UNIVERSITY OF CALIFORNIA

Los Angeles

Response Specificity and Epigenomic Programmability in Innate Immune Sentinels

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Human Genetics

by

Katherine May Sheu

2021

ABSTRACT OF THE DISSERTATION

Response Specificity and Epigenomic Programmability
in Innate Immune Sentinels

by

Katherine May Sheu
Doctor of Philosophy in Human Genetics
University of California, Los Angeles, 2021
Professor Leonid Kruglyak, Co-Chair
Professor Alexander Hoffmann, Co-Chair

Innate immune sentinels serve to both safeguard the organism from foreign invaders and appropriately manage internal damage signals. These cells are distributed throughout the body – circulating the blood, lining the mucosal membranes and the blood vessels, surveying the connective tissue. Their role is to sense the environment, and then decide whether to respond and how best to respond, to maintain the health of the organism. While sentinels include several different cell types, including fibroblasts, endothelial cells, macrophages, and dendritic cells, we here focus on macrophages and examine two properties that may define their sentinel function: response-specificity and epigenomic programmability.

Immune sentinel macrophages initiate the immune response and interact with T-cells, B-cells, NK cells, etc. to coordinate downstream immune system activation. Considering the thousands of potential human pathogens, hundreds of cell damage molecules, and dozens of host-cytokines that may signify inflammatory danger, it should be evolutionarily unsurprising that macrophages respond differently to each signal. However, quantifying this specificity requires single cell data that reveals response distributions. In Chapter 2, we measured single-

cell transcriptomic profiles of macrophages responding to diverse bacterial, viral, and cytokine stimuli to quantify the response-specificity of macrophages. Using information theory, we uncovered distinct functional gene groups that differed in their degree of response-specificity. We found that response-specificity was context-dependent, as some genes lost specificity when macrophages were responding from IFNγ or IL4 cytokine contexts. Examining peritoneal macrophages from old and obese mice, which may harbor inflammatory microenvironments, revealed diminished response-specificity compared to healthy mice. These findings suggest that macrophage response-specificity quantified by single cell mRNA measurements after stimulation may be the basis for an innate immune health score.

Beyond responding to an immediate threat by deploying the appropriate transcriptomic programs, immune sentinels may also retain a stimulus-specific memory of the ligand encountered. Prior experimental studies suggested that stimulus-activated transcription factors when bound to DNA could provide the trigger for epigenomic reprogramming, but the mechanisms remained unclear. In Chapter 3, we developed mathematical models of the nucleosome to investigate mechanisms by which the spontaneous kinetics of nucleosomal-DNA interactions interfaced with the stimulus-specific dynamics of transcription factor activity. First focusing on the signal-dependent transcription factor NFκB, our model predicted that presence or absence of oscillations in NFκB signaling was the primary determinant of alterations to epigenomic state. We next developed more detailed Markov models of the nucleosome with base-pair resolution. Fitting the models to time-course stimulus-response ATAC-seq data, we found that kinetic rates of nucleosome unwrapping were slower *in vivo* than reported *in vitro* rates, and that the propensity for nucleosome eviction was greatest if transcription factor binding sites sat close to but not on the nucleosome dyad. This work uncovered biophysical principles governing nucleosome dynamics *in vivo,* enabling the prediction of epigenomic alterations during inflammation at single nucleosome resolution.

The dissertation of Katherine May Sheu is approved.

Thomas G. Graeber

Maureen A. Su

Leonid Kruglyak, Committee Co-Chair

Alexander Hoffmann, Committee Co-Chair

University of California, Los Angeles

2021

To all my friends, family, and mentors.
You inspire me to live with a sense of adventure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to the many people who contributed to this thesis, directly or indirectly. My sincere thanks to my advisor Alex Hoffmann, whose energy created a great environment to experience the art of science. There was always something exciting going on, a pace that set me on a path to make the most of my graduate years. But balanced with that, there was also enough space for the boredom necessary to generate new and innovative ideas. Outside of the science, I also appreciated Alex as a person – smart and sharp, yet more importantly, warm and caring towards other people, and steadfastly optimistic in the face of setbacks. I infinitely valued his generous mentorship, thoughtful emails, the life lessons and jokes, and the opportunities he opened.

I am also deeply grateful for long-running mentorship from Tom Graeber, who more than anyone set me up to pursue computational biology. When I was busy in medical school, I remember Tom would be available in his office at all sorts of evening hours to give feedback on my work and talk science. My thanks for all those cycles of examining results, looking up literature, and wrangling algorithms in real time, which helped me further develop skills and intuition for computational research. When I first started in the lab, one of Tom's students told me, 'Tom likes long meetings.' And that was often the case when Tom, Niko, Favour, and I sat down to discuss the small cell project we worked on together. But I never thought I'd leave another 3hour meeting even more energized than at the start, and in my own interactions with mentees, I hope to reflect Tom's inspiring enthusiasm and passion for science.

Of course, none of the science here would have been possible without the numerous friends and collaborators who contributed to it, whether generating data, analyzing data, or best of all, discussing data and ideas over an unrushed lunch or a record-long email thread. Specific contributions to projects included here are detailed at the close of each section, but I also want to acknowledge all the members of the Hoffmann and Graeber labs over the course of the past five years for the invigorating atmosphere that made the lab a place to play.

For these projects in particular, several groups of people patiently listened to my periodic updates and gave feedback whenever possible: I thank the Modeling and Gene Expression subgroups hosted by Alex every week – the energetic meetings were always an enjoyment. The Microscopy & Modeling group also provided fruitful exchange of ideas among several labs, and cross-pollination of methods and approaches. My committee members Tom Graeber, Maureen Su, and Leonid Kruglyak listened to various iterations of these projects and provided suggestions, thoughts about the implications, and opportunities for collaboration.

In thinking about how I ended up here to begin with, I need to first thank my undergraduate research advisor Guoping Fan. When I first interviewed for a spot in his lab as a freshman undergraduate, I had no research experience and confidently told him my goal was to go to medical school and be a physician. But over the next couple years, my research experience in his lab and his subtle encouragement to consider MD/PhD programs launched me onto this even more interesting path. The second thanks goes to my family – over the course of twenty years they gifted me the love of learning and instilled in me the importance of serving others, and then gave me freedom to follow any path I wanted to pursue.

For me as a graduate student, the end-product of a paper published is not as valuable as the collection of experiences that led up to it. Through working on papers, I have met and interacted closely with many brilliant people and forged many friendships. For every project, I can recall a time capsule of stories and conversations – laughter in the office, debates over statistics, anticipation over reviewer comments, or together mulling over both the expected and the unexpected. Still, some projects were eventually completed, and others are hopefully in the works: Chapter 2 is a paper in preparation (Sheu et al), “Quantifying the response-specificity of macrophages”. Chapter 3 is a collection of mathematical models that are part of a paper in press (Cheng, Ohta et al, 2021) and a paper in review (Kim, Sheu et al), “Stochastic models of nucleosome dynamics reveal regulatory rules of stimulus-induced epigenome remodeling”.

Chapter 2 is in part adapted from the following: Sheu, K.M., Guru, A., Hoffmann A. Quantifying the response-specificity of macrophages by single cell RNA profiling to score the health of innate immune function (*in prep*). KMS performed the experiments and conducted the analyses. AG helped with piloting the experimental approach and assisted with the machine learning analysis. KMS and AH wrote the paper. All authors reviewed the manuscript. AH coordinated and funded the work.

Chapter 3 is in part adapted from the following: Cheng, Q.J.*, Ohta S.*, Sheu, K.M., Spreafico, R., Adelaja A., Taylor, B., Hoffmann, A. NFκB dynamics determine the stimulus-specificity of epigenomic reprogramming in macrophages (*in press*, Science, 2021). QJC, SO, AA, and BT performed the experiments. QJC, SO, KMS, RS, and AA analyzed the data. KMS and BT developed the mathematical model. QJC, KMS, and AH wrote the manuscript. All authors reviewed the manuscript. AH coordinated and funded the work.

Kim J.*, Sheu, K.M.*, Cheng, Q.J., Hoffmann, A.[#], Enciso, G.[#]. Stochastic models of nucleosome dynamics reveal regulatory rules of stimulus-induced epigenome remodeling (*in review*). JK, KMS, AH, and GE conceived the project. JK and KMS developed the model and analyzed the data. JK analyzed the model and performed model computations. KMS performed bioinformatic analyses of the data. QJC provided experimental data and helpful input on the manuscript. JK, KMS, AH, and GE wrote the paper. All authors reviewed the manuscript. JK, KMS, AH, and GE coordinated and funded the work.

VITA

2016-Present Medical Scientist Training Program, University of California, Los Angeles.
(expected PhD 2021, MD 2023)

2012-2016 University of California, Los Angeles, B.S. *Summa Cum Laude*
Major: Molecular, Cell, and Developmental Biology, Highest Honors
Minors: Bioinformatics; Biomedical Research

PUBLICATIONS

14. **Katherine M. Sheu**, Aditya Guru, Alexander Hoffmann. Quantifying the response-specificity of macrophages by single cell RNA profiling to score the health of innate immune function. (in prep)
13. Jinsu Kim*, **Katherine M. Sheu***, Quen J. Cheng, Alexander Hoffmann#, German Enciso#. Stochastic models of nucleosome dynamics reveal regulatory rules of stimulus-induced epigenome remodeling. (in review)
12. Stefanie Luecke*, **Katherine M. Sheu***, Alexander Hoffmann. Signaling Mechanisms regulating Innate Immune Responses: A Systems Biology Perspective. (accepted, *Immunity*, 2021)
11. Quen J. Cheng*, Sho Ohta*, **Katherine M. Sheu**, Roberto Spreafico, Adewunmi Adelaja, Brooks Taylor, Alexander Hoffmann. NFκB dynamics determine the stimulus-specificity of epigenomic reprogramming in macrophages. (in press, *Science*, 2021)
10. Adewunmi Adelaja*, Brooks Taylor*, **Katherine M. Sheu**, Yi Liu, Stefanie Luecke, Alexander Hoffmann. Six distinct NFκB signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. *Immunity,* 2021, 54, no. 5: 916-930.e7.
9. Yeon Joo Kim, **Katherine M. Sheu**, Jennifer Tsoi, Gabriel Abril-Rodriguez, Catherine Grasso, Davis Y. Torrejon, Ameya S. Champhekar, Kevin Litchfield, Charles Swanton, Daniel Speiser, Philip O. Scumpia, Alexander Hoffmann, Thomas G. Graeber, Cristina Puig-Saus, Antoni Ribas. Melanoma de-differentiation induced by interferon-gamma epigenetic remodeling in response to anti-PD-1 blockade therapy. *The Journal of Clinical Investigation*, 2021, 131, no. 12.
8. Supriya Sen*, Zhang Cheng*, **Katherine M. Sheu**, Yu Hsin Chen, Alexander Hoffmann. Gene regulatory strategies that decode the duration of NFκB activity contribute to LPS- vs TNF-specific gene expression. *Cell Systems*, 2020, 10, no. 2: 169-182.e5.
7. **Katherine M. Sheu***, Stefanie Luecke*, Alexander Hoffmann. Stimulus-specificity in the responses of immune sentinel cells. *Current Opinion in Systems Biology*, 2019, 18: 53–61. (*equal contribution)
6. Youjin Hu†#, Jiawei Zhong#, Yuhua Xiao, Zheng Xing, **Katherine Sheu**, Shuxin Fan, Qin An, Pinhong Rao, Yuanhui Qiu, Yingfeng Zheng, Xialin Liu, Guoping Fan, Yizhi Liu†. Single-cell RNA cap and tail sequencing (scRCAT-seq) reveals subtype-specific isoforms differing in transcript demarcation. *Nature Communications,* 2020, 11, no. 1: 5148.

5. Nikolas G. Balanis*, **Katherine M. Sheu***, Favour N. Esedebe, Saahil J. Patel, Bryan A. Smith, Jung Wook Park, Salwan Alhani, Brigitte N. Gomperts, Jiaoti Huang, Owen N. Witte‡, and Thomas G. Graeber‡. Pan-cancer convergence to a small-cell neuroendocrine phenotype that shares susceptibilities with hematological malignancies. *Cancer Cell*, 2019. 36, no. 1: 17-34.e7. (*equal contribution)
4. Jung Wook Park, John K. Lee, **Katherine M. Sheu**, Liang Wang, Nikolas G. Balanis, Kim Nguyen, Bryan A. Smith, Chen Cheng, Brandon L. Tsai, Donghui Cheng, Jiaoti Huang, Siavash K. Kurdistani, Thomas G. Graeber*, Owen N. Witte*. Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. *Science*, 2018, 362, no. 6410: 91–95.
3. Bryan A. Smith ‡, Nikolas G. Balanis ‡, Avinash Nanjundiah, **Katherine M. Sheu**, Brandon L. Tsai, Qingfu Zhang, Jung Wook Park, Michael Thompson, Jiaoti Huang, Owen N. Witte*, Thomas G. Graeber*. A human adult stem cell signature marks aggressive variants across epithelial cancers. *Cell Reports*, 2018, 24, no. 12: 3353-3366.e5.
2. Jennifer Tsoi, Lidia Robert, Kim Paraiso, Carlos Galvan, **Katherine M. Sheu**, Johnson Lay, Deborah J.L. Wong, Mohammad Atefi, Roksana Shirazi, Daniel Braas, Catherine S. Grasso, Nicolaos Palaskas, Antoni Ribas, Thomas G. Graeber. Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell*, 2018, 14;33(5):890-904.e5.
1. Youjin Hu, Qin An, **Katherine Sheu**, Brandon Trejo, Shuxin Fan, Ying Guo. Single cell multi-omics technology: methodology and application. *Frontiers in Cell and Developmental Biology*, 2018, 6:28.

PRESENTATIONS

- UCLA QCBio Research Seminars, April 2020. *Stimulus-specificity of gene expression dynamics in macrophage immune responses.*
- Annual Symposium, NSF Simons Center for Multiscale Cell Fate, Oct 2019. *Stochastic modeling of nucleosome-DNA interactions identifies mechanisms that relate chromatin accessibility to SDTF dynamics.*
- UCLA Dept. of Molecular and Medical Pharmacology Retreat, November 2017. *Highly concordant neuroendocrine dedifferentiation states reveal a convergent pan-cancer therapy resistance route.* Oral presentation award.
- AAAS Pacific Division Conference, June 2015. *Examining the Role of DNA Methylation in Naive Pluripotency Using DNA Methyltransferase Knockout and Reconstitution Mouse Embryonic Stem Cells,* First Place Award in Cell and Molecular Biology Section. Robert I. Larus Award.

AWARDS AND HONORS

2016-present	Medical Scientist Training Program, UCLA
2021	Schering Award for outstanding achievement in research, UCLA MIMG
2018-2020	Systems and Integrative Biology Training Grant, UCLA T32
2020	Sidney C. Rittenberg Award, UCLA MIMG
2018-2019	NSF Simons Center Interdisciplinary Opportunity Award
2012-2016	Regents Scholarship, University of California, Los Angeles
2014-2016	Howard Hughes Undergraduate Research Program, UCLA

CHAPTER 1

Introduction

PREFACE

Ilya Mechnikov provided a first description of the function of macrophages in the late 19th century, when he noted a striking observation: 'white corpuscles' moved to surround a small splinter embedded in a starfish larva (Mechnikov, 1908). Macrophages are first responders to injury or pathogen threats and are the drivers of immune activation (Rivera et al., 2016). While macrophages play a large role in triggering downstream inflammatory events, sentinels in fact include many cell types, which all share the function of sensing the environment and responding appropriately. During an inflammatory event, sentinels alert the immune system to the threat through the direct production of cytokines, upregulation of antimicrobial or cell death proteins, release of second messengers, or antigen presentation that activates cells of adaptive immunity.

Cell types that perform sentinel functions are strategically placed. Bone-marrow derived monocytes circulate the blood, extravasating into tissues and differentiating into macrophages upon sensing a danger signal, whether a pathogen-associated molecular pattern (PAMP), damage-associated molecular pattern (DAMP), or cytokine. Almost every tissue in the body is populated with macrophages, but it is important to distinguish, however, that not all macrophages are monocyte-derived or act as sentinels (Franken et al., 2016). Subsets of tissue-resident macrophages derived from the yolk sac are longer-lived and instead primarily function as regulators of homeostasis (Varol et al., 2015). Dendritic cells such as the Langerhans cells of the mucosa and skin are, like macrophages, also antigen-presenting sentinels (Doebel et al., 2017). The subset of dendritic cells that function as sentinels can be activated by a wide range of PAMP or DAMP stimuli, which trigger cellular transcriptomic programs that result in changes to receptor expression, release of chemokines, or importantly, migration towards local lymphoid tissue (Stockwin et al., 2000).

Beyond these classical immune cell types, certain structural cells like fibroblasts and endothelial cells, are known to have sentinel functions (Krausgruber et al., 2020). Endothelial cells line the blood vessels and play multiple physiological roles, but increasing evidence points to their importance as immune sentinels (Krausgruber et al., 2020; Mai et al., 2013; Pober and Sessa, 2007): Placed in a prime position to respond to circulating endotoxins like lipopolysaccharide (LPS), endothelial cells express TLRs, NLRs, and cytokine receptors (Opitz et al., 2006, 2009), themselves produce inflammatory cytokines to mobilize other immune cells (Anand et al., 2008; Tissari et al., 2005), and participate in antigen presentation (Danese et al., 2007). Indeed, mice that expressed the LPS receptor TLR4 exclusively on endothelial cells were able to clear a systemic LPS infection without the action of TLR4-competent monocytes (Andonegui et al., 2009). Fibroblasts, which form part of the connective tissue, are tissue-resident sentinels, similarly express PAMP, DAMP, and cytokine receptors, and activate immune response genes upon ligand challenge (Bautista-Hernández et al., 2017; Cheng et al., 2017; Davidson et al., 2021; Sen et al., 2020).

Sentinel cells, despite disparate ontogenies, share certain characteristics. These characteristics are less defined by developmental origin or their steady state epigenomic or transcriptomic profiles (Epelman et al., 2014), but rather by the manner in which they respond to environmental inflammatory signals. In this thesis, we investigate two properties which we propose are hallmarks of healthy immune sentinel function. As functional hallmarks, they are emergent properties of responses, or having responded: 1) Response-Specificity, and 2) Stimulus-Specific Memory. Briefly, the first property addresses the capacity of sentinel cells to respond specifically to ligands encountered through a well-regulated signaling and epigenomic network, despite having a limited number of receptors, unlike the cells of the adaptive immune system which can generate infinitely many receptors through genetic recombination. The second property addresses the ability of healthy sentinels to selectively

retain stable memories for prior encounters with specific stimuli, through distinct epigenetic changes.

The mechanisms behind signaling and epigenome regulation of sentinel response have been closely studied and reviewed (Akira et al., 2006; Beutler, 2009; Mogensen, 2009; Murray et al., 2014). However, it has been less obvious how these how these molecularly-detailed biochemical and biophysical mechanisms give rise to these two immunological functions of sentinels. Recent studies quantifying the stimulus-specificity of signaling dynamics (Adelaja et al., 2021), on feedforward and feedback mechanisms (Gottschalk et al., 2019; Kok et al., 2020), or on context- or exposure-dependent epigenomic changes (Cheng et al., 2021; Kang et al., 2017), have further elucidated these connections. Furthermore, new technology such as reliable live-cell microscopy, cost-effective scRNAseq and multi-dimensional flow cytometry, and algorithmic advances in analysis of the data, have allowed us to probe and measure these hallmarks accurately. Because they concern function and not steady states, the hallmarks are a) evident only upon a perturbation that tests sentinel responses, and b) dysregulated in diseases where sentinel function may be challenged. Like other hallmarks – of aging, of cancer – their definition and scope are subject to expansion and revision as more is realized.

RESPONSE-SPECIFICITY

Sentinel cells express a large but limited set of receptors on the cell surface and intracellularly. While the abundance and fractional distribution of each type of receptor varies across cell types, these receptors enable sentinels to sense the environment. PAMP ligands are sensed by the Toll-like receptors (TLRs), NOD-like receptors (NLRs), RIG-I-like-receptors (RLRs), cytokines by cytokine receptors, and DAMPs by cytosolic DNA receptors like cGAS, among others. It is now appreciated that macrophages respond to ligands with stimulus-specific signaling profiles that then activate specific gene expression programs (Cheng et al., 2017,

2019; Ozinsky et al., 2000; Sheu et al., 2019). Yet, early studies of macrophages focused on a common core response, highlighting the common phagocytic and antigen-presenting abilities that empowered downstream antigen-specificity through the genetic diversity of adaptive immune cell receptors (Nau et al., 2002).

Only later did studies instead highlight a high degree of gene expression specificity in macrophages, dendritic cells, and fibroblasts. While these studies have shown that responses can be ligand-specific, the property of response-specificity relies not on the average expression values of a population of sentinels, but rather on the spread in single cell responses and the degree of overlap in ligand-response distributions. Macrophages perform as individual units in their function as sentries; each has the capacity to sense molecular patterns and activate signaling pathways to generate a response. Although macrophages within tissue are comprised of distinct cellular subsets (Dehne et al., 2016), response-specificity refers to the degree of heterogeneity in responses of a single uniform population of cells, rather than that due to cell subsets with distinct developmental origins (Gordon and Taylor, 2005). The diversity of responses to a single stimulus across individual cells with similar baseline states may increase effectiveness of the overall immune response for the organism.

Molecular components: receptors, kinases, and transcription factors

Response-specificity relies on molecular components and pathways that are activated stimulus-specifically. Not only ligand, but also ligand dose, and ligand exposure duration generate distinct responses (DeFelice et al., 2019; Gottschalk et al., 2016; Hoffmann et al., 2002; Zhang et al., 2017). The control of Response-specificity can be summarized into two principles of combinatorial and temporal control: the stimulus-specific deployment of target genes is dependent on 1) different stimuli activating different combinations of signaling regulators, or

2) different stimuli activating the same set of regulators but with different temporal patterns of activity.

These two principles of control emerge through multiple levels of signaling: receptors, adapters, kinases, and transcription factors (Figure 1a). First, receptors enable macrophages and other sentinels to sense the environment. PAMP ligands are sensed by the Toll-like receptors (TLRs), NOD-like receptors (NLRs), RIG-I-like-receptors (RLRs), cytokines by cytokine receptors, and DAMPs by cytosolic DNA receptors like cGAS, among others (Kawai and Akira, 2011; Takeuchi and Akira, 2010). Stimulus-specific temporal control of receptor activity is generated through receptor-ligand half-lives, receptor internalization, or receptor synthesis and degradation. Of note, at the species level, PRRs in humans and mice are evolutionarily divergent, with different numbers or TLRs and promoter sequences affecting their regulation (Rehli, 2002). Human macrophages and monocytes more poorly express TLR3 than mouse bone-marrow-derived monocytes, affecting ligand-receptor interactions for double-stranded RNA such as poly(I:C) (Heinz et al., 2003). These differences in receptor expression and regulation make response-specificity also a function of cell type and species.

