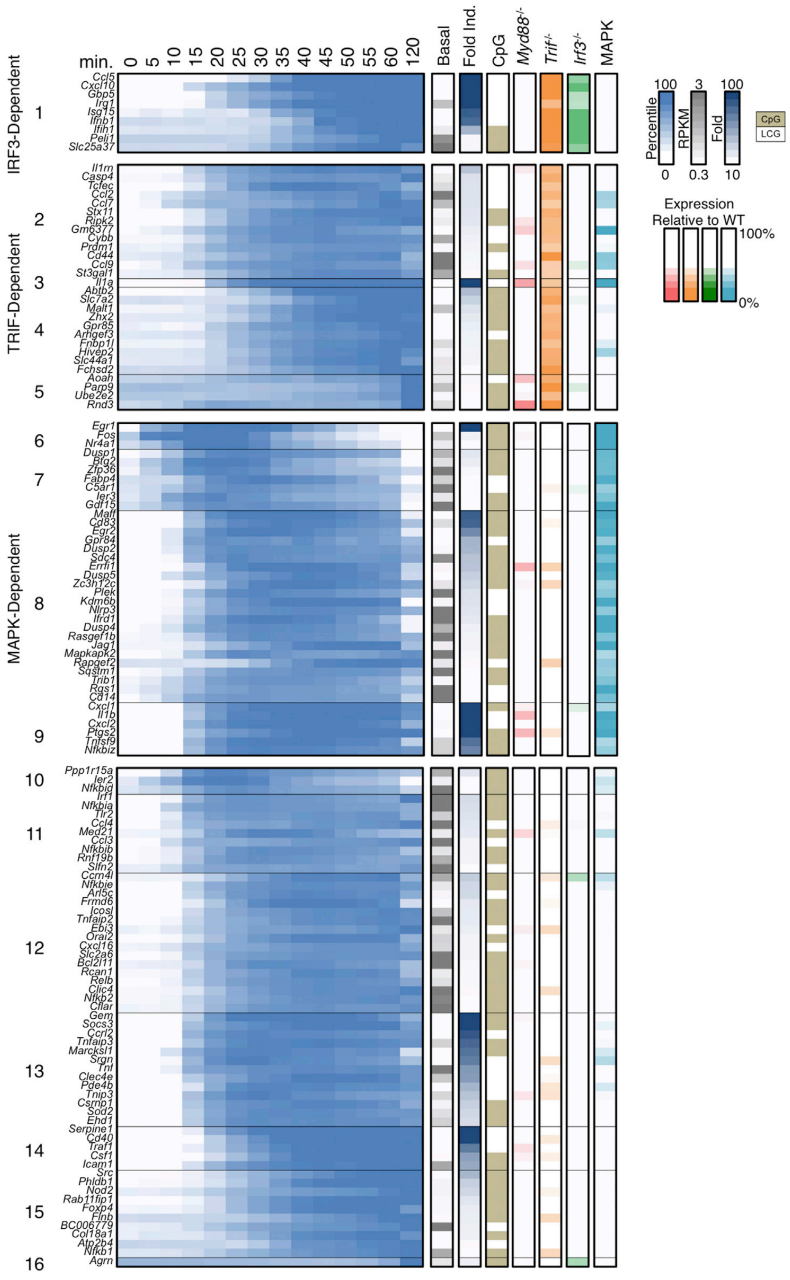


Figure S2. Highly Induced Primary Response Genes Ordered by Their Dependence on Various Signaling Pathways, Related to Figure 3

An expanded version of Figure 3D is shown to include gene names for each 10-fold significantly induced primary response gene.