While the abundance and fractional distribution of each type of receptor varies across macrophage types, across species, and even from cell-to-cell, response-specificity begins with the availability, regulation, and turnover of receptors. The initial expression and regulation of receptors after stimulation subsequently affect the activation patterns of adapter proteins that link ligand-bound receptors to intracellular signaling cascades. For instance, all TLRs engage just two adapters, Myd88 and TRIF. In response to lipopolysaccharide (LPS), TLR4 signals at the plasma membrane and initiates the oligomerization of the adapter Myd88, or can be internalized to signal through the endosome, where it interacts with a different adapter TRIF to initiate TRIF-dependent signaling. LPS dose-response specificity is temporally controlled at the adapter level by the different oligomerization dynamics of Myd88 and TRIF (Cheng et al.,

2015). Combinatorial control can also occur at the level of adapters – while TLR4 activates Myd88 and TRIF, TLR3 activates only TRIF, and TLR2 only Myd88, resulting in stimulus-specific responses for LPS, p(I:C), and Pam3CSK4, respectively. Adapters and associated ubiquitin chains ultimately activate a limited set of kinase-transcription factor modules. Signaling events originating at the receptor level thus converge on only four main kinase-transcription factor (TF) modules in the immune response – TBK1-IRFs, MAPKs-AP1, IKK-NFκB, and JAK-STAT.

Response-specificity is an output of ligand-receptor interactions and can be measured at multiple regulatory layers (Figure 1b). As signaling events converge on only three main transcription factor (TF) families in the immune response, interferon regulatory factors (IRF), AP1, and NFκB, a convenient point of measurement centers around TF activation profiles, by nuclear translocation or phosphorylation. Further informative measurement points are either the upstream kinase regulator of TF activation (MAPK ERK, p38, JNK, or IKK), or the downstream the result of that activity in the form of transcriptional and proteomic profiles.

Heterogeneity of transcription factor activation

Population level studies carried out in the 2000s were performed in fibroblasts and elucidated many molecular components of specific responses (Covert et al., 2005; Shih et al., 2009; Werner et al., 2005). However, not until the 2010's was appropriate single cell technology developed to allow insight into single cell response distributions. One early study performed in immortalized macrophage RAW cells used immunofluorescence to capture NFκB and ATF activation levels across single cells at a fixed timepoint, over 12 different TNF doses spanning 4 orders of magnitude. Dose discrimination of TNF through the nuclear abundance of these TFs at 30 mins was only about 1.2 bits, or about 2.3 distinguishable TNF concentrations (Cheong et al., 2011). The bottleneck mechanism on the degree of dose-response specificity

was modeled to be due to the TNF receptor complex, a single node which then passes information down to multiple signaling pathways. For early timepoints, constitutive expression of a feedback regulator A20 helped decrease noise and increase the range of NFκB activation by suppressing NFκB basal activity levels, generating greater dose-specificity, especially for low doses, than in A20 knockout cells.

Live-cell microscopy on fluorescently tagged proteins, first in immortalized cells and then in primary immune sentinels, were a next advance that allowed insight into the temporal dynamics of signaling activity (Adelaja et al., 2021; Gottschalk et al., 2016; Lane et al., 2019; Selimkhanov et al., 2014). Considering the dynamic profiles of NFκB activation in response to different doses of LPS showed that the information contained in heterogeneous dose-response TF trajectories was higher than for any single timepoint measurement, but still limited by substantial cell-cell heterogeneity (Selimkhanov et al., 2014). Mechanisms of positive feedback from RelA were identified as one component enabling LPS dose-specificity (Sung et al., 2014). Furthermore, the first single cell NFκB signaling study performed in primary macrophages rather than immortalized RAW cells investigated both a range of doses, and an array of immune stimuli targeting both TLRs and cytokine receptors (Adelaja et al., 2021). Just six NFκB dynamical features, termed signaling codons, were shown to be key to maximally facilitating specificity to ligand dose and ligand identity. The degree of IKK activity was modeled to be the mechanism that determined ligand-specificity of oscillatory or non-oscillatory signaling, and a mechanism of paracrine TNF signaling further adjusted the specificity of low-dose CpG responses, but the primary sources of molecular noise with greatest impact on the heterogeneity of NFκB trajectory profiles remain unclear.

However, beyond temporal specificity of a single TF, immune responses are a product of multiple signaling pathways activated in combination, which are a large contributor to response-specificity as many immune response genes are combinatorically regulated by

multiple transcription factors. For immune sentinel responses, NFκB, IRF, and MAPK signaling together determine the expression of gene programs. Simultaneous measurement of NFκB and MAPKp38 in single cell macrophages at a single timepoint revealed that dose-response curves for each pathway were distinct, with MAPK being digitally activated above a ligand concentration threshold, revealing a larger overall dose-discrimination capacity in the resulting gene expression (Gottschalk et al., 2016). Mechanisms for MAPK controlled response-specificity is not only due to transcription, but also MAPK control of mRNA half-life and transport. A further study that measured single cell temporal dynamics for both NFκB and JNK indicated that the two pathways combined were biologically informative and reflective of different levels of threat from pathogenic vs non-pathogenic organisms (Lane et al., 2019).

Heterogeneity of gene expression programs

Moving from studies of transcription factor and kinase activation, a key advance to better measuring response-specificity will be to understand the resulting heterogeneity of gene expression. The cell-to-cell variability present in signaling pathway dynamics and activation levels may be either buffered or amplified by the chromatin-associated or post-transcriptional regulatory mechanisms controlling the expression each gene. Technological developments have been required to link signaling to gene expression in the same cell, such as smFISH and scRNAseq, and advances in microfluidics and image analysis (Lee et al., 2014; Tay et al., 2010; Van Valen et al., 2016). In one study in macrophage RAW cells, after measuring NFκB signaling dynamics in response to LPS and then profiling the transcriptome of resulting cell at the end point, it was found that the expression of some cytokine and feedback regulator genes were correlated to the cell's NFκB dynamics (Lane et al., 2017). In addition, certain pairs of genes possibly controlled by the same enhancer elements maintained correlated expression

levels across single cells. Further work in profiling transcriptomic response-specificity across multiple ligands and doses are needed to understand the extent to which signaling dynamics are decoded stimulus-specifically. As with studies of signaling dynamics in sentinel cells, both statistical analysis and mechanistic modeling may prove insightful in elucidating mechanisms and sources of biological noise in single cell gene regulation.

Cytokine feedback and feedforward as an element of response specificity

Furthermore, regulatory motifs such as feedback loops and cytokine feedforward contribute to response-specificity, especially at late time points. After the initial induction of gene expression in response to a ligand encounter, immune sentinels continue sensing the environment, including cytokines transcribed and secreted in a paracrine or autocrine manner. For example, activation of the IRF pathway by PAMP stimuli such as LPS, poly(I:C), and TNF induce the later production of interferon-β (IFNβ), which acts in a feedforward loop to produce ISGF3-driven gene expression programs in neighboring bystander cells (Ourthiague et al., 2015). A single cell study in dendritic cells responding to LPS showed that early paracrine secretion of IFNβ in just a handful of cells was important for antiviral gene expression in the population; at later timepoints, *Ifnar* and *Stat1*-dependent IFNβ paracrine signaling downregulated inflammatory genes not uniformly, but in a fraction of the cell population (Shalek et al., 2014). In a later study, this presence or absence of negative feedback from Type I interferons was shown to be biologically important for distinguishing Gram-negative versus Gram-positive bacteria in BMDM responses. Bacterial class-specific production of key cytokines such as CXCL1 and TNF was diminished in either IFNAR knockouts, or when IFNβ was exogenously supplied (Gottschalk et al., 2019; Peignier and Parker, 2021).

Like interferons, TNF production also may further amplify or curtail response-specificity. Single cell NFκB signaling dynamics in the presence of TNFRII, a soluble TNF

inhibitor, revealed that more cells became non-oscillatory or non-responsive, compared to low dose CpG or LPS stimulation without the inhibitor (Adelaja et al., 2021). While the effect of paracrine TNF signaling on single cell gene expression is unclear, one outcome may be an increased ability to discriminate different doses of Myd88-ligands. In addition, blocking TNF autocrine signaling decreased the heterogeneity of NFκB signaling profiles in response to LPS, suggesting that cell-to-cell variability of signaling was in part affected by the heterogeneity of cellular secretion of TNF (Lane et al., 2017). Thus, the production, secretion, and responses to soluble cytokines, and the single cell heterogeneity of these processes, may be a feature that can either expand or restrict sentinel ability to discriminate dose or identity of a pathogen or DAMP ligand.

Diseases of impaired response-specificity

For response-specificity to be a property of healthy sentinel function, abnormal regulation ought to contribute to disease phenotypes. Healthy response-specificity may be marked by particular response distributions to each immune threat, and both increases or decreases in the heterogeneity may result in disease (Figure 1b). As such, in these diseases, the behavior of outliers is critical. Rheumatoid arthritis, lupus, multiple sclerosis, and Sjogren's Syndrome all have remitting and relapsing characteristics (Aletaha and Smolen, 2018; Buch et al., 2021; Steinman, 2014). The intermittent nature in the presentation of multiple autoimmune diseases hints that low probability outlier events may underlie its etiology.

Indeed, both aberrant TNF and IFNβ regulation have been implicated as opposing sides of different autoimmune diseases (Banchereau and Pascual, 2006). Excessive IFNβ production from dendritic cells was postulated to be an initiator of the autoimmune disease systemic lupus erythematosus (SLE) (Hall and Rosen, 2010). As IFNβ has both feedforward and negative feedback functions on neighboring cells, the improper production of IFNβ from even a subset

of cells may have significant consequences on response-specificity. A recent single-cell study more directly couched the autoimmune disease Sjogren's Syndrome as involving loss of response-specificity. Loss of the NFκB negative feedback regulator *Nfkbia* diminished the distinguishability of macrophage responses to TNF vs poly(I:C), interestingly through the increased expression of IRF target genes in a fraction of TNF-stimulated cells (Adelaja et al., 2021). TNF-induced IFNβ production through IRF1 has also been implicated in rheumatoid arthritis, which could be corrected through JAK inhibitor drugs (Bonelli et al., 2019). Thus, the misregulation of the TNF vs IFNβ axes in autoimmune disease may provide clues into how to correct or control loss of response-specificity (Vila-del Sol et al., 2008).

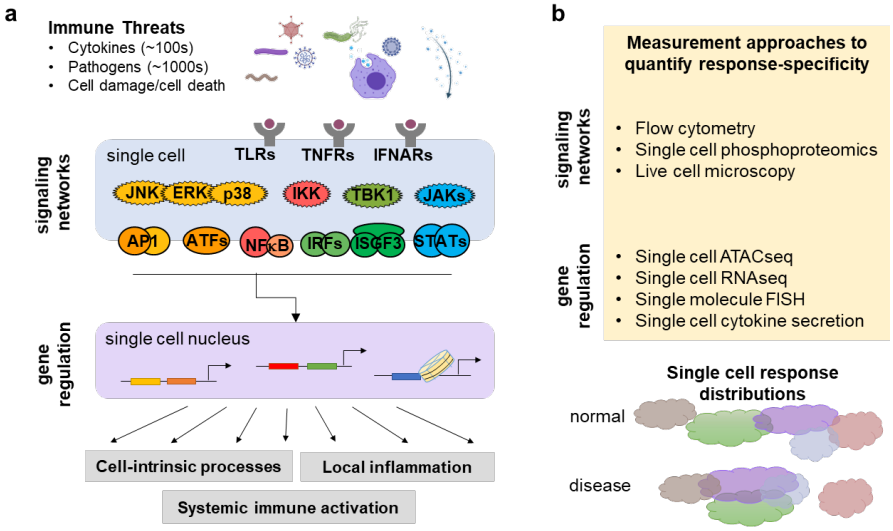


Figure 1.1. Mechanisms and measurement of response-specificity. A) Sources of heterogeneity in signaling network activation and transcriptional regulation that impact response-specificity. B) Single cell measurements of signaling or epigenetic events that can be interrogated to quantify response-specificity, resulting in an understanding of ligand-response distributions in normal vs disease.

EPIGENOMIC PROGRAMMABILITY

Distinct from chromatin remodeling during the immediate immune response, epigenetic memory refers to a longer-term property of sentinel cells to mark past exposures stably within the epigenome. The epigenetic landscape determines cell identity, but plasticity allows for varied functions over a cell's lifespan (Masserdotti and Götz, 2020). Early work on epigenetic cell identity established that differentiated cells were defined by an enhancer landscape, with nucleosomes defining euchromatin and heterochromatin regions held in place by the stable expression of a set of cell-type-specific transcription factors and the resulting deposition of histone marks (Allis and Jenuwein, 2016; Hayes and Wolffe, 1992). This category of transcription factors, termed lineage-determining transcription factors (LDTFs), are pioneer factors (Heinz et al., 2015; Soufi et al., 2015) that have structural elements that enable them to bind to nucleosomal DNA and adjust the enhancer landscape during development (Garcia et al., 2019).

Stimulus-specific epigenetic programming concerns the malleability of this established epigenetic landscape, which results in immune memory when changes are stably stored (Figure 2a). Unlike context-dependence of response-specificity, which affects the immediate response that is influenced by signaling crosstalk or epigenetic effects of cytokines still present in the environment, memory is maintained in the epigenome even after the initial challenge is no longer present. Thus, stimulus-specific epigenomic programming in essence stores marks of previous exposure to influence sentinel cell responses to future stimuli.

Molecular components: transcription factors, nucleosome remodelers, metabolites

Several classes of molecules in the nucleus mediate stimulus-specific epigenetic programming: transcription factors, nucleosome remodelers, metabolites. LDTFs are distinct in each cell type, and their binding to chromatin regions contributes to establishing the cell type's gene

expression potential. In macrophages, PU.1 and C/EBPβ are the LDTFs that are known to promote nucleosome-free regions and establish macrophage identity (Ghisletti et al., 2010). Interestingly, PU.1 has also been shown to cause fibroblasts to switch from a pro-inflammatory to a pro-fibrotic phenotype (Wohlfahrt et al., 2019). LDTFs have the ability to directly perturb nucleosome structure (Zaret, 2020) and thus cooperatively hold open chromatin regions for the subsequent binding of non-pioneer factors and other nucleosome remodeling complexes, resulting in the deposition of histone modifications that produce epigenetic memory.

However, for stimulus-induced epigenetic programming to occur, another category of transcription factors must also be activated. Signal-dependent transcription factors (SDTFs), like AP1, NFκB and IRFs, are activated stimulus-specifically by ligands, but are not cell-type specific. Initially, SDTFs were thought to primarily serve the function of immediately transcribing the appropriate immune response genes. Because the combinations of SDTFs activated are stimulus-specific, epigenetic memory thus also depends on the stimulus. SDTFs also have the capacity to impart more permanent changes on epigenomes of differentiated cells through their DNA binding mechanism (Ostuni et al., 2013a). Patterns of SDTF-DNA binding were shown to enable the stimulus-specific formation of de novo enhancers marked by H3K4me1 deposition, a covalent modification to the chromatin landscape (Ostuni et al., 2013a). Mechanisms by which the temporal dynamics of SDTF activity have also been elucidated. In fibroblasts, long duration NFκB activity was shown to be able to open chromatin (Sen et al., 2020). In macrophages, non-oscillatory NFκB activity allowed the necessary continuous residence time for nucleosome eviction and eventual H3K4me1 deposition (Cheng et al., 2021) (Figure 2b).

Because SDTFs like NFκB and IRF are not pioneer factors, it was originally unclear how SDTF activity could impart long-term changes to the epigenome. Crystallographic structures suggested that SDTFs like NFκB could indeed bind to nucleosomal DNA, and

displace histone H1 (Angelov et al., 2004; Lone et al., 2013a). The spontaneous unwrapping and rewrapping 'breathing' of DNA around the histone core octamer suggested that even SDTFs could invade the nucleosome (Li and Widom, 2004; Li et al., 2005a), by preventing rewrapping of the SDTF bound section of DNA back onto the nucleosome (Zhu et al., 2018). However, the rates of spontaneous rewrapping were rapid enough that NFκB bound only to its cognate motif when the sequence sat at the edge of the nucleosome, and not its interior. Thus, while stimulus-specific epigenetic memory is mediated by the activation of SDTFs, cooperative mechanisms from other proteins may be required (Comoglio et al., 2019). For example in macrophages, the LDTF PU.1 was colocalized with areas of SDTF binding and H3K4me1 deposition upon stimulation (Kaikkonen et al., 2013; Ostuni et al., 2013a). H3K4me1 marks active enhancers and persists even after H3K27ac and H3K4me3, a mark of active enhancers or promoters, is lost (Logie and Stunnenberg, 2016). Of note, the cooperative action of LDTFs also results in cell-type specificity in which regions become epigenetically programmed upon SDTF activation (Barozzi et al., 2014; Zhang and Glass, 2013). Other histone chaperone proteins such as FACT (Liu et al., 2020), and ATP-dependent nucleosome remodelers such as SWI/SNF (human ortholog SMARCA/B) or RSC (human ortholog BAF) may catalyze unwrapping or nucleosome sliding (Lorch et al., 2001; Wagner et al., 2020), thus allowing for SDTF binding (Brahma and Henikoff, 2020).

Metabolites and changes to metabolic circuits are an integral arm of epigenetic programming in immune sentinel cells (Penkov et al., 2019) (Figure 2c). These mechanisms are driven by the reliance of many epigenetic modifications on metabolic processes, such as one-carbon metabolism for histone and DNA methylation and generation of acetate pools from Acetyl-CoA for histone acetylation (Baardman et al., 2015; O'Neill et al., 2016; Van den Bossche et al., 2017). Furthermore, mevalonate and cholesterol biosynthesis pathways, which influence the innate immune response through Pi3K signaling (Akula et al., 2016), are

downstream of Acetyl-CoA production. Glycolysis and glutaminolysis have been both implicated in mediating trained immunity. Therefore, rewiring of metabolic circuits may be necessary to initiate and sustain immune memory (Arts et al., 2016a; Cheng et al., 2014). In human monocytes stimulated with B-glucan, glutaminolysis and cholesterol metabolism were required for trained immunity. In particular, these induced metabolic processes resulted in the accumulation of fumarate, which inhibited KDM5 histone demethylases to promote epigenetic reprogramming (Arts et al., 2016b). Fumarate treatment of monocytes itself also mimicked B-glucan treatment by increasing both H3K4me3 and H3K27ac deposition. Furthermore, in human macrophages exposed to IL4, alpha-ketoglutarate promoted demethylation of H3K27me3 in a manner dependent on Jmjd3, a histone demethylase (Liu et al., 2017).

Memory mechanisms due to prior infection

Stimulus-specific memory of past exposures serve the physiological purpose of changing future gene expression responses. Two main categories of innate immune memory, tolerance and trained immunity, are generated by different stimuli and alter responses in opposing directions (Ifrim et al., 2014). Memory-mediated tolerance was observed in macrophages exposed to a primary stimulus of high concentrations of lipopolysaccharide (LPS), and after a washout of up to five days, stimulated again with a secondary stimulus (Seeley and Ghosh, 2017). The resulting blunted second response, or tolerance, was shown to be mediated by nucleosome repositioning and histone H3 lysine methyltransferase G9a, which generated heterochromatin assembly and epigenetic silencing (El Gazzar et al., 2010). Interestingly however, tolerance is dose dependent: when high doses of LPS, P3CSK, and poly(I:C) are diluted 100-10000 fold, hyper-response replaces tolerance (Ifrim et al., 2014). Epigenetic

changes resulting in tolerance are stimulus-specific, but it remains to be seen how much the responses to heterologous secondary stimulation are also altered stimulus-specifically.

Trained immunity, involving hyper-response upon stimulation, is a key outcome of stimulus-specific epigenetic memory. In response to the fungal wall component beta-glucan, or Candida albicans, monocytes responded to secondary stimulation with much higher production of key cytokines like TNF and IL6 (Quintin et al., 2012; Saeed et al., 2014). Immune training was associated with increases in H3K4me3 and H3K4me1 (marking enhancers), even after loss of H3K27ac (marking active promoters), suggesting a stable epigenetic modification of enhancer regions help maintain trained immune memory. IFNγ secreted after initial challenge with *C. neoformans* was also shown to generate innate immune memory for up to 70 days, resulting in hyper responses of pro-inflammatory cytokines upon a second challenge (Wager et al., 2018)**.** Furthermore, dendritic cells have also shown stimulus-specific trained immune memory. DCs treated with the fungal pathogen *Cryptococcus neoformans*, transplanted into naïve mice, and challenged again, showed increased interferon response gene expression, as well as increased production of C. neoformans cytokines. This apparent memory was inhibited by treatment with histone methylase inhibitors (Hole et al., 2019).

Though studies on the epigenetic plasticity and memory of innate immune responses have focused on immune cells like macrophages, there is emerging evidence that fibroblasts, stromal cells, and hematopoietic stem cells may also be pliable to stimulus-specific epigenetic programming (Mitroulis et al., 2018). These cells have longer lifespans than circulating monocytes, and may thus be the optimal messengers to carrying memory of past exposures (Crowley et al., 2018; Hamada et al., 2019). In fibroblasts, for example, chromatin marks deposited after IFNβ stimulation led to faster and increased expression of interferon genes on a second stimulation (Kamada et al., 2018). Epithelial stem cells were also shown to maintain

memory of a primary response through sustained increase in chromatin accessibility at key inflammatory response genes, heightening responses to subsequent inflammatory stimuli (Naik et al., 2017).

Inducing memory for vaccination or disease treatment

Aside from its role in physiology, attempts have been to harness the training of innate immunity through vaccination. The tuberculosis vaccine BCG (Bacillus Calmette-Guérin) is a well-known example (Calmette, 1931), where vaccination with this bacterium provides cross-stimulus protection against multiple bacterial and fungal organisms through hyper-response upon secondary stimulation (Covián et al., 2019). BCG trained immunity not only affected monocytes via H3K27ac histone modifications (Arts et al., 2018), but also impacted the epigenetic landscape of hematopoietic stem cells (Cirovic et al., 2020; Kaufmann et al., 2018). BCG-trained HSCs led to epigenetically modified monocytes and macrophages that had alterations in H3K4me1, H3K4me3, and H3K27ac, and cleared tuberculosis infections more effectively than naïve macrophages (Kaufmann et al., 2018). Trained immunity of progenitor cells like HSCs may explain the lasting effects of innate immune vaccination. Importantly, while programming the epigenetic landscape is specific to the stimulus, unlike vaccines targeted at adaptive immunity that aim to generate memory B-cells, innate immune vaccination by BCG provides heterologous effects and protects individuals from many other bacterial, viral, or fungal pathogen threats (Chumakov et al., 2021; Pulendran and Ahmed, 2006).

Tolerance and immune training via treatment with LPS or BCG, respectively, has also been suggested as a potential avenue for the modulation of autoimmune diseases like systemic sclerosis (Jeljeli et al., 2019), which is marked by fibrosis as a result of ‘sterile inflammation’ (Dowson et al., 2017). Treatment of macrophages with low-dose LPS generated a tolerized phenotype that reduced inflammation-related fibrosis in a mouse model. On the other hand,

BCG exposure generated a trained phenotype with increased production of pro-inflammatory cytokines, exacerbating the fibrotic process. LPS and BCG generated unique epigenomic changes, with gene-specific changes in chromatin marks, including H3K4me3 (Jeljeli et al., 2019).

Diseases of dysregulated innate immune memory

Severe pathology can result from dysregulated immune memory. For example, sepsis, which involves hyperactivation of immune response as well as immune paralysis that prevents the clearance of bacteria in the bloodstream, affects millions of people yearly and nearly 1/3 of the people who die in hospitals have sepsis. Both tolerance and trained immunity play roles. Tolerance serves to eliminate hyper response on secondary stimulation, but misapplied regulation of tolerance results in poor host defense to secondary exposures to bacterial stimuli. Interestingly, the metabolic output of TCA cycle decarboxylation, itaconate, promoted tolerance in human monocytes, while beta-glucan inhibited IRG1, the enzyme that promotes itaconate synthesis, leading instead to an enhanced secondary immune response (Domínguez-Andrés et al., 2019). The ability of specific stimuli to generate trained immunity and revert disease-causing tolerance could lead to additional methods to modulate the immune system during or after infection.

Another disease of dysregulated immune memory involve patients with hyper IgD syndrome (HIDS), an inborn error of metabolism where mevalonate kinase deficiency leads to accumulation of mevalonate (Drenth et al., 1999). Monocytes and macrophages in these patients produce higher amounts of TNF, IL6, and IL1b, and anti-TNF and antii-IL1 therapies have only been partially effective (Mulders-Manders and Simon, 2015). The metabolite mevalonate was shown to be critical in beta-glucan- and oxLDL-induced trained immunity by driving the mTOR pathyway, activating the TCA cycle, and generating acetyl-CoA needed for

altered H3K27ac at inflammatory genes (Bekkering et al., 2018). The chronic trained immunity state of macrophages due to elevated mevalonate may be a cause for the sterile inflammatory phenotype seen in these patients, such as febrile attacks, arthritis, and skin lesions (Bekkering et al., 2018). Importantly, administration of statins blocked the mevalonate-cholesterol synthesis pathway, attenuating trained immunity and reducing inflammatory attacks (Simon et al., 2004).

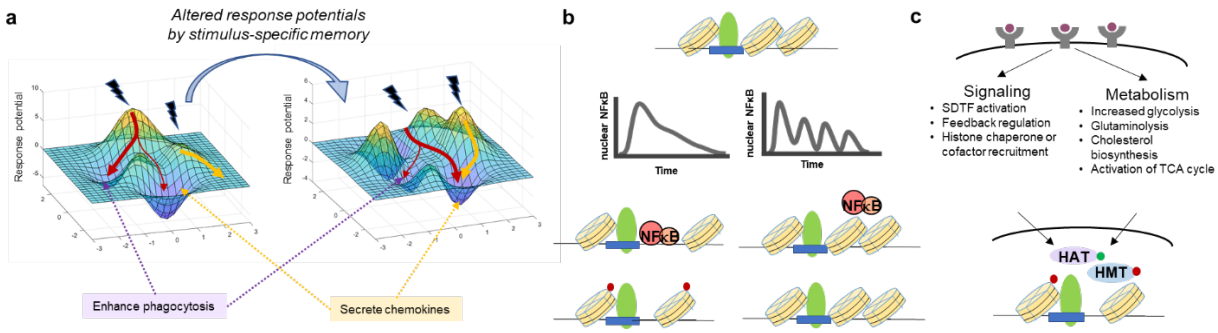


Figure 1.2. Establishment of stimulus-specific epigenetic memory by prior exposure. A) Stimulus-specific memory leaves more permanent changes to the epigenetic landscape, altering response potential after the initial stimulus is gone. B) Stimulus-specific non-oscillatory activity of signal-dependent transcription factors open chromatin in collaboration with cofactors and chromatin remodeling enzymes. C) Both signaling pathway activation and alterations to metabolic pathway activity are critical arms for generating innate immune memory.

OUTLOOK

For each of these hallmarks, addressing the outstanding questions will bring us closer to harnessing and controlling macrophage function either for diagnostics or for the treatment of human diseases.

How might response-specificity inform us about an individual's risk for inflammatory disease? Response-specificity, a property of macrophages that is a function of both cytokine context and immune memory, may prove a convenient metric for measuring health and disease states. In multiple immune diseases, noisy or ineffective recognition of an inflammatory threat lead to autoimmunity or faulty pathogen clearance. The functional health of the innate immune system, which is affected by context or prior exposures, could in the future be measured by perturbing monocytes isolated from peripheral blood, and profiling the resulting transcriptome or epigenome. The diagnosis or prognosis for a wide variety of diseases depends on immune system function, including autoimmune diseases, cancers, or neurodegenerative diseases, to name a few. It remains to be seen to what extent monocyte and macrophage response-specificity is reflective of risk or disease stage of each of these.

How might we control the programmability of innate immune responses for human health? A recent study reported that four-weeks of aerobic exercise prior to surgery created a lasting phenotype of immune tolerance in Kupffer cells, improving ischemia-reperfusion injury outcomes (Zhang et al., 2021). However, further study is needed to understand the physiological consequences of training innate immunity. For example, innate immune memory may play roles in post-COVID19 immunity or inflammatory sequelae. A recent study on convalescing COVID19 patients indicated altered monocyte subsets after COVID, with increased chromatin accessibility at inflammatory genes in patients covering from COVID19, suggestive of trained immunity. CD14+ and CD16+ monocytes from convalescing patients maintained epigenetic modification and had increased IL6 and IL-1beta production on subsequent stimulation with spike-nCoV pseudovirus (You et al., 2021). It remains to be determined whether this trained immunity results in an effect similar to vaccination, protecting the individual from subsequent infection, or whether the hyperinflammatory responses predisposes individuals to syndromes of long-COVID.

In this thesis, we 1) characterize the functional hallmark of response-specificity in health and disease, and 2) identify and mathematically model the mechanisms that enable the hallmark of stimulus-specific epigenomic programmability. These studies thus address a few open questions in the field, and excitingly, open avenues to raise new lines of inquiry.

References

Adelaja, A., Taylor, B., Sheu, K.M., Liu, Y., Luecke, S., and Hoffmann, A. (2021). Six distinct NFκB signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. Immunity *54*, 916-930.e7.

Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen Recognition and Innate Immunity. Cell *124*, 783–801.

Akula, M.K., Shi, M., Jiang, Z., Foster, C.E., Miao, D., Li, A.S., Zhang, X., Gavin, R.M., Forde, S.D., Germain, G., et al. (2016). Control of the innate immune response by the mevalonate pathway. Nat. Immunol. *17*, 922–929.

Aletaha, D., and Smolen, J.S. (2018). Diagnosis and Management of Rheumatoid Arthritis: A Review. JAMA *320*, 1360–1372.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. *17*, 487–500.

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-kappaB binds to the nucleosome. J. Biol. Chem. *279*, 42374–42382.

Arts, R.J.W., Carvalho, A., La Rocca, C., Palma, C., Rodrigues, F., Silvestre, R., Kleinnijenhuis, J., Lachmandas, E., Gonçalves, L.G., Belinha, A., et al. (2016a). Immunometabolic Pathways in BCG-Induced Trained Immunity. Cell Rep. *17*, 2562–2571.

Arts, R.J.W., Novakovic, B., Ter Horst, R., Carvalho, A., Bekkering, S., Lachmandas, E., Rodrigues, F., Silvestre, R., Cheng, S.-C., Wang, S.-Y., et al. (2016b). Glutaminolysis and Fumarate Accumulation Integrate Immunometabolic and Epigenetic Programs in Trained Immunity. Cell Metab. *24*, 807–819.

Arts, R.J.W., Moorlag, S.J.C.F.M., Novakovic, B., Li, Y., Wang, S.-Y., Oosting, M., Kumar, V., Xavier, R.J., Wijmenga, C., Joosten, L.A.B., et al. (2018). BCG Vaccination Protects against Experimental Viral Infection in Humans through the Induction of Cytokines Associated with Trained Immunity. Cell Host Microbe *23*, 89-100.e5.

Baardman, J., Licht, I., de Winther, M.P.J., and Van den Bossche, J. (2015). Metabolic-epigenetic crosstalk in macrophage activation. Epigenomics *7*, 1155–1164.

Banchereau, J., and Pascual, V. (2006). Type I interferon in systemic lupus erythematosus and other autoimmune diseases. Immunity *25*, 383–392.

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014). Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers. Mol. Cell *54*, 844–857.

Bekkering, S., Arts, R.J.W., Novakovic, B., Kourtzelis, I., van der Heijden, C.D.C.C., Li, Y., Popa, C.D., Ter Horst, R., van Tuijl, J., Netea-Maier, R.T., et al. (2018). Metabolic Induction of Trained Immunity through the Mevalonate Pathway. Cell *172*, 135-146.e9.

Beutler, B.A. (2009). TLRs and innate immunity. Blood *113*, 1399–1407.

Bonelli, M., Dalwigk, K., Platzer, A., Olmos Calvo, I., Hayer, S., Niederreiter, B., Holinka, J., Sevelda, F., Pap, T., Steiner, G., et al. (2019). IRF1 is critical for the TNF-driven interferon response in rheumatoid fibroblast-like synoviocytes. Exp. Mol. Med. *51*, 1–11.

Brahma, S., and Henikoff, S. (2020). Epigenome Regulation by Dynamic Nucleosome Unwrapping. Trends Biochem. Sci. *45*, 13–26.

Buch, M.H., Eyre, S., and McGonagle, D. (2021). Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. Nat. Rev. Rheumatol. *17*, 17–33.

Calmette, A. (1931). Preventive Vaccination Against Tuberculosis with BCG. Proc. R. Soc. Med. *24*, 1481–1490.

Cheng, C.S., Behar, M.S., Suryawanshi, G.W., Feldman, K.E., Spreafico, R., and Hoffmann, A. (2017). Iterative Modeling Reveals Evidence of Sequential Transcriptional Control Mechanisms. Cell Syst. *4*, 330-343.e5.

Cheng, Q., Behzadi, F., Sen, S., Ohta, S., Spreafico, R., Teles, R., Modlin, R.L., and Hoffmann, A. (2019). Sequential conditioning-stimulation reveals distinct gene- and stimulus-specific effects of Type I and II IFN on human macrophage functions. Sci. Rep. *9*, 5288.

Cheng, Q.J., Ohta, S., Sheu, K.M., Spreafico, R., Adelaja, A., Taylor, B., and Hoffmann, A. (2021). NF-κB dynamics determine the stimulus specificity of epigenomic reprogramming in macrophages. Science *372*, 1349–1353.

Cheng, S.-C., Quintin, J., Cramer, R.A., Shepardson, K.M., Saeed, S., Kumar, V., Giamarellos-Bourboulis, E.J., Martens, J.H.A., Rao, N.A., Aghajanirefah, A., et al. (2014). mTOR- and HIF-1α-mediated aerobic glycolysis as metabolic basis for trained immunity. Science *345*, 1250684.

Cheng, Z., Taylor, B., Ourthiague, D.R., and Hoffmann, A. (2015). Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation. Sci. Signal. *8*, ra69–ra69.

Cheong, R., Rhee, A., Wang, C.J., Nemenman, I., and Levchenko, A. (2011). Information transduction capacity of noisy biochemical signaling networks. Science *334*, 354–358.

Chumakov, K., Avidan, M.S., Benn, C.S., Bertozzi, S.M., Blatt, L., Chang, A.Y., Jamison, D.T., Khader, S.A., Kottilil, S., Netea, M.G., et al. (2021). Old vaccines for new infections: Exploiting innate immunity to control COVID-19 and prevent future pandemics. Proc. Natl. Acad. Sci. *118*.

Cirovic, B., de Bree, L.C.J., Groh, L., Blok, B.A., Chan, J., van der Velden, W.J.F.M., Bremmers, M.E.J., van Crevel, R., Händler, K., Picelli, S., et al. (2020). BCG Vaccination in Humans Elicits Trained Immunity via the Hematopoietic Progenitor Compartment. Cell Host Microbe *28*, 322-334.e5.

Comoglio, F., Simonatto, M., Polletti, S., Liu, X., Smale, S.T., Barozzi, I., and Natoli, G. (2019). Dissection of acute stimulus-inducible nucleosome remodeling in mammalian cells. Genes Dev.

Covert, M.W., Leung, T.H., Gaston, J.E., and Baltimore, D. (2005). Achieving stability of lipopolysaccharide-induced NF-kappaB activation. Science *309*, 1854–1857.

Covián, C., Fernández-Fierro, A., Retamal-Díaz, A., Díaz, F.E., Vasquez, A.E., Lay, M.K., Riedel, C.A., González, P.A., Bueno, S.M., and Kalergis, A.M. (2019). BCG-Induced Cross-Protection and Development of Trained Immunity: Implication for Vaccine Design. Front. Immunol. *10*.

Crowley, T., Buckley, C.D., and Clark, A.R. (2018). Stroma: the forgotten cells of innate immune memory. Clin. Exp. Immunol. *193*, 24–36.

DeFelice, M.M., Clark, H.R., Hughey, J.J., Maayan, I., Kudo, T., Gutschow, M.V., Covert, M.W., and Regot, S. (2019). NF-κB signaling dynamics is controlled by a dose-sensing autoregulatory loop. Sci Signal *12*, eaau3568.

Dehne, N., Jung, M., Mertens, C., Mora, J., and Weigert, A. (2016). Macrophage Heterogeneity During Inflammation. In Compendium of Inflammatory Diseases, M.J. Parnham, ed. (Basel: Springer), pp. 865–874.

Domínguez-Andrés, J., Novakovic, B., Li, Y., Scicluna, B.P., Gresnigt, M.S., Arts, R.J.W., Oosting, M., Moorlag, S.J.C.F.M., Groh, L.A., Zwaag, J., et al. (2019). The Itaconate Pathway Is a Central Regulatory Node Linking Innate Immune Tolerance and Trained Immunity. Cell Metab. *29*, 211-220.e5.

Dowson, C., Simpson, N., Duffy, L., and O'Reilly, S. (2017). Innate Immunity in Systemic Sclerosis. Curr. Rheumatol. Rep. *19*, 2.

Drenth, J.P.H., Cuisset, L., Grateau, G., Vasseur, C., van de Velde-Visser, S.D., de Jong, J.G.N., Beckmann, J.S., van der Meer, J.W.M., and Delpech & contributing members of the International Hyper-IgD Study Group, M. (1999). Mutations in the gene encoding mevalonate kinase cause hyper-IgD and periodic fever syndrome. Nat. Genet. *22*, 178–181.

El Gazzar, M., Liu, T., Yoza, B.K., and McCall, C.E. (2010). Dynamic and selective nucleosome repositioning during endotoxin tolerance. J. Biol. Chem. *285*, 1259–1271.

Epelman, S., Lavine, K.J., and Randolph, G.J. (2014). Origin and Functions of Tissue Macrophages. Immunity *41*, 21–35.

Garcia, M.F., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H., and Zaret, K.S. (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. Mol. Cell *0*.

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.

Gordon, S., and Taylor, P.R. (2005). Monocyte and macrophage heterogeneity. Nat. Rev. Immunol. *5*, 953–964.

Gottschalk, R.A., Martins, A.J., Angermann, B.R., Dutta, B., Ng, C.E., Uderhardt, S., Tsang, J.S., Fraser, I.D.C., Meier-Schellersheim, M., and Germain, R.N. (2016). Distinct NF-κB and

MAPK Activation Thresholds Uncouple Steady-State Microbe Sensing from Anti-pathogen Inflammatory Responses. Cell Syst. *2*, 378–390.

Gottschalk, R.A., Dorrington, M.G., Dutta, B., Krauss, K.S., Martins, A.J., Uderhardt, S., Chan, W., Tsang, J.S., Torabi-Parizi, P., Fraser, I.D., et al. (2019). IFN-mediated negative feedback supports bacteria class-specific macrophage inflammatory responses. ELife *8*.

Hall, J.C., and Rosen, A. (2010). Type I interferons: crucial participants in disease amplification in autoimmunity. Nat. Rev. Rheumatol. *6*, 40–49.

Hamada, A., Torre, C., Drancourt, M., and Ghigo, E. (2019). Trained Immunity Carried by Non-immune Cells. Front. Microbiol. *9*.

Hayes, J.J., and Wolffe, A.P. (1992). The interaction of transcription factors with nucleosomal DNA. BioEssays *14*, 597–603.

Heinz, S., Haehnel, V., Karaghiosoff, M., Schwarzfischer, L., Müller, M., Krause, S.W., and Rehli, M. (2003). Species-specific Regulation of Toll-like Receptor 3 Genes in Men and Mice *. J. Biol. Chem. *278*, 21502–21509.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154.

Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science *298*, 1241–1245.

Hole, C.R., Wager, C.M.L., Castro-Lopez, N., Campuzano, A., Cai, H., Wozniak, K.L., Wang, Y., and Wormley, F.L. (2019). Induction of memory-like dendritic cell responses in vivo. Nat. Commun. *10*, 2955.

Ifrim, D.C., Quintin, J., Joosten, L.A.B., Jacobs, C., Jansen, T., Jacobs, L., Gow, N.A.R., Williams, D.L., van der Meer, J.W.M., and Netea, M.G. (2014). Trained Immunity or Tolerance: Opposing Functional Programs Induced in Human Monocytes after Engagement of Various Pattern Recognition Receptors. Clin. Vaccine Immunol. CVI *21*, 534–545.

Jeljeli, M., Riccio, L.G.C., Doridot, L., Chêne, C., Nicco, C., Chouzenoux, S., Deletang, Q., Allanore, Y., Kavian, N., and Batteux, F. (2019). Trained immunity modulates inflammation-induced fibrosis. Nat. Commun. *10*, 5670.

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., et al. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell *51*, 310–325.

Kamada, R., Yang, W., Zhang, Y., Patel, M.C., Yang, Y., Ouda, R., Dey, A., Wakabayashi, Y., Sakaguchi, K., Fujita, T., et al. (2018). Interferon stimulation creates chromatin marks and establishes transcriptional memory. Proc. Natl. Acad. Sci. *115*, E9162–E9171.

Kang, K., Park, S.H., Chen, J., Qiao, Y., Giannopoulou, E., Berg, K., Hanidu, A., Li, J., Nabozny, G., Kang, K., et al. (2017). Interferon-γ Represses M2 Gene Expression in Human

Macrophages by Disassembling Enhancers Bound by the Transcription Factor MAF. Immunity *47*, 235-250.e4.

Kaufmann, E., Sanz, J., Dunn, J.L., Khan, N., Mendonça, L.E., Pacis, A., Tzelepis, F., Pernet, E., Dumaine, A., Grenier, J.-C., et al. (2018). BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. Cell *172*, 176-190.e19.

Kawai, T., and Akira, S. (2011). Toll-like Receptors and Their Crosstalk with Other Innate Receptors in Infection and Immunity. Immunity *34*, 637–650.

Kok, F., Rosenblatt, M., Teusel, M., Nizharadze, T., Gonçalves Magalhães, V., Dächert, C., Maiwald, T., Vlasov, A., Wäsch, M., Tyufekchieva, S., et al. (2020). Disentangling molecular mechanisms regulating sensitization of interferon alpha signal transduction. Mol. Syst. Biol. *16*, e8955.

Lane, K., Van Valen, D., DeFelice, M.M., Macklin, D.N., Kudo, T., Jaimovich, A., Carr, A., Meyer, T., Pe'er, D., Boutet, S.C., et al. (2017). Measuring Signaling and RNA-Seq in the Same Cell Links Gene Expression to Dynamic Patterns of NF-κB Activation. Cell Syst. *4*, 458-469.e5.

Lane, K., Andres-Terre, M., Kudo, T., Monack, D.M., and Covert, M.W. (2019). Escalating Threat Levels of Bacterial Infection Can Be Discriminated by Distinct MAPK and NF-κB Signaling Dynamics in Single Host Cells. Cell Syst. *8*, 183-196.e4.

Lee, R.E., Walker, S.R., Savery, K., Frank, D.A., and Gaudet, S. (2014). Fold change of nuclear NF-kappaB determines TNF-induced transcription in single cells. Mol. Cell *53*, 867–879.

Li, G., and Widom, J. (2004). Nucleosomes facilitate their own invasion. Nat. Struct. Mol. Biol. *11*, 763–769.

Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. Nat. Struct. Mol. Biol. *12*, 46–53.

Liu, P.-S., Wang, H., Li, X., Chao, T., Teav, T., Christen, S., Di Conza, G., Cheng, W.-C., Chou, C.-H., Vavakova, M., et al. (2017). α-ketoglutarate orchestrates macrophage activation through metabolic and epigenetic reprogramming. Nat. Immunol. *18*, 985–994.

Liu, Y., Zhou, K., Zhang, N., Wei, H., Tan, Y.Z., Zhang, Z., Carragher, B., Potter, C.S., D'Arcy, S., and Luger, K. (2020). FACT caught in the act of manipulating the nucleosome. Nature *577*, 426–431.

Logie, C., and Stunnenberg, H.G. (2016). Epigenetic memory: A macrophage perspective. Semin. Immunol. *28*, 359–367.

Lone, I.N., Shukla, M.S., Charles Richard, J.L., Peshev, Z.Y., Dimitrov, S., and Angelov, D. (2013). Binding of NF-κB to Nucleosomes: Effect of Translational Positioning, Nucleosome Remodeling and Linker Histone H1. PLoS Genet. *9*, e1003830.

Lorch, Y., Zhang, M., and Kornberg, R.D. (2001). RSC Unravels the Nucleosome. Mol. Cell *7*, 89–95.

Masserdotti, G., and Götz, M. (2020). A decade of questions about the fluidity of cell identity. Nature *578*, 522–524.

Mechnikov, I. (1908). The Nobel Prize in Physiology or Medicine 1908.

Mitroulis, I., Ruppova, K., Wang, B., Chen, L.-S., Grzybek, M., Grinenko, T., Eugster, A., Troullinaki, M., Palladini, A., Kourtzelis, I., et al. (2018). Modulation of Myelopoiesis Progenitors Is an Integral Component of Trained Immunity. Cell *172*, 147-161.e12.

Mogensen, T.H. (2009). Pathogen Recognition and Inflammatory Signaling in Innate Immune Defenses. Clin. Microbiol. Rev. *22*, 240–273.

Mulders-Manders, C.M., and Simon, A. (2015). Hyper-IgD syndrome/mevalonate kinase deficiency: what is new? Semin. Immunopathol. *37*, 371–376.

Murray, P.J., Allen, J.E., Biswas, S.K., Fisher, E.A., Gilroy, D.W., Goerdt, S., Gordon, S., Hamilton, J.A., Ivashkiv, L.B., Lawrence, T., et al. (2014). Macrophage Activation and Polarization: Nomenclature and Experimental Guidelines. Immunity *41*, 14–20.

Naik, S., Larsen, S.B., Gomez, N.C., Alaverdyan, K., Sendoel, A., Yuan, S., Polak, L., Kulukian, A., Chai, S., and Fuchs, E. (2017). Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. Nature *550*, 475–480.

Nau, G.J., Richmond, J.F.L., Schlesinger, A., Jennings, E.G., Lander, E.S., and Young, R.A. (2002). Human macrophage activation programs induced by bacterial pathogens. Proc. Natl. Acad. Sci. *99*, 1503–1508.

O'Neill, L.A.J., Kishton, R.J., and Rathmell, J. (2016). A guide to immunometabolism for immunologists. Nat. Rev. Immunol. *16*, 553–565.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell *152*, 157–171.

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

Ozinsky, A., Underhill, D.M., Fontenot, J.D., Hajjar, A.M., Smith, K.D., Wilson, C.B., Schroeder, L., and Aderem, A. (2000). The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between Toll-like receptors. Proc. Natl. Acad. Sci. *97*, 13766–13771.