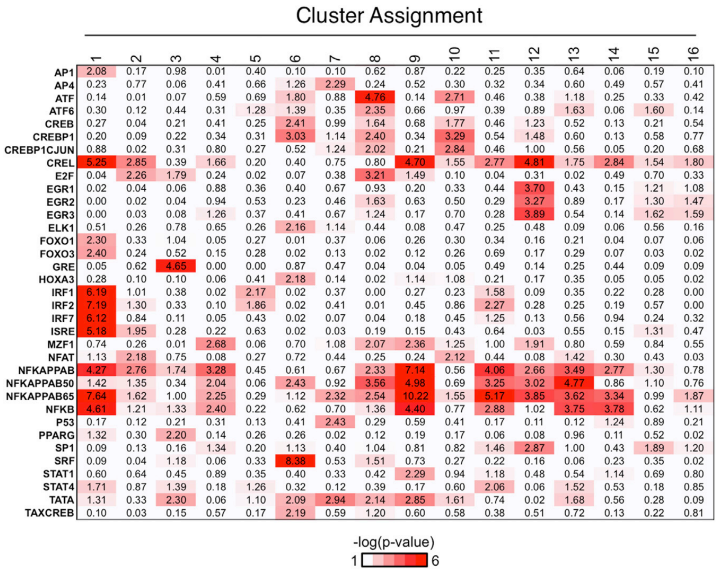


Figure S3. Promoter Motif Analysis of Primary Response Gene Clusters, Related to Figure 3

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

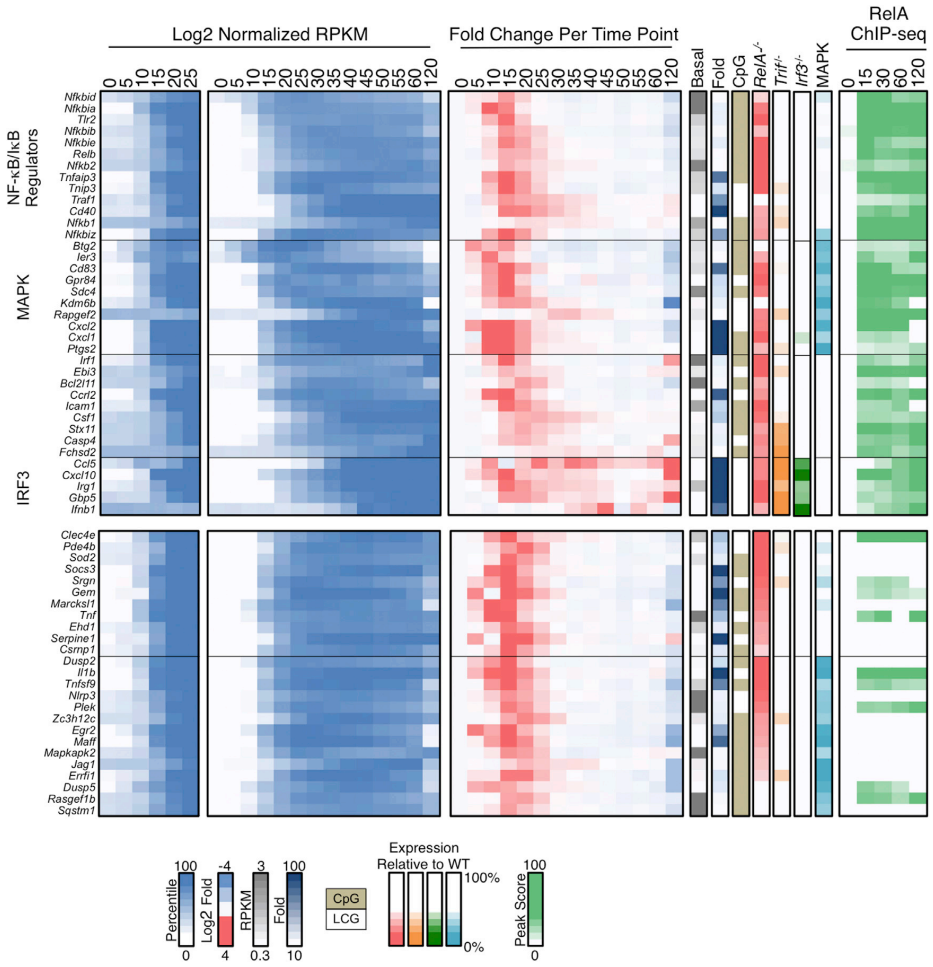


Figure S4. Putative NF-κB Target Genes and the Genes that Exhibit Similar Kinetics and/or RelA Dependence, Related to Figure 5

An expanded version of Figures 5A and 5B is shown to include gene names for each putative NF-κB target and other genes that may be enhancer regulated NF−κB target genes.