Peignier, A., and Parker, D. (2021). Impact of Type I Interferons on Susceptibility to Bacterial Pathogens. Trends Microbiol.

Penkov, S., Mitroulis, I., Hajishengallis, G., and Chavakis, T. (2019). Immunometabolic Crosstalk: An Ancestral Principle of Trained Immunity? Trends Immunol. *40*, 1–11.

Pulendran, B., and Ahmed, R. (2006). Translating Innate Immunity into Immunological Memory: Implications for Vaccine Development. Cell *124*, 849–863.

Quintin, J., Saeed, S., Martens, J.H.A., Giamarellos-Bourboulis, E.J., Ifrim, D.C., Logie, C., Jacobs, L., Jansen, T., Kullberg, B.-J., Wijmenga, C., et al. (2012). Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. Cell Host Microbe *12*, 223–232.

Rehli, M. (2002). Of mice and men: species variations of Toll-like receptor expression. Trends Immunol. *23*, 375–378.

Rivera, A., Siracusa, M.C., Yap, G.S., and Gause, W.C. (2016). Innate cell communication kick-starts pathogen-specific immunity. Nat. Immunol. *17*, 356–363.

Saeed, S., Quintin, J., Kerstens, H.H.D., Rao, N.A., Aghajanirefah, A., Matarese, F., Cheng, S.-C., Ratter, J., Berentsen, K., van der Ent, M.A., et al. (2014). Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science *345*, 1251086.

Seeley, J.J., and Ghosh, S. (2017). Molecular mechanisms of innate memory and tolerance to LPS. J. Leukoc. Biol. *101*, 107–119.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. (2014). Accurate information transmission through dynamic biochemical signaling networks. Science *346*, 1370–1373.

Sen, S., Cheng, Z., Sheu, K.M., Chen, Y.H., and Hoffmann, A. (2020). Gene Regulatory Strategies that Decode the Duration of NFκB Dynamics Contribute to LPS- versus TNF-Specific Gene Expression. Cell Syst. *10*, 169-182.e5.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Sheu, K.M., Luecke, S., and Hoffmann, A. (2019). Stimulus-specificity in the responses of immune sentinel cells. Curr. Opin. Syst. Biol. *18*, 53–61.

Shih, V.F., Kearns, J.D., Basak, S., Savinova, O.V., Ghosh, G., and Hoffmann, A. (2009). Kinetic control of negative feedback regulators of NF-kappaB/RelA determines their pathogen- and cytokine-receptor signaling specificity. Proc. Natl. Acad. Sci. U. S. A. *106*, 9619–9624.

Simon, A., Drewe, E., van der Meer, J.W.M., Powell, R.J., Kelley, R.I., Stalenhoef, A.F.H., and Drenth, J.P.H. (2004). Simvastatin treatment for inflammatory attacks of the hyperimmunoglobulinemia D and periodic fever syndrome. Clin. Pharmacol. Ther. *75*, 476–483.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. Cell *161*, 555–568.

Steinman, L. (2014). Immunology of Relapse and Remission in Multiple Sclerosis. Annu. Rev. Immunol. *32*, 257–281.

Sung, M.-H., Li, N., Lao, Q., Gottschalk, R.A., Hager, G.L., and Fraser, I.D.C. (2014). Switching of the relative dominance between feedback mechanisms in lipopolysaccharide-induced NF-κB signaling. Sci. Signal. *7*, ra6.

Takeuchi, O., and Akira, S. (2010). Pattern Recognition Receptors and Inflammation. Cell *140*, 805–820.

Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell NF-κB dynamics reveal digital activation and analogue information processing. Nature *466*, 267–271.

Van den Bossche, J., O'Neill, L.A., and Menon, D. (2017). Macrophage Immunometabolism: Where Are We (Going)? Trends Immunol. *38*, 395–406.

Van Valen, D.A., Kudo, T., Lane, K.M., Macklin, D.N., Quach, N.T., DeFelice, M.M., Maayan, I., Tanouchi, Y., Ashley, E.A., and Covert, M.W. (2016). Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments. PLOS Comput. Biol. *12*, e1005177.

Vila-del Sol, V., Punzón, C., and Fresno, M. (2008). IFN-gamma-induced TNF-alpha expression is regulated by interferon regulatory factors 1 and 8 in mouse macrophages. J. Immunol. Baltim. Md 1950 *181*, 4461–4470.

Wager, C.M.L., Hole, C.R., Campuzano, A., Castro-Lopez, N., Cai, H., Dyke, M.C.C.V., Wozniak, K.L., Wang, Y., and Jr, F.L.W. (2018). IFN-γ immune priming of macrophages in vivo induces prolonged STAT1 binding and protection against Cryptococcus neoformans. PLOS Pathog. *14*, e1007358.

Wagner, F.R., Dienemann, C., Wang, H., Stützer, A., Tegunov, D., Urlaub, H., and Cramer, P. (2020). Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. Nature *579*, 448–451.

Werner, S.L., Barken, D., and Hoffmann, A. (2005). Stimulus specificity of gene expression programs determined by temporal control of IKK activity. Science *309*, 1857–1861.

Wohlfahrt, T., Rauber, S., Uebe, S., Luber, M., Soare, A., Ekici, A., Weber, S., Matei, A.-E., Chen, C.-W., Maier, C., et al. (2019). PU.1 controls fibroblast polarization and tissue fibrosis. Nature.

You, M., Chen, L., Zhang, D., Zhao, P., Chen, Z., Qin, E.-Q., Gao, Y., Davis, M.M., and Yang, P. (2021). Single-cell epigenomic landscape of peripheral immune cells reveals establishment of trained immunity in individuals convalescing from COVID-19. Nat. Cell Biol. *23*, 620–630.

Zaret, K.S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. Annu. Rev. Genet. *54*, 367–385.

Zhang, D.X., and Glass, C.K. (2013). Towards an understanding of cell-specific functions of signal-dependent transcription factors. J. Mol. Endocrinol. *51*, T37–T50.

Zhang, H., Chen, T., Ren, J., Xia, Y., Onuma, A., Wang, Y., He, J., Wu, J., Wang, H., Hamad, A., et al. (2021). Pre-operative exercise therapy triggers anti-inflammatory trained immunity of Kupffer cells through metabolic reprogramming. Nat. Metab. 1–16.

Zhang, Q., Gupta, S., Schipper, D.L., Kowalczyk, G.J., Mancini, A.E., Faeder, J.R., and Lee, R.E.C. (2017). NF-κB Dynamics Discriminate between TNF Doses in Single Cells. Cell Syst. *5*, 638-645.e5.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81.

CHAPTER 2

Response Specificity

2.1 Quantifying the Response-Specificity of macrophages by single cell RNA profiling to score the health of innate immune function

ABSTRACT

Immune sentinel macrophages initiate responses to immune threats using hundreds of immune response genes. Each immune threat demands a tailored immune response, but quantifying the capacity of macrophages to mount stimulus-response-specific gene expression responses requires single cell data that reveals response distributions. Here we measured single-cell transcriptomic profiles of macrophages responding to diverse bacterial, viral, and cytokine stimuli. Information theoretic analyses of the data uncovered distinct functional gene groups that differed in their response specificity index. Comparing each gene's response specificity to the information contained in NFκB signaling dynamics, we found that target genes associate with different subsets of the six known NFκB signaling codons. The response specificity of other genes was tunable by polarizing cytokines IFNγ and IL4, and peritoneal macrophages from old and obese mice showed diminished response specificity. These findings suggest that macrophage response-specificity quantified by single cell mRNA measurements after stimulation may be the basis for an innate immune health score.

INTRODUCTION

Macrophages reside in almost all tissues of the body, where they perform diverse functions depending on their location and differentiation states (Murray and Wynn, 2011). In one function central to innate immunity, macrophages are tasked with sensing the environment through only a few dozen available receptors and responding appropriately to hundreds of potential pathogens, danger-associated molecular patterns, and cytokines. As macrophages are

first responders to pathogens and cellular damage, the precise deployment of functions that either directly target the threat, or initiate the activation of other cell types, is critical for preventing abnormal immune sequelae.

Transcriptomic profiling of resting macrophages identified a few hundred genes that defined different macrophage types, including genes that regulate phagocytic function (Gautier et al., 2012). Other early transcriptome profiling studies focused on responses rather than steady state, and suggested macrophages had a common core response to different immune ligands (Nau et al., 2002). However, later studies instead highlighted that macrophage responses were in fact highly stimulus-specific (Cheng et al., 2017; Xue et al., 2014). Unlike adaptive immunity, this specificity arose from the signaling and epigenetic networks downstream of receptor-ligand interactions, rather than being generated genetically by recombination.

Because immune responses are costly and detrimental to the host, immune mechanisms counteracting a potential threat should be produced only to the extent necessary. Both weak or overactive immune responses may be a result of immune sentinel cells failing to produce responses appropriately specific to the stimulus (Alleva et al., 2000; Galani et al., 2021; McKechnie and Blish, 2020; Moore and June, 2020). The ability to measure robustness of this function *in vitro* may thus be important to quantifying susceptibility to inflammatory disease. Such quantification of response specificity requires single cell measurements, as immune sentinels are not organs, but rather operate as single cells that each sense the environment (Muldoon et al., 2020). Two stimuli generating highly distant average responses at the population level may also produce high single cell heterogeneity, resulting in overall poor distinguishability; in contrast, two stimuli with close average responses may also have very small variation in single cell responses, leaving little confusion as to which ligand was encountered. Single cell sequencing of pathogen responses in dendritic cells have pointed the

outlier cells (Shalek et al., 2014), suggesting that single cell macrophages may also exhibit extremes of responses that impact ligand identifiability.

Response specificity may also be a function of microenvironmental context. Signaling cross-talk from polarizing cytokines affects network wiring and the activity of key immune response transcription factors (Adelaja and Hoffmann, 2019; Piccolo et al., 2017), and both Type I and Type II interferons have been shown to enhance the activity of NFκB through distinct mechanisms (Mitchell et al., 2019). At the transcriptomic level, average responses are gene-specifically altered by prior conditioning through epigenetic changes (Cheng et al., 2019; Kang et al., 2017, 2019; Murray, 2017). Recent work has characterized the heterogeneity of macrophages polarized by inflammatory cytokines (Muñoz-Rojas et al., 2021). However, the influence of these conditioning cytokines on the heterogeneity of subsequent responses is unclear.

Here, we quantify the functional property of response specificity in macrophages by profiling single cell transcriptomes during immune responses to an array of immune ligands. Although average transcriptomic profiles are distinct, we observed unexpected amounts of overlap in response distributions between cytokine and viral ligands. Employing information theory, we identified genes and gene combinations that were most informative for relaying information about the ligands encountered. For NFκB target genes, each gene had a different propensity for decoding each of the NFκB signaling codons. We found that a small combination of genes can represent most of the information, and that response specificity was altered in specialized macrophages signaling from polarizing cytokine environments, and in aging and obesity.

RESULTS

Single cell RNAseq reveals heterogeneity of macrophage responses to diverse stimuli

To investigate the extent of stimulus-specificity in macrophage responses when they are exposed to different immune ligands, we developed an experimental workflow to collect single cell gene expression measurements (Fig. 1a) (Shum et al., 2019). We chose a targeted gene approach to reduce the technical variance due to high drop-out rates of genome-wide approaches. To identify a set of stimulus-specific genes and a set of stimuli to probe them with, we first analyzed bulk RNAseq data from macrophages responding to 14 different pathogen or cytokine ligands to determine the ligands that induce diverse macrophage responses (Fig. S1a) (Cheng et al., 2017). Using tensor components analysis (Kolda and Bader, 2009), we found that each ligand occupied a non-redundant location, indicating a distinct transcriptomic response for population means, though some ligands sat closely adjacent in the tensor decomposed space (Fig. S1b). From the 14 stimuli, we selected 6 as representing a spectrum of gram-positive (P3C) and gram-negative (LPS) bacteria, bacterial DNA (CpG), viral nucleic acids (PIC), and host cytokines activating either the interferon (IFNβ) or NFκB (TNF) pathways.

We used a PCA-based framework to identify a set of 500 stimulus-specific macrophage genes (Methods, Table S1, Fig. S1c), which were selectively amplified and sequenced using the BD Rhapsody platform (Fig 1a). While all stimulus-induced genes (1502) showed enrichment for NFκB, IRF, MAPK/AP1, ETS, and Zinc finger (Zf) motifs in their regulatory regions, the set of 500 genes showed less enrichment in ETS and Zf motifs and a more even enrichment in motifs associated with the three immune response signaling pathways (NFκB, IRF, MAPK), suggesting that they primarily drive response specificity (Fig. 1b). Gene set enrichment analysis to identify the top three gene ontology terms for selected vs unselected genes indicated similar biological functions, with further enrichment of migration, chemotaxis, and cell adhesion genes in the selected gene set (Fig. 1c).

To assess response specificity at different timepoints of stimulation, we calculated distances in pairwise stimulus comparisons at 1, 3, and 8hrs. For most pairs, transcriptomic responses at 3hrs were more distinct than at 1hr and similar to 8hrs (Fig. S1d). To minimize secondary cytokine effects, we chose the 3hr timepoint for single cell analysis. A heatmap comparison of the newly single cell data at the 3hr timepoint showed good concordance with the published bulk RNAseq data, while revealing the cell-to-cell heterogeneity in expression (Fig. 1d). Inspecting a few important cytokines (Ccl5, Tnf, Cxcl10) we found that multiple stimuli resulted in a Fano factor >1 (e.g. Cxcl10 PIC response), indicating a greater heterogeneity in gene expression than expected by a simple stochastic process (Fig. 1e).

Gene expression patterns from the Rhapsody platform paralleled those from 10x genomics (Fig. S2ab), with similar distributions in genes/cell and counts/gene to the 10x data subset to the 500 selected genes (Fig. S2c). The differences in counts/gene distributions in 10x vs 10x:500genes suggested that the genes outside the custom panel had low or zero (dropout) counts (Fig. S2c bottom). The Rhapsody platform resulted in a smaller percentage of genes with all 0 counts across cells (Fig. S2d). Furthermore, replicates of the single cell data for five stimuli were concordant in means (avg. Pearson's r = 0.86) (Fig. S3a) and distributions (Fig. S3b). Taken together, the experimental approach provided a reliable, cost-effective means for measuring heterogeneous single cell macrophage responses across diverse immune ligands, upon which we could quantify stimulus-response specificity.

Statistical classifiers quantify unexpected overlap in stimulus-response distributions

To assess the response distributions, we performed PCA of all gene expression data associated with each cell and found that IFNβ and LPS response distributions were best distinguished, with minimal overlap of their 95% confidence regions with other stimuli, although outlier cells were evident when examining the first two PCs (which capture 44% of the variance in the data)

(Fig. 2a). However, the other stimuli, PIC, TNF, CpG, and P3C, were overlapping. To determine whether these other stimuli could be separated on lower PCs, we calculated pairwise Bhattacharyya distances between stimulus pairs, which corroborated that CpG and P3C distributions were most similar, and IFNβ most distinct (Fig. 2b left). UMAP visualization of the top 20 PCs (capturing 61% of the variance of the data) (Fig. 2b right) revealed that IFNβ cells were perfectly distinguished, and LPS cells were well distinguished, though with a handful of CpG cells intermixed. On the other hand, CpG vs P3C, and TNF vs PIC distributions still could not be well separated.

To further quantify the extent of ligand identifiability, we trained a random forest classifier on a portion of the single cell gene expression data and tested the classifier on the held-out data. Based on its transcriptome, each cell in the test set was given a probability of having encountered a particular ligand (Fig. 2c left, Fig. S4a), and the highest probability ligand was assigned as the prediction (Fig. 2c right). Soft probability predictions helped distinguish weak versus strong assignments: for instance, correct IFNβ predictions consistently had prediction probabilities close to 1, whereas correct CpG predictions had prediction probabilities of ~0.6-0.7 with the second best choice being P3C. Overall, we found that IFNβ could be perfectly predicted (F1 score = 100%), while bacterial ligands CpG and P3CSK were more poorly classified (F1 = 85% and 74%, respectively) (Fig. 2d). Other machine learning models resulted in similar overall accuracies (Fig. S4b). Higher FPR and FDR were associated with pathogen ligands than with cytokine ligands (Fig. 2e), suggesting that a single cell's ability to consistently generate differential responses to IFNβ versus TNF is more physiologically important than doing so for LPS versus CpG.

Hierarchical clustering of the union of top 20 differentially induced genes for each stimulus vs all other stimuli showed that IFNβ-stimulated macrophages strongly expressed interferon response genes such as *Mx1* and *Ifit3* (as did LPS and PIC more heterogeneously),

but was primarily distinguished from other stimuli by consistent lack of expression of NFκB genes such as *Tnfaip3* (Fig. 2e). Other key immune response genes *Ccl5* and *Il6* were expressed uniformly in LPS-stimulated macrophages, but with much more variation for the other bacterial stimuli CpG and P3C. Inflammatory response genes *Socs3, Tnf, Il1a, Il1b* were expressed on average in response to all bacterial ligands, but *Il1a* and *Il1b* much more heterogeneously.

The hierarchical clustering of individual cells (Fig. 2e) suggested that although highly stimulus-specific and consistent transcriptomes were generated across single cells for some stimuli, other stimuli elicited much more overlapping response distributions. We may imagine that some stimuli may require physiologically similar responses, as for example responses to diverse bacterial ligands, whether they are intracellular or extracellular. Other observations are more surprising: we found that ~5-10% of TNF-responding macrophages induce interferon genes (resulting in a partial overlap with the PIC response), suggesting that it may be favorable that even this inflammatory cytokine elicits the production of type I interferon in a small fraction of macrophages, the innate immune cytokine thought to be restricted to responses to pathogen-associated molecular patterns.

Quantification of response-specificity for gene subsets and individual genes suggests a physiological importance in maintaining either specific or non-specific gene expression

To succinctly quantify response specificity across ligands, and for different subsets of genes, we employed an information theoretic approach (Cover and Thomas, 2012; Shannon, 1948). Using this approach, ligand information was considered as transmitted through a channel comprised of the signaling and epigenetic networks, both affected by molecular noise, to produce gene expression responses (Fig. 3a) (Suderman et al., 2017; Tkačik et al., 2008). The channel capacity describes the maximum mutual information (max MI) possible between input and output. For 2 ligands, 1 bit equates perfect distinguishability of the two response

distributions (Fig. 3a), and for 6 ligands, the maximum theoretical mutual information is 2.58 bits ($2^{2.58} = 6$).

To identify which stimuli were distinguished by key cytokine genes, we calculated the maximum mutual information associated with pairs of stimuli for three cytokines, Ccl5, Tnf, and Cxcl10 (Fig. 3b). With the exception of LPS vs TNF induced expression of Cxcl10, no two stimuli could be perfectly distinguished (max MI = 1.0) with just one of these cytokines. Furthermore, each of these genes distinguished different classes of immune threats. Ccl5 expression distribution best distinguished LPS from host cytokines (max MI: 0.8-0.9), while more moderately distinguishing LPS from other pathogen ligands (max MI: 0.4-0.65). On the other hand, Tnf performed very poorly at distinguishing any two bacterial ligands (max MI: 0-0.1), but separated very well bacterial from viral or cytokine ligands(max MI: 0.6-0.85). Though each gene played a distinct role in distinguishing stimuli, these three cytokine genes highlighted the importance of concerted gene programs to achieve an appropriate stimulus response: CpG-stimulated cells express Ccl5 very heterogeneously, causing severe loss of distinguishability from LPS (max MI = 0.48), but very uniformly show lack of expression of Cxcl10, thus allowing distinguishability of most CpG cells from LPS through Cxcl10 instead (max MI = 0.9) (Fig. 3b).

Examining individual genes, we found that when considering all 6 stimuli as input, 95% of the genes measured conveyed no more than 1 bit of information (Fig. 3c top, Table S2). The highest scoring single gene allowed for a maximum mutual information of 1.4 bits, or ~54% of the theoretical maximum (Fig. 3c bottom). No single gene had sufficient dynamic range or a sufficiently tight distribution to define a stimulus-specific response of more than 2-3 groups, suggesting the importance of concerted gene programs. We further investigated the relative contribution of ‘differences in means’ or ‘low heterogeneity’ to response-specificity (Fig. 3d), as genes with low information content due to high heterogeneity may indicate an advantageous

'bet-hedging' function (Chaturvedi et al., 2020; Dragotakes et al., 2020). We found that while mean squared deviation (MSD) correlated more strongly to max MI (r = 0.8) than average Fano factor (r = -0.5), genes such as cytokines Cxcl10, Ccl5, and the acute phase reactant Saa3 had lower max MI than expected based on their MSD. In contrast, metabolic and non-secreted genes such as Cmpk2 and Ifit3 had higher max MI than expected by the trend line, due to tight distributions (i.e. lower avg. Fano factor) (Fig. 3d). These genes illustrated a trend where cell-intrinsic functions had tighter stimulus-specific distributions, while secreted proteins tended to have more heterogenous overlapping distributions despite large differences in mean expression.

To further understand the information carrying capacity of gene combinations that represented important biological functions, we grouped small sets of genes into functional classes important for either immune system activation or cell fate decisions (Fig. 3e). Considering all 6 stimuli, proinflammatory and antiviral functions had relatively high response specificity, while anti-inflammatory functions were less stimulus-specific (Fig. 3e). High response specificity was attributable to certain pairs of stimuli: For pro-inflammatory genes, LPS vs PIC were well distinguished, but not PIC vs cytokines. In contrast, antiviral genes did distinguish PIC vs cytokines, but not LPS vs PIC. In an analysis of cell fate-related gene sets, cell death genes were more stimulus-specific than antioxidant genes, with cell death genes following a pairwise pattern similar to that of the proinflammatory gene set.

Information conveyed in NFκB signaling dynamics generates gene specific effects

Information about the ligand is conveyed through a signaling network comprised of four signal response pathways and decoded by the epigenetic gene regulatory network (Fig. 4a). We thus wondered how the information capacity of signaling dynamics compared to that of gene expression. While previous studies of NFκB dynamics measured only one transcription factor,

gene expression response-specificity is the product of the four pathways activated in stimulus-specific combinatorial and temporal patterns (Fig. 4b) (Adelaja et al., 2021; Cheng et al., 2017). Furthermore, for any single pathway, not all of the information in the signaling network may be ultimately reflected to gene expression.

To investigate how much of gene expression response specificity could be retained from NFκB signaling dynamics specificity, for each ligand pair, we calculated the information accumulation in the NFκB time series data, as well as the pairwise max MI in each gene (Fig. 4c-e). Max MI from NFκB signaling for every pair was consistently between 0.5-0.75 bits; however, max MI for expression was highly gene-specific. We assigned each gene to a gene regulatory mechanism based on a curation of prior literature (Table S3). For LPS vs TNF, genes in the IRF or "NFκB or IRF" clusters retained the most response specificity, as only LPS activates IRF signaling (Fig. 4c). Although LPS and TNF activate NFκB, with different dynamics, the corresponding max MI of most NFκB target genes was lower than the max MI of the NFκB time series. Examining TNF vs P3CSK, which both do not activate IRF, only three genes from the NFkB&p38 cluster had a max MI above 0.5 bits (Fig. 4d). These genes included *Nfkbiz*, important for the production of IL6 (Yamamoto et al., 2004); *Tnf*, whose half-life is controlled by p38 activity (Cheng et al., 2017); and the RNase *Zc3h12a*, responsible for the decay of IL6 mRNA (Matsushita et al., 2009). Comparing PIC to P3CSK also showed that most NFκB target genes utilized only a fraction of the information provided from signaling (Fig. 4e). Interestingly, IRF target-gene specificity was also lower in PIC vs P3CSK compared to TNF vs LPS, due to the high heterogeneity of PIC responses (Fig. 2f). An analysis of the remaining pairs of ligands revealed concordant patterns, with IRF gene specificity being greatest for LPS vs TNF, but lower for PIC vs TNF, or LPS vs PIC, due to the heterogeneity of PIC-induced IRF target gene expression (Fig. S5a-c). Notably, no NFκB genes had high

specificity in TNF vs PIC (Fig. S5b), potentially due to the more similar temporal dynamics of NFκB signaling in response to these two ligands.

To investigate which NFκB signaling codon may be associated with the response specificity of NFκB target gene expression, we analyzed transcriptomic data from an NFκB negative feedback mutant (Adelaja et al., 2021). This mutant shows diminished the stimulus-specificity of the NFκB signaling codon "oscillatory content". Using both control and mutant data in response to TNF, LPS, and PIC, we correlated the response specificity of NFκB signaling codons with the response specificity of NFκB target genes. Interestingly, we found that NFκB target genes could be grouped into several categories: those whose specificity tracked changes in Duration and Total Activity (e.g. *Fpr2, Il12b, Saa3, Ptges*), those whose specificity tracked changes in Speed, Peak Amplitude, and Early vs. Late activity (e.g. *Cd14, Nfkb1, Trim36*), and a small group associated with the Oscillatory content (e.g. *VcamI, IcamI, Icosl*) (Fig. 4f). These distinctions may be reflective of target genes regulated by non-linear dose responses mediated by an incoherent feedforward loop (Lee et al 2014), long mRNA half-life, slow chromatin opening steps (Sen et al., 2020), or requiring de novo enhancers (Cheng et al., 2021).

Response Specificity is a function of microenvironmental contexts

A hallmark of macrophages is that the microenvironment (e.g. the cytokine milieu) influences their function (Murray, 2017), and this context-dependent change in function can be reflective of both physiological and pathological states. We thus hypothesized that specialized macrophages such as those polarized by prior cytokine exposure might have a different capacity for stimulus-specific responses, and that quantifying this change may allow us to infer what microenvironments the macrophages are signaling from. We treated naive macrophages with IFNγ or IL4 for 24hrs to polarize them into M1-IFNγ and M2-IL4 states (Fig. S6a), which

canonically represent opposing ends of the macrophage polarization spectrum, and generated single cell gene expression for six ligands as before (Fig. 5a). At baseline, polarized macrophages all expressed macrophage marker *Adgre1* and did not exhibit differential NFκB pathway activation (Fig. S56c). Polarized single cell macrophages were enriched for the appropriate polarization markers (Fig. S6d-e).

PCA and UMAP projection of cells from all three states at 3hrs revealed that M1 and M2 macrophage response distributions were more overlapping compared to M0 (Fig. 5b, Fig. S6f). Calculation of the Bhattacharyya distance on each stimulus vs all others confirmed that M1 response distributions became less distinct upon polarization, especially for LPS vs TNF responses (Fig 5c). Interestingly however, pairwise comparisons of the percent change in distance compared to M0 revealed that for both M1 and M2 macrophages, PIC and TNF responses became slightly more distinct, as did CpG vs P3C (Fig. 5d), suggesting that both loss and gain of response specificity occurred in a stimulus-pair-dependent manner.

To compare the response specificity of individual genes across the macrophage types, we plotted the distribution of max MI values when considering all 6 stimuli. Max MI of M1 macrophage response genes formed a distribution with a lower mean, as illustrated by *Cxcl10*, which no longer carried stimulus-specific information about any pairs of stimuli (Fig. 5e). To investigate where stimulus-confusion occurred due to these differences in response-specificity, we tested the M1 and M2 single cells using the random forest classifier trained on M0 cells (Fig. S7a-b). The loss of specificity in genes like *Cxcl10* in M1 cells resulted in a greater proportion of CpG and P3C-responding cells being more LPS-like, and a higher fraction of TNF-responding cells being more PIC-like (Fig. S7a, Fig. 5f). In contrast, in M2 cells, the loss of specificity of IRF and NFκB target genes caused CpG and P3C responses to be more likely confused with TNF (Fig. S7b, Fig. 5g).

Motif analysis further supported that M1 cells had decreased response-specificity of IRF genes. M1 and M2 genes that both lost response specificity represented the biological functions 'response to virus' or 'response to IFNγ' (Table S4, Fig. S7c) and were enriched for IRF motifs (p-value <1e-10, p-value <1e-7, respectively) (Fig. 5h). NFκB motifs were second-highest enriched in genes that lost specificity in M2 cells (p-value < 0.001) but were, on the other hand, enriched in genes that gained specificity in M1 cells (p-value <1e-5). These observations together suggested that prior exposure to two different polarizing cytokines altered macrophage response-specificity distinctly, indicating that quantifying response specificities of macrophages could be a means for evaluating their microenvironmental contexts.

Development of an index for Response Specificity to assess innate immune health

Building on the observation that response-specificity was uniquely altered by cytokine context, we next wanted to understand whether a subset of the stimulus-specific genes in the M0 condition could be used to score the microenvironment. To identify the best combinations of genes that distinguish stimuli in the M0 condition, we used a forward selection approach to add one gene at a time to the mutual information analyses, starting from the most informative single gene (Methods). We found that the max MI increased rapidly for the first 5 genes, and then more incrementally as less informative genes were added (Fig. 6a). Doing the same on M1 and M2 cells revealed that M0 consistently had higher max MI across all dimensions tested (Table S5). For M0 macrophages, 15 genes carried ~2.2 bits of information (ability to distinguish ~4.6 ligands). These genes included a few cytokine genes, but also intracellular proteins controlling metabolism, cell death, and phagocytosis (Fig. 6b). This suggested that measuring response specificity of macrophages requires multiple genes but can be accomplished with a small number of genes.

Training the machine learning classifier for different subsets of genes, we found that just the top 15 genes identified by mutual information analyses could closely match the accuracy, false positive, and false discovery rates of a model using all genes (Fig. 6c-d, Fig. 2c). The 15 gene set biomarker identified produced lower False Discovery Rates than random sets of 15 genes drawn from the 500 gene panel (Fig. 6e), supporting the utility of information theoretic analyses to identify biomarker genes most important for Response Specificity.

To develop an score for response specificity, we calculated max MI for different sets of stimuli, using the 15 gene combination identified for M0 macrophages. The index that formed the score was comprised of stimulus subsets representing potential biological functions: ability to distinguish different bacterial ligands (LPS, P3C, CpG); different IRF activating stimuli (LPS, PIC, IFNβ); host from viral from bacterial ligand (TNF, PIC, LPS); different cytokines (IFNβ, TNF); and viral from host (TNF, PIC). Testing this index on M0, M1, and M2 macrophages, we found that M1 and M2 macrophages had an index 'fingerprint' distinct from that of M0 (Fig. 6f). Particularly, M1 macrophages had a lower score for distinguishing bacterial stimuli (LPS, P3C, CpG), as well as a lower score for distinguishing one each of bacterial, viral and cytokine stimuli (LPS, TNF, PIC). M2 macrophages scored higher in distinguishing TNF and PIC.

To make the method threshold free and remove bias toward M0 cells, we also performed PCA on all three conditions, now using all genes, and then calculated max MI on the PCA scores of top three components, called the Response Specificity Score (RSS). This second method, which weighted all macrophage types equally and allowed the PC loadings to dictate the gene weights (Table S6), achieved an index fingerprint similar to that of the top 15 M0 genes (Fig. 6g, Methods). We further collapsed the index to a single numerical score by summing the squared differences from M0 (Fig. 6g), which we called the delta Response Specificity Score (ΔRSS). We found that the overall response specificity M1 macrophages

scored slightly worse than M2 macrophages (higher ΔRSS), primarily due to greater loss of specificity in distinguishing bacterial stimuli, and in distinguishing bacterial, viral, and host cytokine ligands.

Decreased response specificity in old and obese mouse peritoneal macrophages

Next we tested whether the Response Specificity Score (RSS) might help reveal aberrations in macrophages derived from mice at risk of chronic inflammatory disease. We took peritoneal macrophages (PMs) from young and healthy mice, old mice, and obese mice, and performed the stimulus-specificity assay using five ligands: LPS, TNF, PIC, P3C, and IFNβ (Fig. 7a). tSNE visualization on the resulting single cell data suggested that old mice had aberrant IFNβ response distributions, while obese mice had aberrant PIC response distributions (Fig. 7b). We trained a random forest model on a subset of the healthy cells, and made predictions using the remaining healthy cells, and those from old and obese (Fig. 7c). The results supported that macrophages from aged mice had lower sensitivity in predicting IFNβ, while macrophages from obese mice had lower sensitivity in predicting PIC. We then calculated max MI for every individual gene given all stimuli and compared values from old and obese to healthy (Fig. 7d). The two inflammatory conditions were different from healthy in gene-specific ways. Both old and obese mice lost specificity in Tnf expression, as well as metabolic gene Acod1. Interestingly, a couple genes had high higher stimulus-specificity in either old or obese mice: for old, Cxcl10 and AW112010; and for obese, Cav1 and Slamf8, which play roles in macrophage differentiation and migration.

Computing the Bhattacharyya distance on ligand distributions, for each ligand vs all others (Fig. 7e), or for ligand pairs (Fig. 7f), we found that both old and obese mice macrophages lost specificity overall in IFNβ, LPS, and TNF responses (Fig. 7e). For TNF responses, however, old macrophages had an increase in distance between TNF and PIC

distributions compared to healthy (Fig. 7f). We applied the RSS to macrophages from healthy, old, and obese mice, for the stimuli available. We noted that the PMs from young healthy mice had a ΔRSS most similar to M0 bone marrow-derived macrophages, while high-fat diet mice PMs were most different (Fig. 7g). A closer inspection of the values in the index showed that PMs from old mice showed higher scores for stimuli subsets that included TNF and PIC, reflecting the greater separation between these two stimuli, paralleling M2 macrophages. In contrast, macrophages from high-fat diet obese mice had decreased scores across several stimuli groups tested, particularly due to a higher overlap of PIC and TNF responses.

Taken together, quantifying the response specificity of macrophages revealed that this hallmark immune sentinel function is affected not only by polarizing cytokines used in pre-conditioning regimes in vitro, but also by the microenvironments in situ that are evidently distinct in obese and old mice. The observed differences in Response Specificity suggest that quantifying post-stimulation single cell response distributions could be valuable for assessing the health of macrophage function.

DISCUSSION

As macrophages orchestrate the activation of downstream immune responses (Rivera et al., 2016), a critical function is to produce tailored, stimulus-appropriate responses when threatened by pathogens or host damage signals. Misregulation of immune responses, as seen in severe COVID-19 (Brodin, 2021; Moore and June, 2020; Schulte-Schrepping et al., 2020), may be predictable through assays that challenge the functional response of immune sentinels. Here, we assessed an emergent functional property of macrophages, Response Specificity, and developed an approach to quantify it. In doing so, we found that stimulus-response distributions held unexpected overlap, and that different genes and genesets were responsible for maintaining distinct responses to stimuli. By quantifying the overlap of distributions, we

ascribed a score to the function of macrophages derived from different physiological or pathological conditions.

Measurements of NFκB dynamics have provided an initial clue into the capacity of the signaling network in allowing macrophages to mount specific responses to bacterial, viral, or cytokine ligands (Adelaja et al., 2021). Indeed, alterations to the distributions of two signaling features (NFκB codons) decreased ligand identifiability in a mouse model of Sjogren's autoimmunity. Presumably for target genes controlled only by NFκB, the information received about the ligand will be equal to or less than the quantity contained in NFκB dynamics. However, gene expression ligand-specificity is a product of at least four signaling pathways acting in concert (Cheng et al., 2017). These other signaling pathway effectors AP1, IRF, MAPK-p38 interplay with NFκB dynamics to ultimately produce the gene products defining each cell's response.

Signaling studies also did not reveal what genes were regulated by different dynamic features. As a single cell's NFκB dynamics controls hundreds of target genes, gene-specific differences in response-specificity may further arise from epigenetic mechanisms, where each gene may only be able to interpret certain NFκB codons. For example, signaling duration has been shown to be decoded by either chromatin opening or mRNA half-life in MEFs (Sen et al., 2020), while genes with nucleosomes in their enhancers or promoters may be better equipped to decode presence of signaling activity oscillations (Cheng et al., 2020). In addition, genes are regulated by both temporal and combinatorial activity of signaling pathways, and gene expression Response Specificity reveals the outcome of all signaling pathways acting in unison. However, one limitation of this study here is that unlike measurements of NFκB dynamics, single cell transcriptomic profiles are by nature measured at snapshot timepoints. Thus, we scored Response Specificity on a single timepoint, and it remains still unclear how

gene expression temporal dynamics contribute to specificity, as has been studied for signaling dynamics (Selimkhanov et al., 2014).

Our analysis of macrophages from extracted from animal models of aging and obesity suggested that Response-Specificity, a measurement of stimulus-response distributions, could be a means to assess the presence of different inflammatory conditions. High body mass index (BMI) or advanced age in people has been correlated to worse outcomes in viral infections, for example, as seen with Sars-CoV2 (Tay et al., 2020). However, whether or not the propensity for outlier responses causally contributes to disease pathology remains unclear. One possibility is that the presence of a small number of outlier responses in a population of responding innate immune cells triggers improper activation of the downstream systemic immune response. Quantifying this functional hallmark of innate immunity may thus eventually become a useful clinical blood test for people at risk for inflammatory disease.

The Response Specificity Score leverages a statistical framework based on a survey of ligand-responses for *in vitro* macrophages states, namely naïve macrophages and macrophages from opposing ends of the macrophage polarization spectrum. Response Specificity of samples from new conditions are then compared to this landscape, and information theory is used to quantify the extent to which ligands can be distinguished given single cell transcriptomes. Either a higher or lower Response Specificity from normal healthy macrophages may be indicative of misregulation, hence the utility of the summary score ΔRSS. Interestingly, we saw that normal healthy mice macrophages had a score most similar to naïve macrophages, suggesting that the microenvironment milieu within old or obese mice caused a deviation from the naïve state in a parallel manner as *in vitro* conditioning with polarizing cytokines.

Despite its importance, there is currently no metric to measure the functional health of the innate immune system. Like blood pressure measurements, which are two numbers that measure risk for various cardiovascular-related diseases, an innate immune health index has

potential to assess risk for inflammatory diseases, including ability to clear an infection, but also for autoimmunity, cancer, or neurodegenerative disease, which all have an immunological component. However, unlike blood pressure measurements, which are taken at steady states, Response Specificity is a functional property that a requires a perturbation to immune cells, in order to assess whether they respond appropriately. As a genomic assay, its high dimensionality not only provides prognostic markers but may also help elucidate the underlying etiology of different diseases.

Importantly though, Response-Specificity may not be the only measurement that comprises an ultimate Immune Health Index. Other measurements, such as cell type composition or amounts of circulating cytokines in the bloodstream may be other informative measurements. However, we here show that Response Specificity is a quantifiable property of macrophages that changes with conditioning cytokines both *in vitro* and *in vivo*¸ and may contribute to our ability to measure innate immune health in the clinic.

ACKNOWLEDGEMENTS

We thank lab members Ying Tang and Diane Lefaudeux for helpful comments, and Supriya Sen and Carolina Chavez for reagents and experimental and analysis assistance. KMS was supported by the UCLA Medical Scientist Training Program (NIH NIGMS T32-GM008042) and Systems and Integrative Biology Training Grant (T32-GM008185). AH was funded by R01AI127864. Sequencing data from this paper was performed with the services of the UCLA Technology Center for Genomics and Bioinformatics Sequencing Core.

AUTHOR CONTRIBUTIONS

KMS performed the experiments and conducted the analyses. AG helped with piloting the experimental approach and assisted with the machine learning analysis. KMS and AH wrote the paper. All authors reviewed the manuscript. AH funded the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURES AND FIGURE LEGENDS

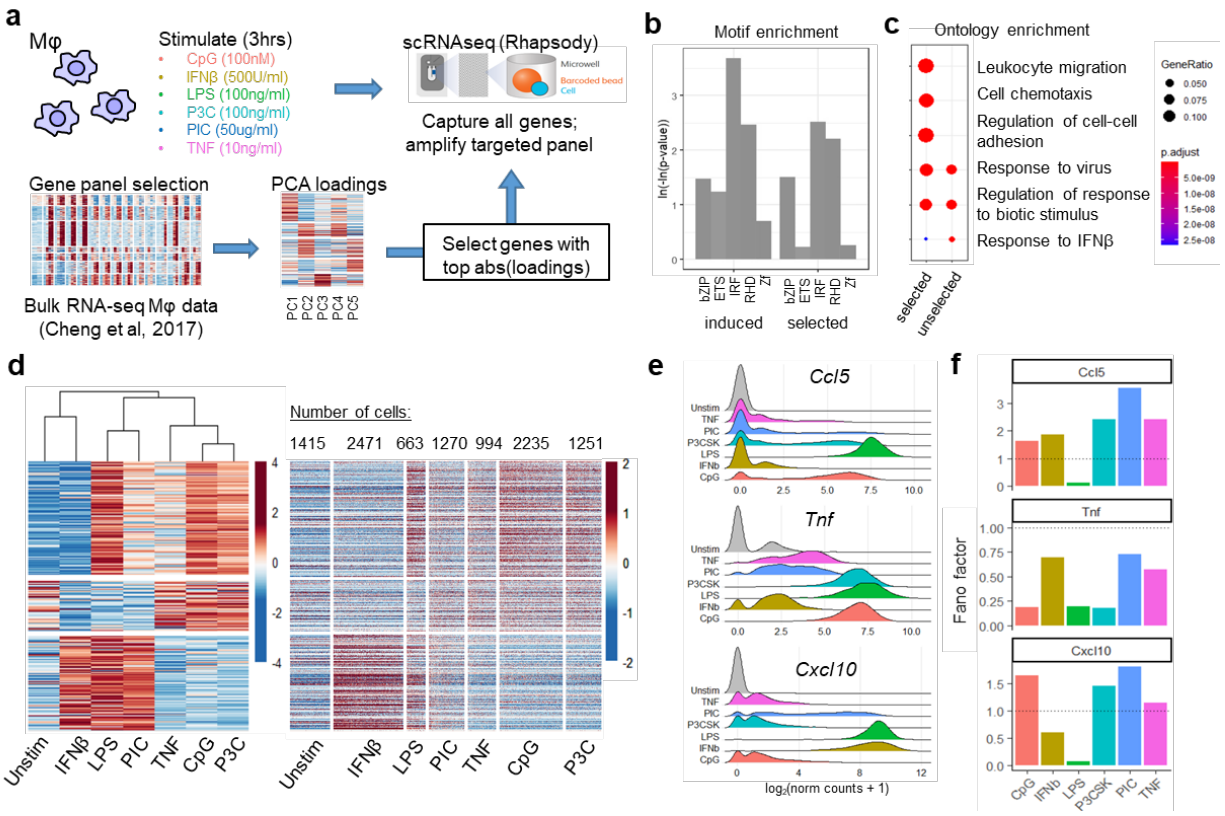


Figure 2.1. Single cell RNAseq reveals the heterogeneity of macrophage responses to diverse stimuli.

A) Experimental strategy for profiling the response specificity of macrophages. **B)** Enriched motifs in all induced genes vs the 500 selected. **C)** Top 3 gene ontology terms for selected vs unselected genes. **D)** K-means clustered heatmap of bulk RNA-seq gene expression patterns versus single cell RNAseq at 3hrs, for the same stimuli. **E)** Three genes with stimulus-specific gene expression and different amounts of single cell heterogeneity at the 3hr timepoint. **F)** Fano factor (variance/mean) for each stimulus for the three genes.

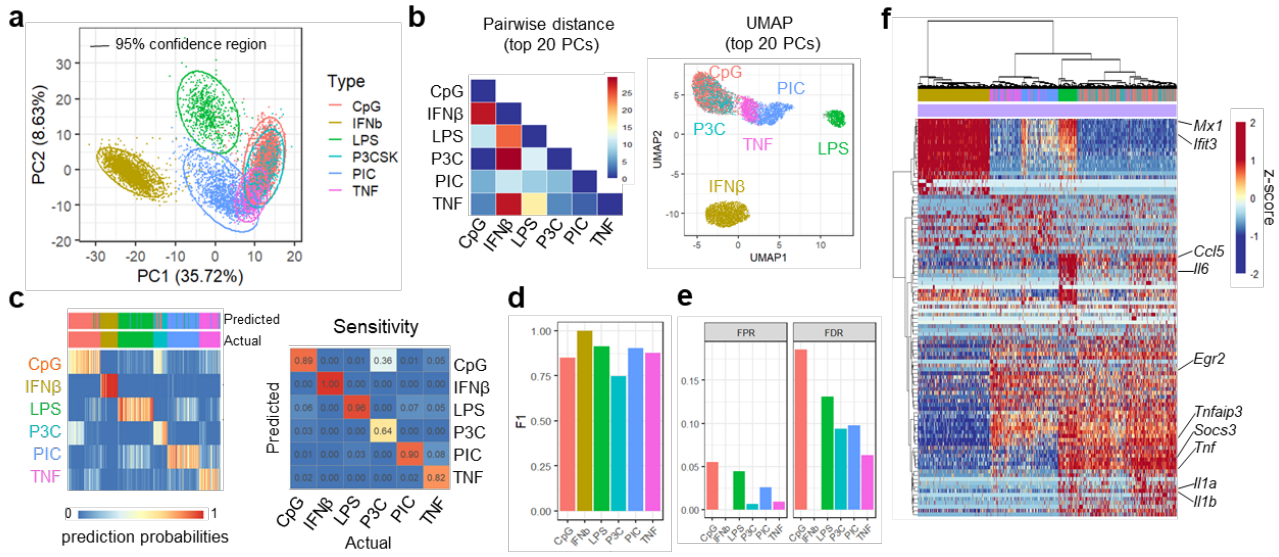


Figure 2.2. Statistical classifiers quantify unexpected overlap in stimulus-response distributions.

A) PCA of single cell transcriptomic responses. Ellipses represent 95% confidence regions. **B)** Pairwise Bhattacharyya distances between multivariate probability distributions, using the top 20 PCs, and corresponding UMAP visualization. **C)** Random forest statistical classifier trained on 70% of single cells and tested on 30% of held out data. Left: Prediction probabilities for each single cell. Right: Confusion matrix of the final predictions. Machine learning was performed on the two replicate datasets combined. **D)** F1 score (a measure of accuracy; harmonic mean of sensitivity and positive predictive value). **E)** FDR (false discovery rate) and FPR (false positive rate) for each stimulus. **F)** Heatmap of top 20 differentially induced genes (based on average log2FC) that distinguishes each stimulus.