A

Peak Locations Relative to Annotated Genes

Number of Genes

Gene Category	Promoter	10kb	20kb	100kb	> 100kb	ChIP-seq Peak Score
	69	56	102	84	121	0
Primary (n=132)	6	9	7	15	4	< 19
	57	67	23	33	7	> 19
	72	64	82	75	91	0
Secondary (n=94)	5	10	4	5	1	< 19
	17	20	8	14	2	> 19
	598	428	577	490	681	0
2 to 10 Fold (n=732)	26	85	59	88	20	< 19
	108	219	96	154	31	> 19
	1584	1333	1559	1450	1673	0
Not Induced (n=1723)	47	156	77	113	19	< 19
	92	234	87	160	31	> 19
	17903	17379	17866	17313	18151	0
Low Expression (n=18487)	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
	52.3	42.4	77.3	63.6	91.7	0
Primary (n=132)	4.5	6.8	5.3	11.4	3.0	< 19
	43.2	50.8	17.4	25.0	5.3	> 19
	76.6	68.1	87.2	79.8	96.8	0
Secondary (n=94)	5.3	10.6	4.3	5.3	1.1	< 19
	18.1	21.3	8.5	14.9	2.1	> 19
	81.7	58.5	78.8	66.9	93.0	0
2 to 10 Fold (n=732)	3.6	11.6	8.1	12.0	2.7	< 19
	14.8	29.9	13.1	21.0	4.2	> 19
	91.9	77.4	90.5	84.2	97.1	0
Not Induced (n=1723)	2.7	9.1	4.5	6.6	1.1	< 19
	5.3	13.6	5.0	9.3	1.8	> 19
	96.8	94.0	96.6	93.6	98.2	0
Low Expression (n=18487)	1.2	3.1	1.8	3.1	0.9	< 19
	2.0	2.9	1.6	3.2	0.9	> 19

B

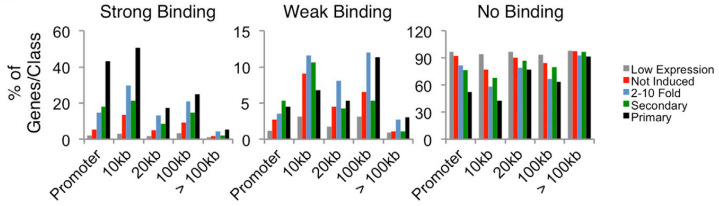


Figure S5. The Position of RelA Peaks Relative to the Transcription Start Sites of All Genes, Related to Figure 4