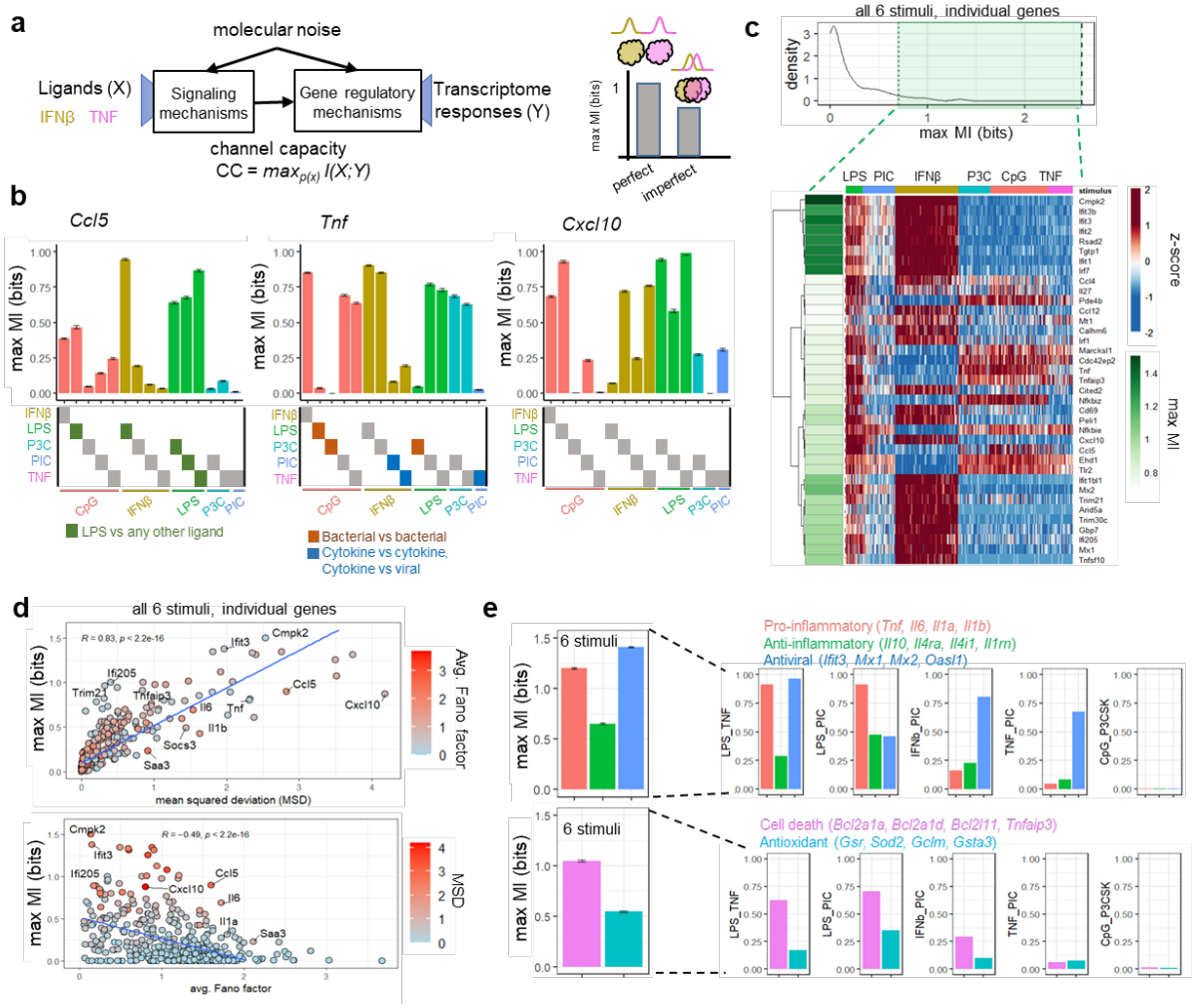
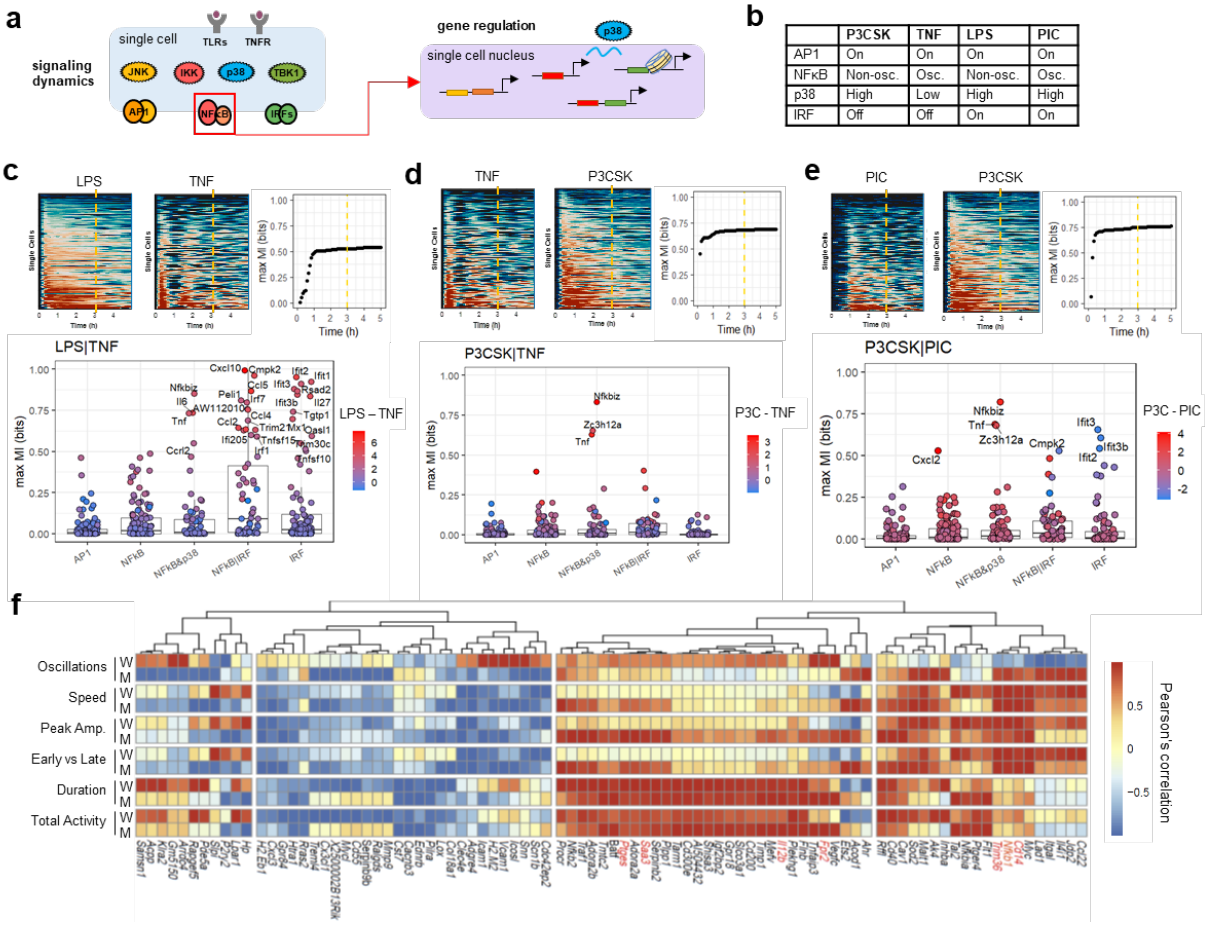


Figure 2.3. Quantification of response-specificity for gene subsets and individual genes suggests a physiological importance in maintaining either specific or non-specific gene expression.

A) Schematic of the biological channel that may be able to distinguish stimuli to produce stimulus-specific transcriptomes. **B)** Max MI for every pairwise combination of 2 stimuli for three cytokine genes, *Ccl5, Tnf, Cxcl10*. **C)** Max MI for individual genes but considering all 6 stimuli, with the heatmap of single cell expression values juxtaposed. Genes shown have a max MI >0.7, and theoretical maximum for 6 stimuli is log(6)/log(2)=2.6. **D)** Correlation of mean squared deviation and average Fano Factor to max MI. **E)** Different biological functions show distinct response specificities. (top) Cell intrinsic antiviral (*Ifit3, Mx1, Mx2, Oasl1*) and systemic immune activating pro-inflammatory (*Tnf, Il6, Il1a, Il1b*) categories both have the highest stimulus-specificity. (bottom) Max MI of cell death (*Bcl2a1a, Bcl2ald, Bcl2l11, Tnfaip3*) and antioxidant genes (*Gsr, Sod2, Gclm, Gsta3*) for select pairwise stimuli.



	P3CSK	TNF	LPS	PIC
AP1	On	On	On	On
NFκB	Non-osc.	Osc.	Non-osc.	Osc.
p38	High	Low	High	High
IRF	Off	Off	On	On

Figure 2.4. Information conveyed in NFκB signaling dynamics have gene specific effects.

A) Schematic of signaling and gene regulatory systems that control response specificity. Red box: Stimulus-specificity of NFκB signaling activity has previously been measured. **B)** Summary of known IRF3, AP1, and NFκB signaling profiles in response to four stimuli. **C)** Max MI in NFκB signaling dynamics vs max MI in expression of single genes, for LPS vs TNF, **D)** TNF vs P3CSK, **E)** PIC vs P3CSK. **F)** Correlation of signaling codon specificity to gene expression specificity at 8hrs in WT and NFκB feedback mutant macrophages, for NFκB target genes with at least one correlation p-value < 0.25.

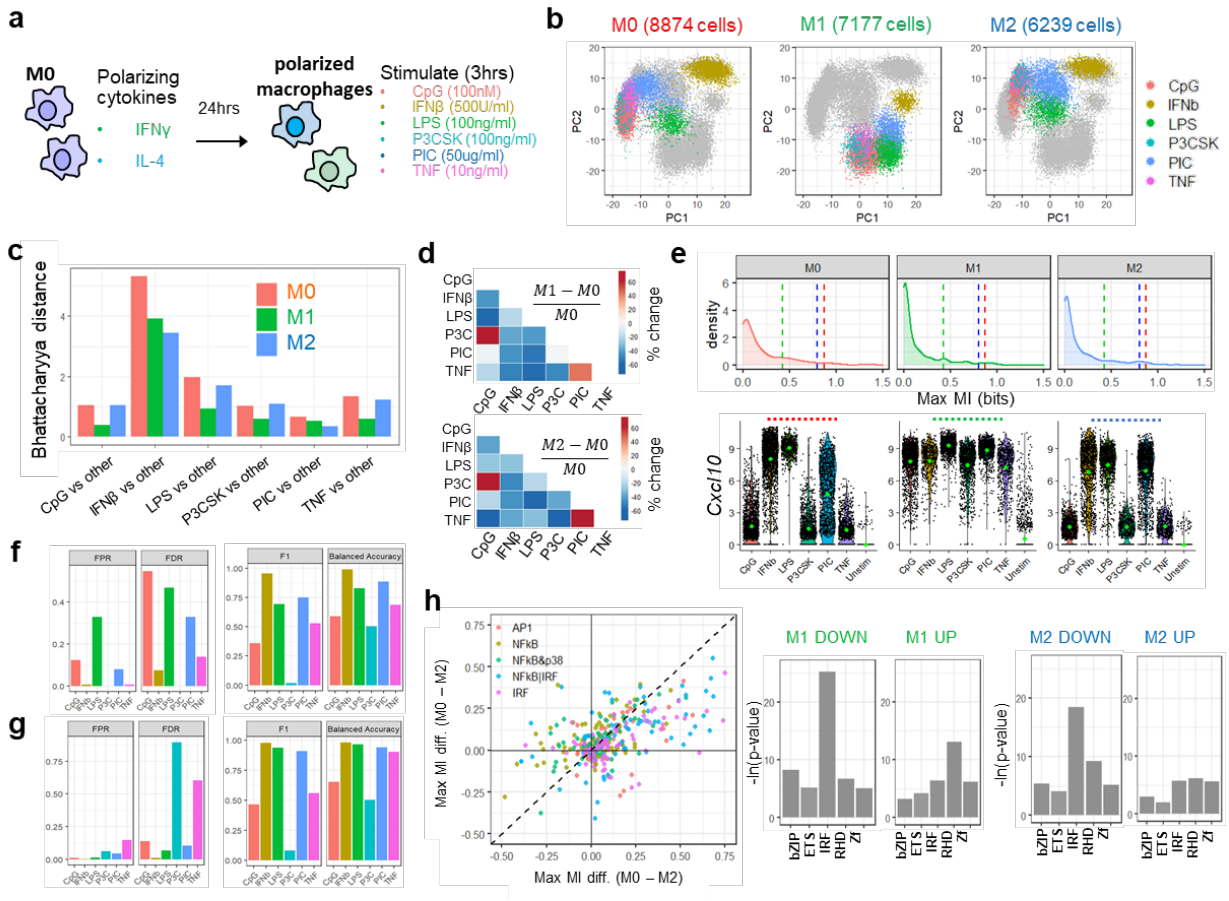


Figure 2.5. Macrophages exposed to different microenvironment cytokines lose transcriptomic capacity for response-specificity in different ways.

A) Single cell RNAseq for multiple stimuli in M1-IFNγ and M2-IL4 polarized macrophages. B) PCA of M0, M1, M2 macrophages at 3hrs. PC1 (19.9%), PC2 (15.7%). **C)** Bhattacharyya distance for M0, M1, M2 macrophages using top 20 PCs for each ligand vs others. D) Comparison of pairwise % change in distance between stimuli distributions after M1 or M2 polarization. **E)** Distribution of max MI values of individual genes when considering all 6 stimuli. Marked with a dotted line is the channel capacity quantification of *Cxcl10*. Log_2(normalized counts+1) expression values for M0 (left), M1 (center) and M2 (right) states are shown. **F-G)** FPR, FDR, and overall F1 score and balanced accuracy for **F)** M1, and **G)** M2 macrophages tested on the M0 machine learning model. **H)** Scatterplot of difference in channel capacity values between M0 and M1/M2 macrophages. Points are colored by assigned gene regulatory strategy. Motif enrichment for (M0-M1) or (M0-M2) channel capacity difference for individual genes.

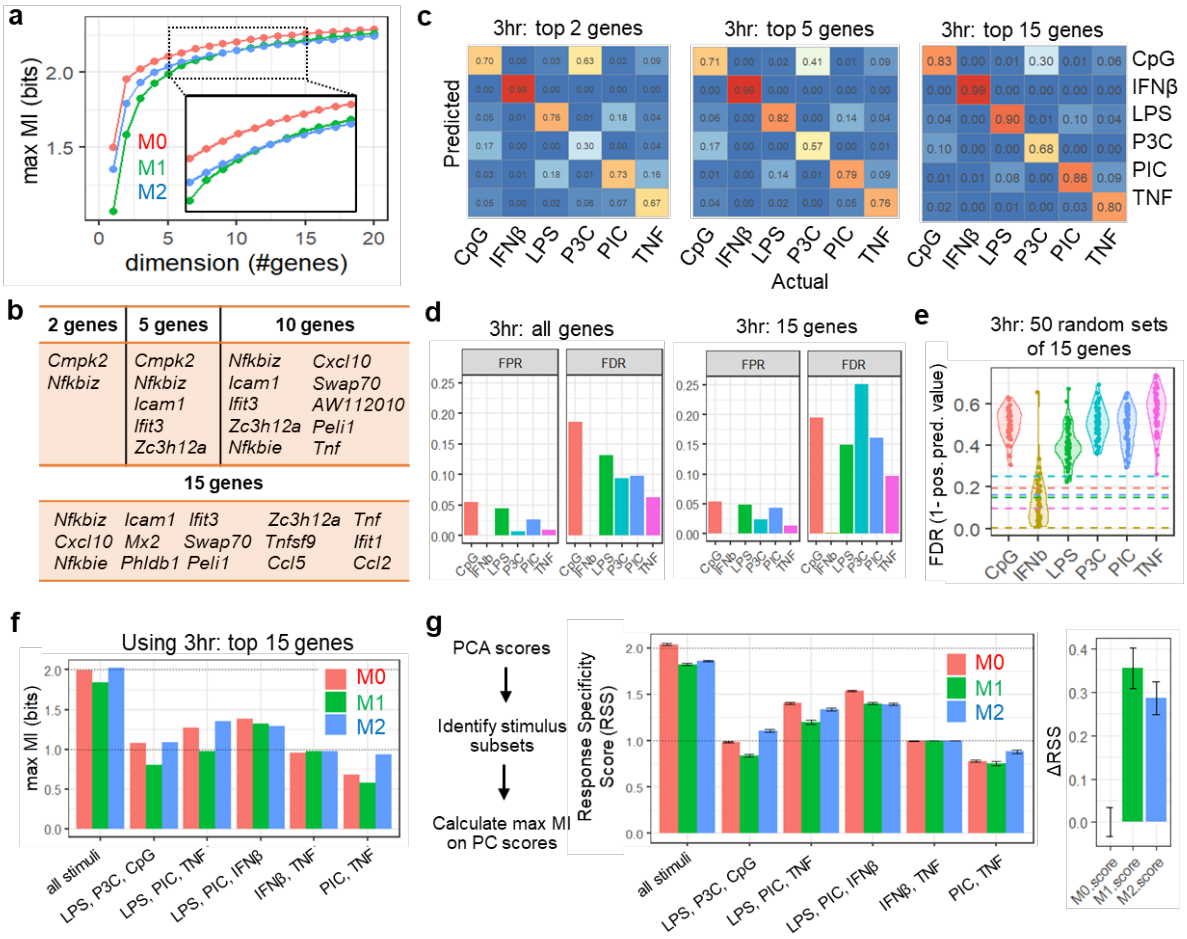


Figure 2.6. Development of a scoring metric for response specificity.

A) Gain in channel capacity as additional numbers of informative genes are added, for each macrophage type. **B)** Table of genes in each dimension for M0. **C)** Confusion matrices for random forest models trained on different numbers of genes. Trained on 70% of the M0 data as done before. **D)** False positive rate (FPR) and false discovery rate (FDR) when using all genes, or best 15 gene combination by max MI. **E)** Distribution of FDRs when using random sets of 15 genes at 3hrs. Dotted lines represent the FDR for the top 15 genes. **F)** Max MI using the top 15 genes at 3hrs for biologically categorized subsets of stimuli. **G)** Channel capacity calculated on the first 3 PCs of M0, M1, M2 macrophages for subsets of stimuli form the index (Response Specificity Index). Sum of the squared deviation from M0 across all stimulus subsets displayed was calculated to obtain the single-value Response Specificity Score. Higher numbers indicate a larger deviation in Response Specificity from M0.

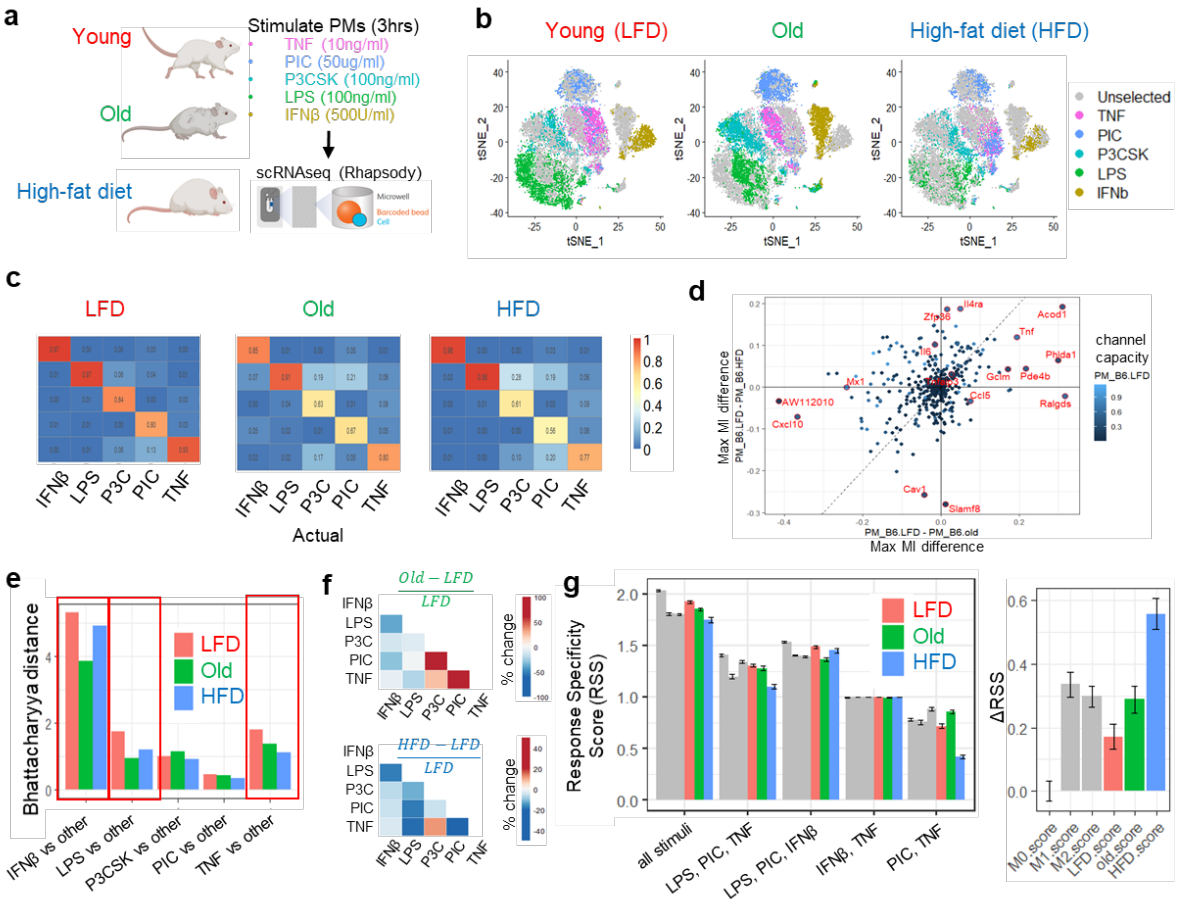


Figure 2.7: Decreased response specificity in old or obese mouse peritoneal macrophages.

A) Schematic of the stimuli tested. **B)** tSNE on mouse peritoneal macrophages: healthy (16wks), old (>90wks), and high-fat diet (16wks). **C)** Random forest model trained on 70% LFD macrophages, using the top 15 stimulus-specific genes. Old and HFD cells were tested on the LFD model. **D)** Channel capacity comparison of individual genes for (Old - LFD) and (HFD - LFD). **E)** Bhattacharyya distance of each stimuli vs all other stimuli, on 3 projected PCs. Red boxes highlight stimuli with greatest differences. **F)** Difference from LFD for pairwise distances between multivariate distributions, using the top 20 PCs. **G)** Scoring of disease models by channel capacity using data projected onto the M0/M1/M2 PCA framework for different stimulus subsets to obtain the index (Response Specificity Index). Subsets that included CpG were not included in the calculation. Sum of the squared deviation from M0 across all stimulus subsets was calculated to obtain the single-value Response Specificity Score.