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

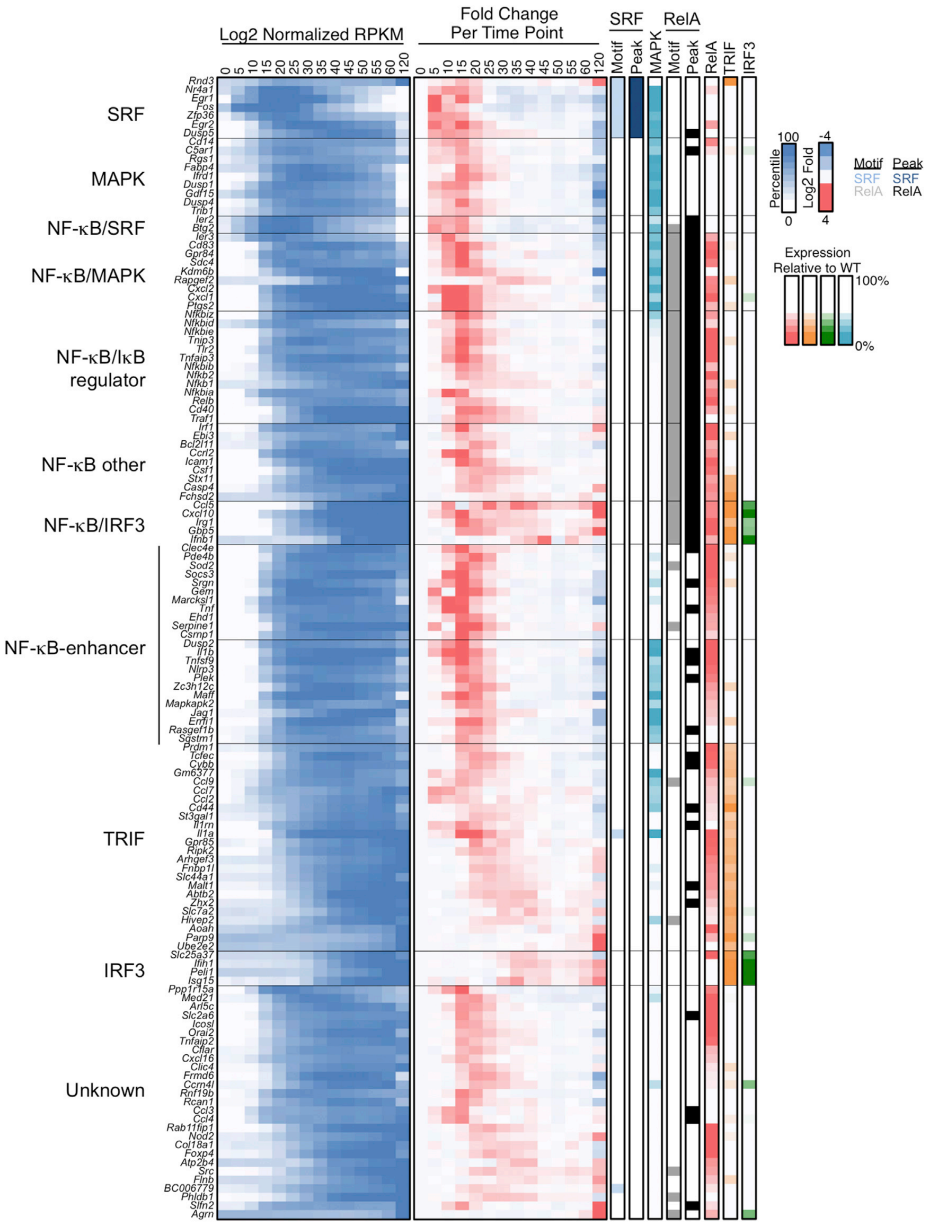


Figure S6. Final Classification of the Primary Response Genes, Related to Figure 7