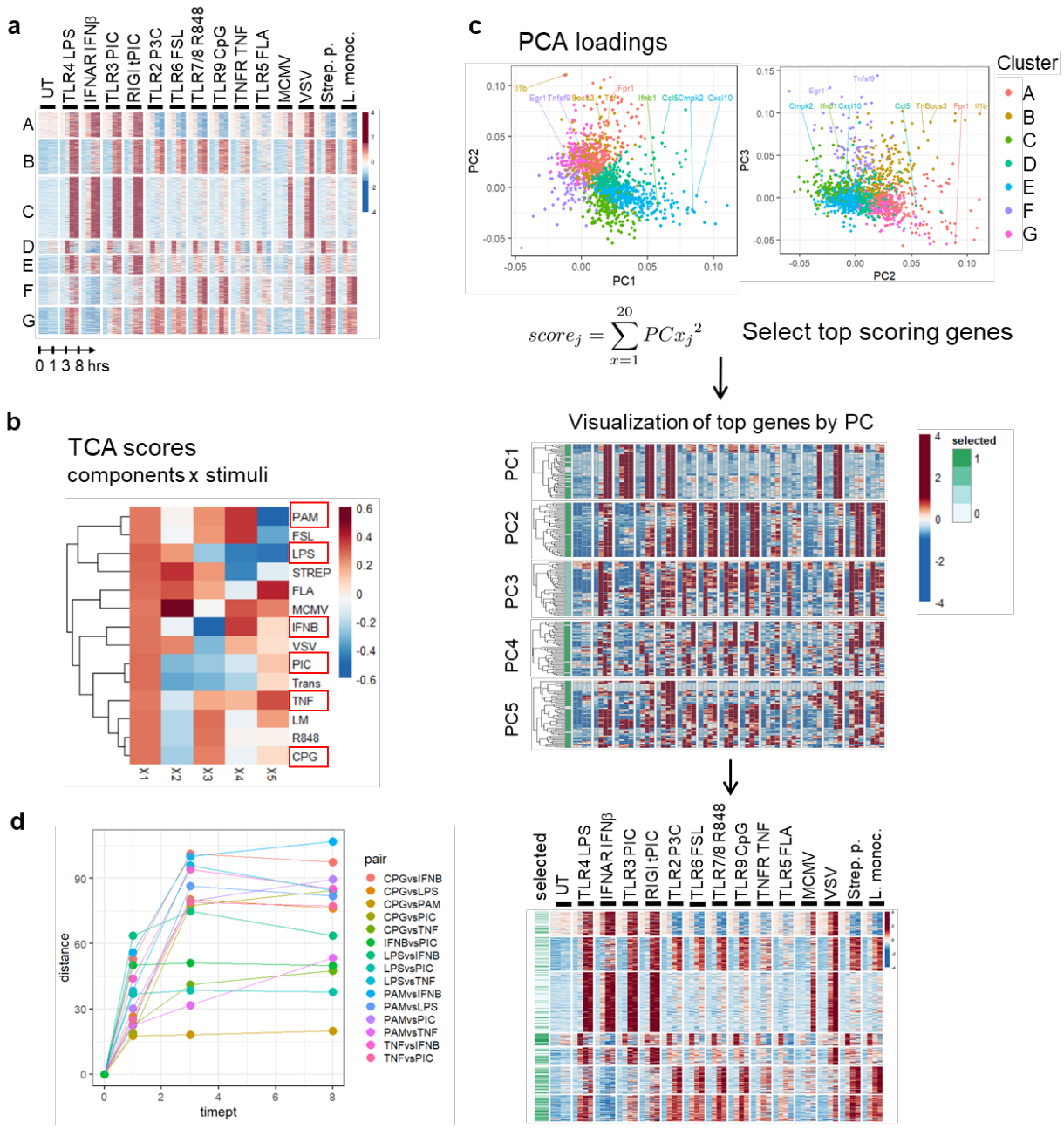


Figure S2.1. Selection of stimuli and gene panel.

A) Heatmap of all genes induced with $\log_2(FC)>2$ in any stimulus from bulk RNAseq data (Cheng et al, 2017). **B)** Tensor components analysis on all stimuli, all timepoints. Red boxes highlight stimuli chosen. **C)** PCA loadings from bulk RNAseq data. Genes were selected by scoring each gene based on loadings from the top 20 PCs. Visualization of genes with top loadings from each PC supports that each PC highlights gene sets with different stimulus-specific expression patterns. Selected genes are marked on the full heatmap. **D)** PCA performed on all data points together. Pairwise Euclidean distances between points in PCA space.

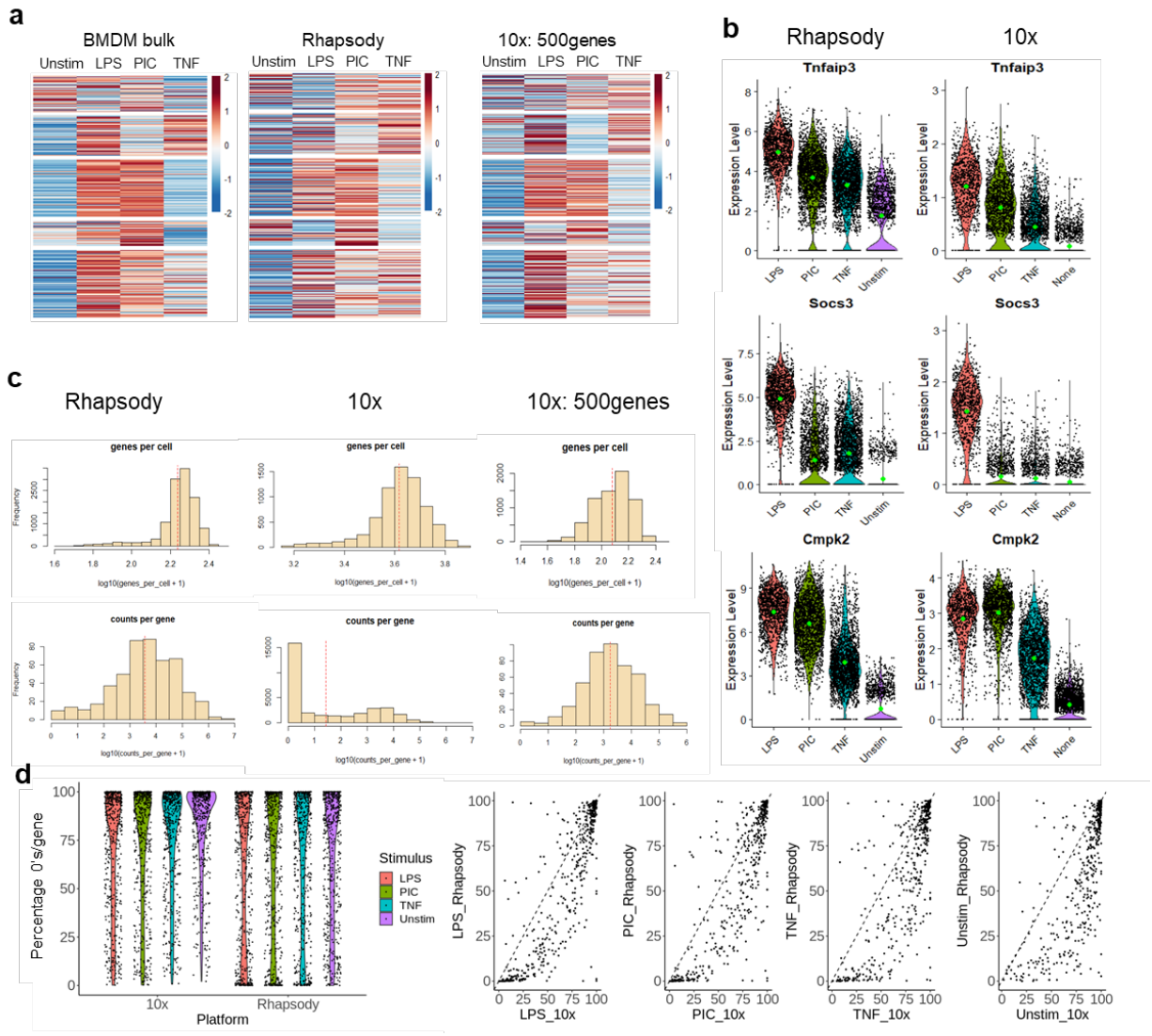


Figure S2.2. Comparison of single cell RNAseq platforms, Rhapsody vs 10x.

A) Heatmap of bulk RNAseq, and pseudobulk values from Rhapsody or 10x, for the same set of genes. **B)** Violin plots of scRNAseq data for a few examples genes for Rhapsody vs 10x. **C)** Distributions of genes measured per cell, or counts per gene for Rhapsody, 10x for all genes, and 10x for the 500 genes in the Rhapsody panel. **D)** Violin plot of percentage of 0 counts for each gene in the 10x or Rhapsody platforms. 100%: all cells are measured at 0 for that gene. Scatterplot comparing 10x vs Rhapsody across all common genes.

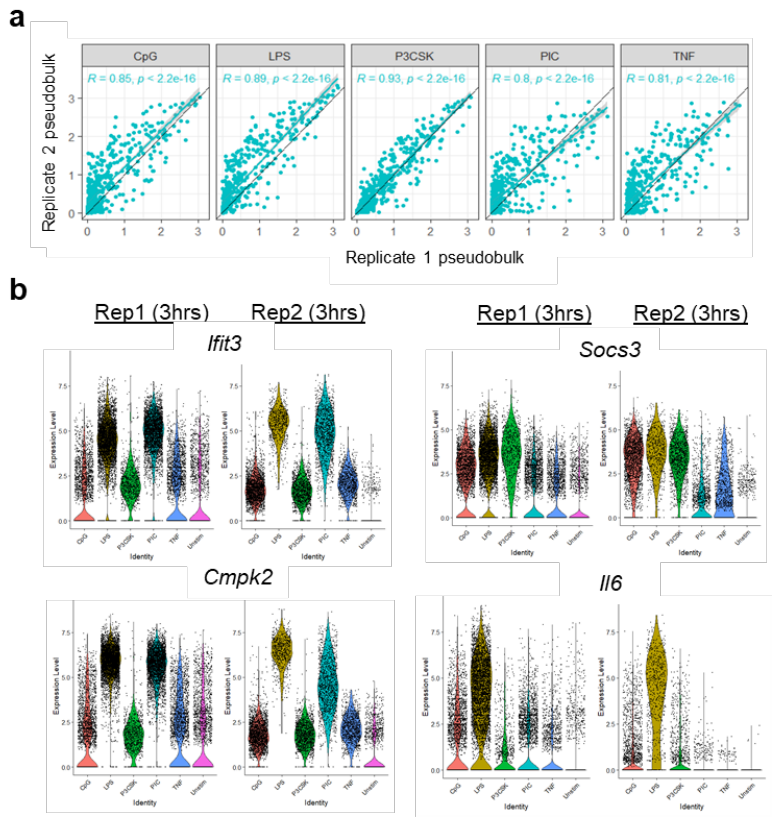


Figure S2.3. Reproducibility of the macrophage experimental system and measurement platform.

A) Scatterplot of pseudobulk values across 5 stimuli and 2 timepoints for 2 replicates from the Rhapsody scRNAseq platform. **B)** Single cell distributions for example genes from each replicate.

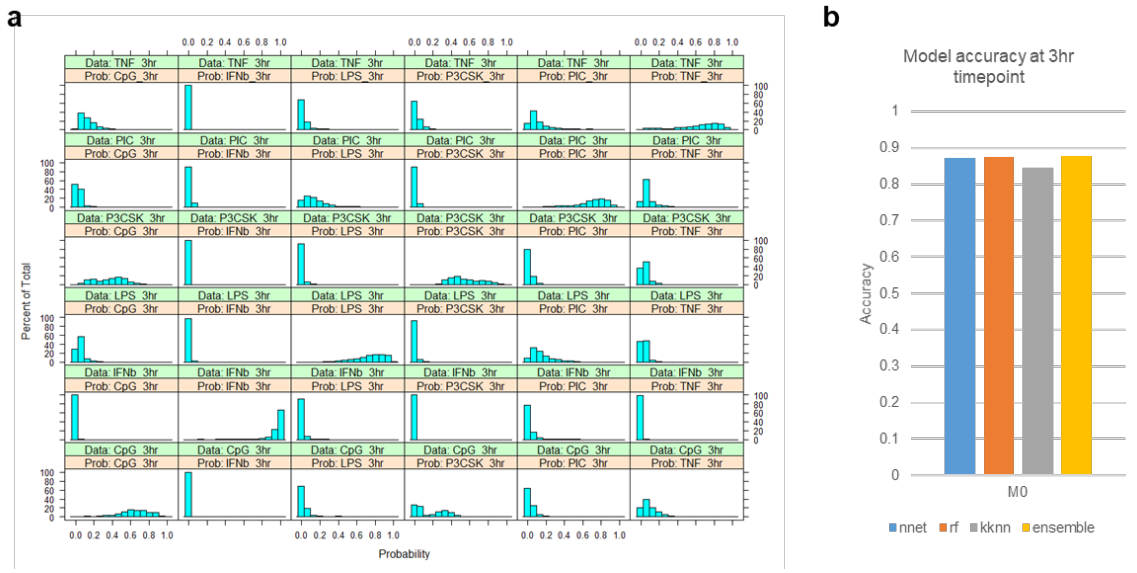


Figure S2.4. Machine learning test probabilities.

A) Distribution of prediction probabilities across cells for the test data. **B)** Comparison of overall accuracy for three different model types, and an ensemble prediction of majority voting from all three.

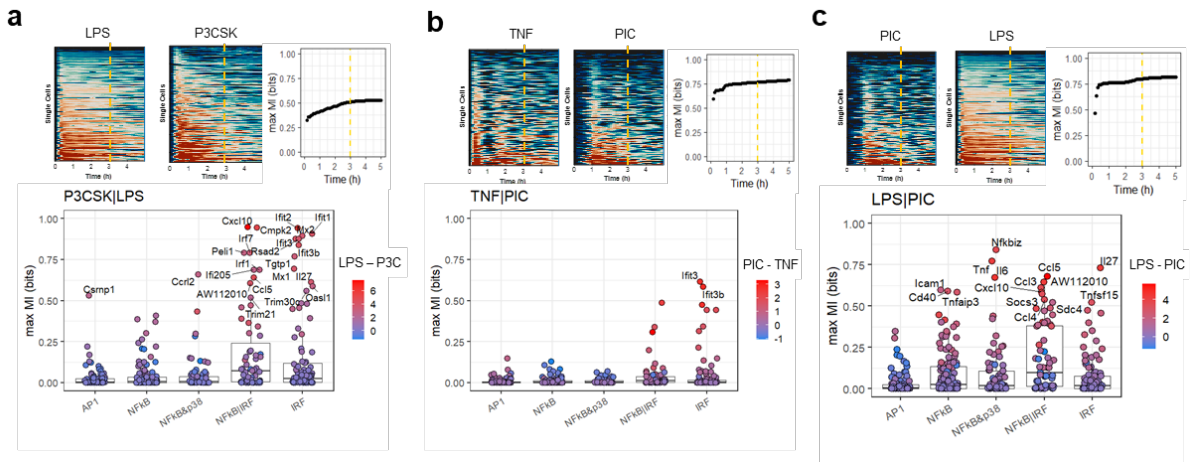


Figure S2.5. Relationship between pairwise specificity of NFκB signaling dynamics and pairwise specificity of gene expression.

A) Max MI in NFκB signaling dynamics vs max MI in expression of single genes, for LPS vs P3CSK, **B)** TNF vs PIC, and **C)** PIC vs LPS.

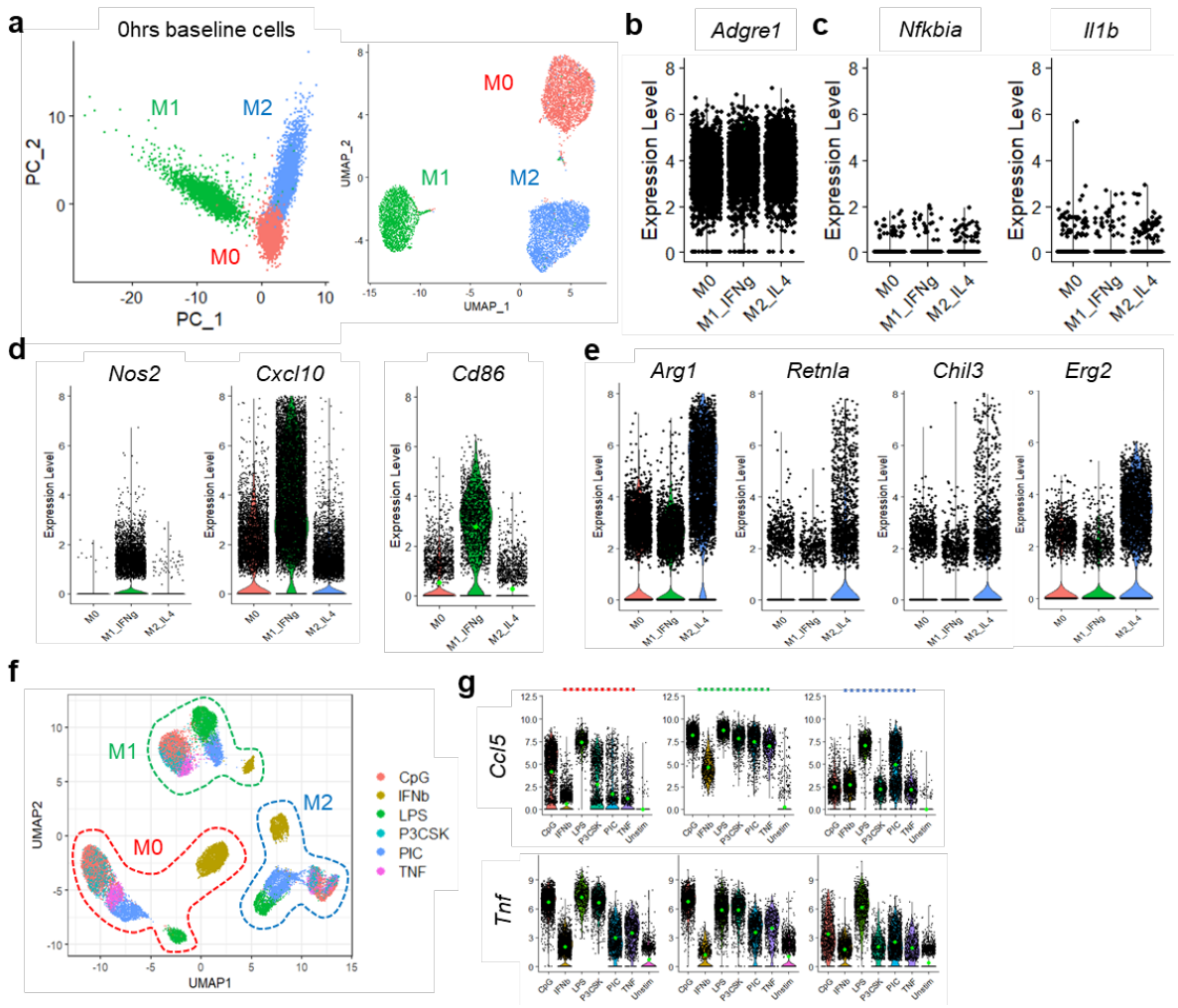


Figure S2.6. Polarized macrophages express the appropriate polarization markers.

A) PCA of M0, M1, and M2 macrophages at baseline 0hrs. M1 and M2 macrophages have been treated with IFNγ or IL4 for 24hrs. **B)** Macrophage marker gene *Adgre1* is expressed in all three conditions. **C)** NFκB target genes are not expressed at baseline. **D)** M1 marker *Nos2* or *Cxcl10* at 0.25-0.5 hour, all stimuli. M1 marker *Cd86* at 0hr. **E)** M2 marker genes at 0hrs. **F)** UMAP of M0, M1, M2 cells together. **G)** Gene expression values for M0, M1, and M2 macrophages for *Ccl5* and *Tnf* at 3 hours.

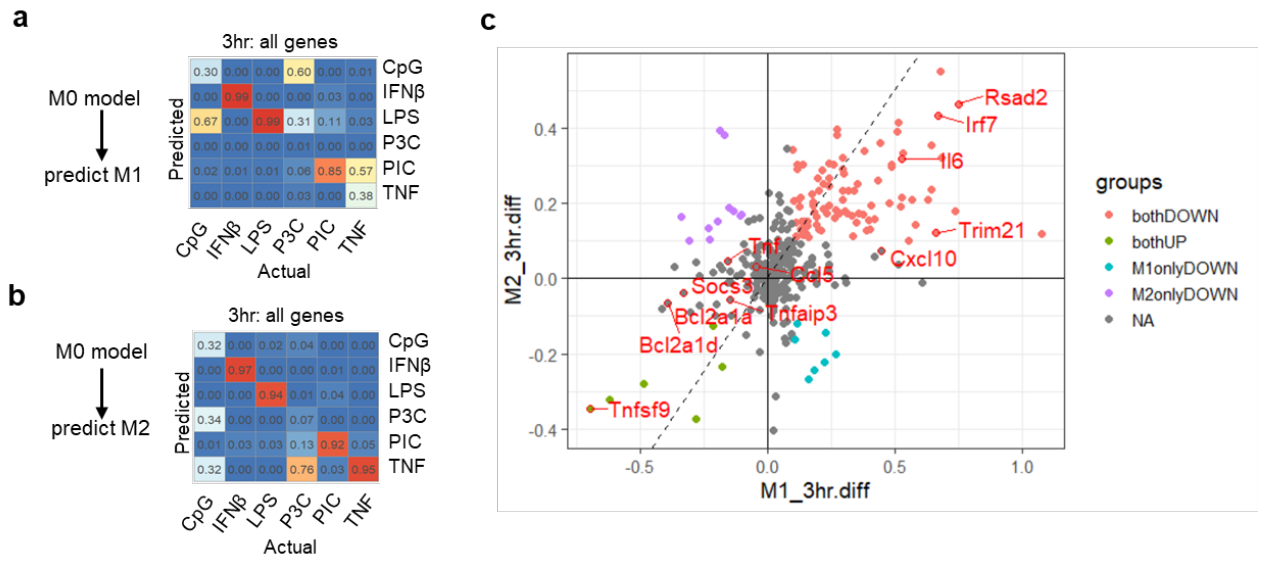


Figure S2.7. Statistical classifiers to predict M1 and M2 data identify which stimuli are more confused in polarized conditions.

A) Random forest model confusion matrices. M1 macrophages tested against the M0 model. Loss of specificity of IFN response genes cause more M1 cells responding to other stimuli to adopt an M0 LPS-like or PIC-like response. **B)** M2 macrophages tested against the M0 model. Loss of specificity in NFκB genes result in greater confusion among TNF, CpG and P3C. **C)** Correlation of the change in max MI for M1 vs M2 macrophages, each compared to M0.