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

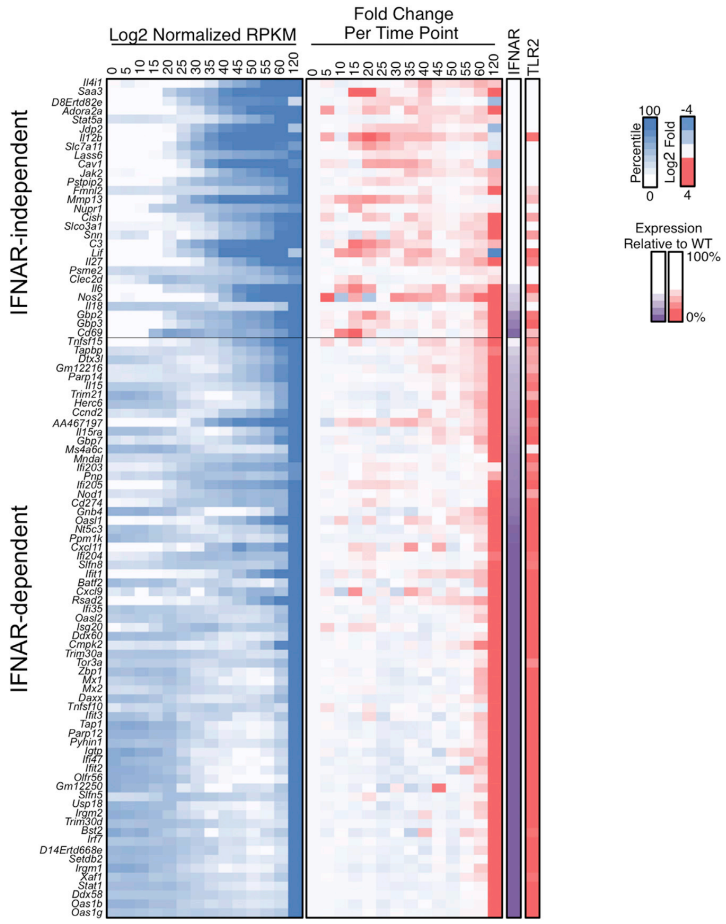


Figure S7. Final Classification of the Secondary Response Genes, Related to Figure 7

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

CHAPTER 5

Concluding Discussion

The preceding chapters have attempted to investigate mechanisms responsible for coordinating innate immune cell function as well as better understand the role of IL10 in whole body metabolism and in shaping the transcriptional response to lipid A. A number of findings have been highlighted in the discussions at the end of each chapter, and future work and open questions have been previously presented on the individual topics. This concluding chapter serves as a placeholder for a brief discussion on the future prospects of studies focused on understanding immune regulation and how new technologies will continue to play a role in shaping research questions as well as in the training of future researchers.

Even with the vast amount of sequencing data that has been generated during the LPS response in mouse immune cells, it can be argued that we have not fully appreciated the complexity of an artificial in-vitro defined system, let alone begin to understand how immune cells interact in-vivo during a live infection. This acknowledgment of our limitations in understanding should not be confused with any indication that we should not be excited by the situation we find ourselves in. The genomics revolution continues to hold promise for researchers interested in understanding basic principles of gene regulation, and will continue to propel us forward as we advance towards applying this knowledge to more complex questions related to systems biology and ultimately, to human development, pathology, and therapeutic interventions. If the prior 30 years are any indication of the types of advances we may see throughout the next 30 years, the future is likely one of tremendous discovery.

Sequencing methods continue to improve in sensitivity, cost, read length, and accuracy. Furthermore, additional sequencing platforms have emerged which rely upon the detection of electrical currents based on nucleobases traversing artificial membranes rather than fluorescent imaging based on chemical synthesis (Branton et al., 2008). If the newer sequencing platforms are able to achieve comparable levels of accuracy per price point, these current-based sequencers could dramatically change the way data is collected and analyzed, as the new method has the potential to monitor base pair modification directly on endogenous DNA and

RNA, a feat not possible with conventional commercial sequencers (Simpson et al., 2017). The newer sequencing devices continue to shrink in size; it is now possible to purchase USB-powered, portable sequencers, thus increasing the availability of new sequencing techniques to researchers around the world. Indeed, there is a growing number of non-traditional “citizen scientists” who analyze and generate sequencing data for personal interests. It is possible that if this trend continues there may be either crowd sourcing of research projects collectively funded and targeted towards specific interests of people, or a greater emphasis on bringing these new methods into K-12 education to continue preparing new students for careers in science through interactive discovery (Bonney et al., 2009).

References

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., and Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. Bioscience *59*, 977–984.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., et al. (2008). The potential and challenges of nanopore sequencing. Nat. Biotechnol. *26*, 1146–1153.

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. Nat Meth *advance on*, 407–410.