TABLES

selection_criteria	gene
0.0000	*Cd14*
0.0000	*Nfkb1*
0.0000	*Nfkb2*
0.0000	*Rela*
0.0000	*Il4*
0.0000	*Il13*
0.0000	*Mmp8*
0.0000	*Mmp3*
0.0000	*Trim21*
0.0000	*Ccl1*
0.0000	*Tlr7*
0.0000	*Tlr4*
0.0000	*Arg1*
0.0000	*Ccl17*
0.0000	*Retnla*
0.0000	*Chil3*
0.0000	*Cd163*
0.0000	*Adgre1*
0.0000	*Irf3*
0.0000	*Irf7*
0.4205	*Ifnb1*
0.3084	*Il12b*
0.3055	*Ifna5*
0.2988	*Ifna2*
0.2976	*Thbs1*
0.2969	*Il10*
0.2908	*Cxcl1*
0.2887	*Il1b*
0.2793	*Ptgs2*
0.2764	*Ccl4*
0.2739	*Egr1*
0.2679	*Ifna4*
0.2595	*Ch25h*
0.2575	*Cxcl2*
0.2546	*Trem1*
0.2542	*Ifna1*
0.2535	*Cxcl10*
0.2482	*Fos*
0.2480	*Il6*
0.2479	*Ccl3*
0.2460	*Mt2*
0.2424	*Nfkbiz*
0.2409	*Edn1*
0.2408	*Fosb*
0.2399	*Nr4a1*
0.2390	*Ifna6*
0.2383	*Il1a*
0.2369	*Tnf*
0.2363	*Socs1*
0.2352	*Mmp13*
0.2333	*Gng4*
0.2332	*Ccl7*
0.2308	*Socs3*
0.2298	*Nr4a3*
0.2282	*Cav1*
0.2252	*Ctla2a*
0.2251	*Pdgfb*
0.2244	*Egr2*
0.2224	*Gdf15*
0.2219	*Kcnj2*
0.2210	*Cxcl11*
0.2196	*U90926*
0.2188	*Slc13a3*
0.2160	*Gpr84*
0.2151	*Nos2*
0.2148	*Ccl2*
0.2143	*Pde3b*
0.2138	*Sele*
0.2137	*Tdg-ps2*
0.2136	*Ppp1r15a*
0.2110	*Plat*
0.2094	*Serpine1*
0.2073	*Cd69*
0.2066	*Olr1*
0.2063	*Tmeff1*
0.2042	*Tnfsf9*
0.2035	*Nfil3*
0.2026	*Lcn2*
0.2015	*Errfi1*
0.2010	*Ppp1r12b*
0.2008	*Atp2b4*
0.2005	*Rnd1*
0.1998	*Adgb*
0.1998	*Cxcl3*
0.1983	*Ahr*
0.1983	*Lif*
0.1976	*Ccl5*
0.1959	*Adora2b*
0.1954	*Osm*

0.1949	*Cxcl9*
0.1931	*Marco*
0.1928	*Bmp2*
0.1926	*Hilpda*
0.1918	*Tnc*
0.1897	*Mycl*
0.1892	*Sdc2*
0.1878	*Ptch1*
0.1870	*Serpinb2*
0.1869	*Timp1*
0.1866	*Phlda1*
0.1863	*Saa3*
0.1855	*Gpr35*
0.1845	*Pim2*
0.1838	*Gbp2b*
0.1836	*Bcor*
0.1830	*Ptges*
0.1826	*Slc7a11*
0.1825	*Plagl1*
0.1824	*Gja1*
0.1823	*Tmem178*
0.1821	*Plau*
0.1820	*Il1rn*
0.1817	*Pmaip1*
0.1816	*Spsb1*
0.1807	*Oasl1*
0.1801	*Hhex*
0.1794	*Dnase1l3*
0.1790	*Flrt3*
0.1789	*Tmem26*
0.1788	*Fabp7*
0.1787	*Dusp5*
0.1786	*Slco3a1*
0.1784	*Csf3*
0.1783	*Ptger4*
0.1776	*Ifit1bl1*
0.1774	*Cx3cl1*
0.1772	*Cd83*
0.1772	*Ccl12*
0.1772	*H60b*
0.1769	*Ctla2b*
0.1766	*Phldb1*
0.1764	*H2-M2*
0.1763	*Fpr1*
0.1762	*Cd300lf*
0.1757	*Il4i1*

0.1753	*Lrp8*
0.1750	*Serpina3g*
0.1748	*Rgs1*
0.1746	*Cd300e*
0.1745	*Egr3*
0.1745	*Dusp14*
0.1743	*Ifi205*
0.1734	*Odc1*
0.1732	*Tagap*
0.1730	*Gfi1*
0.1724	*B3gnt5*
0.1723	*Cx3cr1*
0.1722	*Vcam1*
0.1713	*Clec4e*
0.1710	*Rsad2*
0.1709	*Shisa3*
0.1705	*Ccl22*
0.1704	*Rcan1*
0.1696	*Dusp2*
0.1692	*Trib3*
0.1685	*Syne2*
0.1682	*Flt1*
0.1682	*Siglece*
0.1682	*Plk2*
0.1682	*Rhou*
0.1679	*Ak4*
0.1677	*Pmepa1*
0.1676	*Vcan*
0.1676	*Col27a1*
0.1675	*Hp*
0.1674	*D16Ertd472e*
0.1672	*Ccl8*
0.1671	*Mmp9*
0.1671	*Snn*
0.1668	*Jag1*
0.1667	*Itga9*
0.1665	*Gem*
0.1664	*Tnfsf15*
0.1663	*Angpt1*
0.1662	*Cnn3*
0.1660	*Znrf3*
0.1660	*Irs2*
0.1660	*Gsta3*
0.1660	*Hspa1b*
0.1660	*Procr*
0.1658	*Jdp2*

0.1653	*Lox*
0.1650	*Acod1*
0.1650	*Prag1*
0.1649	*Casp4*
0.1648	*A530064D06Rik*
0.1646	*Tnfaip3*
0.1645	*Frmd6*
0.1644	*Pfkfb3*
0.1644	*Hmox1*
0.1637	*Mmp14*
0.1636	*Cish*
0.1636	*Asap3*
0.1635	*Dmwd*
0.1631	*Hcar2*
0.1630	*Cd40*
0.1626	*Cp*
0.1624	*Bambi*
0.1621	*Prdm1*
0.1620	*Dusp6*
0.1618	*Arhgap26*
0.1616	*Csrnp1*
0.1614	*Btg2*
0.1611	*Serpina3f*
0.1608	*Ddit4*
0.1607	*Ednrb*
0.1605	*Dusp4*
0.1602	*Chst11*
0.1602	*Tfec*
0.1602	*Ppfia3*
0.1601	*Rffl*
0.1601	*Pim3*
0.1598	*Rasgrp1*
0.1594	*Irf4*
0.1593	*Dusp1*
0.1591	*Ier3*
0.1590	*Arid5a*
0.1588	*Lad1*
0.1588	*S100a8*
0.1587	*Il27*
0.1586	*Ifit1*
0.1581	*Mpp7*
0.1576	*Slc16a1*
0.1574	*Fam43a*
0.1572	*Arg2*
0.1566	*Sema3c*
0.1565	*Card6*

0.1557	*Gda*
0.1556	*H1f2*
0.1556	*Gm5150*
0.1550	*Il12rb1*
0.1550	*Mt1*
0.1547	*Marcksl1*
0.1546	*Nfkbie*
0.1546	*Ccrl2*
0.1544	*Klra2*
0.1543	*Maff*
0.1542	*A630072M18Rik*
0.1542	*Tnfsf4*
0.1541	*Gm47283*
0.1540	*Mir155hg*
0.1534	*Tnip3*
0.1534	*Serpinb9b*
0.1533	*G530011O06Rik*
0.1533	*Zc3h12c*
0.1533	*Rgcc*
0.1530	*Cdk5r1*
0.1530	*Htr2a*
0.1530	*Zfp36*
0.1526	*Pde4b*
0.1524	*Tlr9*
0.1523	*Itpr1*
0.1522	*Gpr132*
0.1519	*Pdlim7*
0.1518	*Ifrd1*
0.1516	*Cebpd*
0.1515	*Gclm*
0.1514	*Itgb8*
0.1507	*Clic5*
0.1506	*Batf*
0.1502	*Csf1*
0.1502	*Gpr85*
0.1502	*Idi1*
0.1502	*Sphk1*
0.1497	*Zfp811*
0.1496	*Fzd1*
0.1496	*Tma16*
0.1495	*Noct*
0.1494	*Cst7*
0.1493	*Igf2bp2*
0.1492	*Bcl2a1d*
0.1492	*Ier2*
0.1486	*Kdm6b*

0.1482	*Cdc42ep2*
0.1481	*Klf8*
0.1477	*Nfkbid*
0.1473	*Ltbp2*
0.1471	*Otud1*
0.1469	*Hspa1a*
0.1465	*Cd38*
0.1464	*Dusp16*
0.1463	*Ccr1*
0.1463	*Tlr2*
0.1461	*Prrg4*
0.1461	*Tmem200b*
0.1461	*Kdr*
0.1461	*Gm6377*
0.1458	*Carmil1*
0.1458	*Slpi*
0.1457	*Tarm1*
0.1456	*Pou2f2*
0.1455	*Rras2*
0.1453	*Icosl*
0.1449	*Ets2*
0.1447	*Ifit3b*
0.1445	*Fgl2*
0.1444	*Dusp8*
0.1441	*Iigp1*
0.1439	*Plk3*
0.1436	*Ciita*
0.1433	*Cacnb3*
0.1432	*Fcgr2b*
0.1432	*Upp1*
0.1429	*Slfn4*
0.1428	*Gm19026*
0.1425	*Trib1*
0.1423	*Fmnl2*
0.1423	*Gpr68*
0.1423	*Igsf9*
0.1423	*AI504432*
0.1419	*Armcx6*
0.1418	*Myc*
0.1415	*Inhba*
0.1415	*Cacna1d*
0.1414	*H2-K2*
0.1413	*Zc3h12a*
0.1411	*Fscn1*
0.1409	*Acpp*
0.1409	*Hivep3*
0.1409	*Ccr3*
0.1407	*Lipg*
0.1405	*1110002J07Rik*
0.1403	*Adora2a*
0.1402	*Clcf1*
0.1400	*Ralgds*
0.1399	*Tle3*
0.1397	*Cd34*
0.1397	*Cmpk2*
0.1395	*Xcr1*
0.1394	*Plac8*
0.1393	*Grk3*
0.1391	*Olfml3*
0.1388	*Pla2g7*
0.1388	*Slamf9*
0.1385	*Taf4b*
0.1385	*Scimp*
0.1381	*Ly6a*
0.1379	*Gpr18*
0.1379	*Serpinb6b*
0.1376	*Slc16a3*
0.1375	*Abtb2*
0.1375	*Pde5a*
0.1374	*Gpr141*
0.1373	*Il18rap*
0.1372	*Fpr2*
0.1371	*Gna15*
0.1369	*Mtus1*
0.1367	*Rhoh*
0.1366	*Tmc3*
0.1364	*Spic*
0.1360	*Rnd3*
0.1360	*Mfap3l*
0.1358	*Il4ra*
0.1358	*Gsr*
0.1356	*Mxd1*
0.1356	*Irf1*
0.1354	*Ldlr*
0.1354	*Zeb1*
0.1353	*Ahrr*
0.0000	*Chil4*
0.1352	*Mx2*
0.1352	*Cd86*
0.1350	*P2ry2*
0.1350	*Slc7a2*
0.1346	*Cxcl14*

0.1345	*Plpp1*
0.1344	*Tnfsf14*
0.1343	*Pcx*
0.1341	*Zhx2*
0.1341	*B430306N03Rik*
0.1340	*Dmxl2*
0.1337	*Foxp4*
0.1334	*Htr2b*
0.1334	*Dnmt3aos*
0.1333	*Lpar1*
0.1330	*Dio2*
0.1329	*Trim30c*
0.1327	*Tnfsf10*
0.1324	*Pdpn*
0.1321	*Adgre4*
0.1318	*Junb*
0.1318	*Tmtc2*
0.1315	*Glipr2*
0.1313	*Ehd1*
0.1312	*Calhm6*
0.1312	*Sdc4*
0.1311	*AW112010*
0.1310	*Mefv*
0.1308	*Cd200*
0.1307	*Zfp324*
0.1307	*Spaca6*
0.1305	*Pilra*
0.1305	*Nqo1*
0.1305	*Rapgef5*
0.1302	*Nek6*
0.1300	*Klf2*
0.1299	*Swap70*
0.1298	*Adgrg6*
0.1297	*Ppm1l*
0.1296	*Hmga2*
0.1295	*P2ry13*
0.1293	*Gcnt2*
0.1287	*Trim36*
0.1285	*Bcl2a1a*
0.1283	*Slamf8*
0.1282	*Adgrl2*
0.1281	*Srxn1*
0.1279	*Eno2*
0.1275	*Pilrb1*
0.1275	*Myo1b*
0.1275	*Plekhg1*

0.1274	*Notch1*
0.1268	*Arhgef3*
0.1266	*Rhof*
0.1266	*Sla*
0.1266	*Adrb2*
0.1265	*Prokr1*
0.1265	*Peli1*
0.1262	*Tal2*
0.1262	*Stat4*
0.1262	*Ttc39b*
0.1262	*Gm16712*
0.1259	*Tnfsf8*
0.1259	*Slc28a2*
0.1257	*Ccnd2*
0.1256	*Ms4a4a*
0.1256	*Slc25a25*
0.1252	*Rab19*
0.1251	*Plpp3*
0.1251	*Itgal*
0.1248	*Ifit2*
0.1248	*Samsn1*
0.1248	*Gbp7*
0.1247	*Jun*
0.1246	*Cited2*
0.1245	*Hsd17b7*
0.1243	*Clec4n*
0.1243	*Ccnd1*
0.1241	*Flnb*
0.1239	*Cdc42ep4*
0.1238	*Cemip2*
0.1237	*Pde7b*
0.1237	*Ccne1*
0.1234	*Mtfr2*
0.1230	*Asns*
0.1230	*Oaf*
0.1230	*Serpinb9*
0.1229	*Treml4*
0.1229	*Ptgir*
0.1225	*Lratd2*
0.1222	*Clec5a*
0.1219	*Ptafr*
0.1219	*Hopx*
0.1214	*Mmd*
0.1214	*Nfkbia*
0.1214	*Mx1*
0.1212	*Nes*

0.1212	*Sod2*
0.1211	*Malt1*
0.1210	*Batf2*
0.1210	*Bst1*
0.1209	*Scn1b*
0.1207	*Arrdc4*
0.1203	*Clmp*
0.1203	*Spata13*
0.1202	*Icam1*
0.1200	*4930430E12Rik*
0.1196	*Intu*
0.1196	*P4ha2*
0.1193	*Alas1*
0.1192	*Ifitm1*
0.1192	*Degs2*
0.1191	*Ifit3*
0.1190	*H2-Eb1*
0.1189	*Cd80*
0.1189	*Slc30a4*
0.1186	*Vegfc*
0.1186	*Bcl2l11*
0.1185	*Htra1*
0.1184	*Cd55*
0.1184	*Klf7*
0.1183	*Atm*
0.1181	*Gm1966*
0.1180	*Heatr9*
0.1180	*Tsku*
0.1180	*Rgs3*
0.1180	*Irf8*
0.1179	*Gm4951*
0.1179	*Rab11fip1*
0.1179	*Ly75*
0.1178	*Ifi208*
0.1177	*Egln3*
0.1175	*Prdm9*
0.1175	*Orai2*
0.1175	*Ripk2*
0.1174	*Col18a1*
0.1173	*C920009B18Rik*
0.1173	*Tgtp1*
0.1172	*2500002B13Rik*
0.1172	*Traf1*

Table S2.1. Genes selected for amplification in targeted scRNAseq.

Selection criteria of 0.0 represents a manually selected gene. Other genes were chosen by radial distance of PCA loadings over the top 20 components. PCA was performed on all macrophage bulk RNAseq data available in Cheng et al, 2017.

timept	max MI	gene
3hr	1.500658	*Cmpk2*
3hr	1.380432	*Ifit3*
3hr	1.347347	*Irf7*
3hr	1.344224	*Ifit1*
3hr	1.333213	*Tgtp1*
3hr	1.31326	*Rsad2*
3hr	1.312613	*Ifit2*
3hr	1.253021	*Ifit3b*
3hr	1.121859	*Mx2*
3hr	1.07674	*Ifit1bl1*
3hr	1.013656	*Tnfsf10*
3hr	1.007489	*Mx1*
3hr	1.001306	*Gbp7*
3hr	0.999981	*Ifi205*
3hr	0.976623	*Arid5a*
3hr	0.975356	*Trim30c*
3hr	0.964883	*Trim21*
3hr	0.942231	*Peli1*
3hr	0.914971	*Cd69*
3hr	0.89603	*Ccl5*
3hr	0.89313	*Ehd1*
3hr	0.890749	*Tlr2*
3hr	0.875546	*Cxcl10*
3hr	0.85037	*Nfkbie*
3hr	0.823115	*Cited2*
3hr	0.812637	*Nfkbiz*
3hr	0.801045	*Tnfaip3*
3hr	0.799623	*Tnf*
3hr	0.793563	*Cdc42ep2*
3hr	0.784585	*Marcksl1*
3hr	0.782528	*Irf1*
3hr	0.765144	*Calhm6*
3hr	0.73712	*Il27*
3hr	0.730743	*Ccl4*
3hr	0.719536	*Pde4b*
3hr	0.70619	*Mt1*
3hr	0.704109	*Ccl12*
3hr	0.699614	*G530011O06Rik*
3hr	0.690546	*Il6*

Table S2.2. Max MI for individual genes, top 40 genes displayed.

Individual genes that provide greatest maximum mutual information between ligand and gene expression.

MATERIALS AND METHODS

Macrophage culture and Rhapsody scRNAseq

Macrophages were obtained by differentiating immortalized myeloid progenitors (iMPs) in DMEM/10% FBS + 30% L929 supernatant for a total of 10 days. Cells were replated into 6cm plates with new media on day 7, at a density of ~20k cells/cm^2. On day 10, cells were stimulated with 100ng/mL lipopolysaccharide (LPS, Sigma Aldrich), 10ng/mL murine TNF (R&D), and 50μg/mL low molecular weight polyinosine-polycytidylic acid (Poly(I:C)), 100nM synthetic CpG ODN 1668 (CpG), 500U/ml IFNβ, or media only Untreated control. For polarized macrophages, cells were incubated in 50ng/ml IFNγ or 50ng/ml IL4 for 24 hours prior to stimulation on day 10. To collect for scRNAseq, cells were washed 1x with cold PBS, then lifted into suspension by incubating at 37C for 5 minutes with Accutase, which resulted in cell viability typically >85%. Cells were centrifuged at 4C, 400g for 5 minutes, and resuspended in PBS + 2% FBS. Cells were hashtagged with anti-CD45-hashtags (BD Rhapsody # 633793) and loaded onto the cartridge following manufacturer's instructions (BD Rhapsody # 633771), with the following modifications: Incubation with hashtags was performed for 30mins on ice, instead of 20mins at room temperature; only two washes were performed after hashtag incubation. Each cartridge was then loaded with a total of ~36k cells across 12 hashtagged samples (~3k cells/sample). Libraries were prepared according to manufacturer's instructions (BD Rhapsody # 633771) and sequenced 2x100 on Novaseq 6000.

Peritoneal macrophage experiments

C57Bl/6 mice were ordered from Jackson labs: Two male mice were combined for each condition: 90wks old (000664 C57BL/6J), 17wks old (380050 C57BL/6J/DIO high fat diet (60% fat diet)), 17wks old (380056 C57BL/6J/DIO controls (10% fat diet)). Peritoneal macrophages were extracted by injecting 10mL PBS +1% FBS into the peritoneal space, shaking gently, and then pulling out as much fluid as possible, typically ~8ml. Macrophages

were plated in DMEM +10%FBS and allowed to rest 24hrs. Floating cells after that time were washed away, and remaining adherent macrophages were stimulated with the same ligand concentrations as for BMDMs: 100ng/mL lipopolysaccharide (LPS, Sigma Aldrich), 10ng/mL murine TNF (R&D), and 50μg/mL low molecular weight polyinosine-polycytidylic acid (Poly(I:C)), 500U/ml IFNβ, or media only Untreated control. Cells were washed once with cold PBS after 3hrs and lifted into suspension for the Rhapsody scRNAseq assay.

Selection of gene panel and stimuli

Bulk RNAseq data from Cheng et al, 2017 was obtained from GEO GSE68318. Counts were converted to counts per million (cpm) using the package *edgeR* (Robinson et al., 2010), and genes with cpm>4 in at least three samples were retained. Induced genes were gathered by comparing each of the 14 stimulus conditions, at each timepoint, against the unstimulated controls, and keeping genes with log2FC>2 and p-value < 10^{-5}, which resulted in 1502 genes. PCA was performed centered and unscaled on the induced genes across the three time points and the 14 stimuli in the dataset. For each gene, the loadings matrix was used to calculate a rank score, the radial distance of each gene j from the origin, over the top 20 PCs ($\text{score}_j = \sum_{x=1}^{20}\left(PCx_j\right)^2$). As a visualization of the approach, K-means clustering was performed on all induced genes and loadings were colored by cluster. Genes with the highest absolute loadings values in each PC tended to be from different clusters. The top genes in each principal component also showed distinct patterns across stimuli, with the first few principal components being most distinct. The top 480 ranked genes were included in the panel, and the remaining 20 genes were manually selected to add genes such as cell type markers, macrophage polarization markers, and transcription factors (Table 1). To identify distinct stimuli from the set of 14 in the bulk data, tensor components analysis was performed by folding the gene expression matrix into a genes x stimuli x timepoint tensor, using the package *rtensor*. Tucker

decomposition was performed, and the stimuli loadings matrix was hierarchically clustered using the first five components, which explained 92% of the variance in the data.

Motif Enrichment and Gene Ontology

Motif enrichment of induced and selected genes was performed using HOMER (Heinz et al., 2010), searching from -1000 to +100 of the TSS of each gene. Individual motif hits were categorized into five categories: bZIP (AP1 family TF motifs), IRF (IRF and ISRE motifs), RHD (Rel Homology Domain NFκB family motifs), ETS (Erythroblast Transformation Specific family TF motifs), and Zf (Zinc finger motifs). The average -ln(p value) of motifs in each category was calculated, and another log was taken for plot visualization. Gene ontology was performed on selected and unselected genes using *clusterProfiler* (Yu et al., 2012) against a background of all genes. Cutoff values of p-value < 0.01, Benjamini-Hochberg q-value <0.05, and minimum geneset size >5 were used. Ontologies were grouped if they had a similarity proportion greater than 0.7. Top three Biological Processes ontology terms for each group were plotted.

scRNAseq data processing

Fastq files were processed using the BD Rhapsody™ Targeted Analysis Pipeline (version v1.0) (Shum et al., 2019) hosted on Seven Bridges Genomics. DBEC corrected UMI counts (molecules per cell) were used in the downstream analysis. Multiplets, cells with undetermined barcodes, and cells with less than 80 features were removed from the analysis. Due to the selected 500 gene panel comprised of largely inducible genes, the assumption that the total number of RNAs per cell is constant does not hold. Counts were therefore normalized using the package *ISnorm* (Lin et al., 2020), rather than the more standard approach of dividing by total counts per cell. PCA was performed centered and unscaled using the R function *prcomp*, and UMAP and tSNE were performed on the top 20 PCs.

Machine Learning

Machine learning classification models were implemented using the R package *caret (classification and regression training)* (Kuhn, 2008). The data was split 70/30 into a training group and a testing group. Using the training data, a random forest model was trained using repeated 10-fold cross validation, with 3 repeats. The parameter mtry, which is the number of variables randomly selected as candidate features for each decision tree split, was set to sqrt(total number of variables). After the model was training, the remaining held-out data was tested, with each cell assigned a soft probability for each ligand. The highest probability ligand was the final prediction.

Mutual Information Analysis

Estimation of maximum mutual information was performed using the R package SLEMI (Jetka et al., 2019). Error bars were calculated using 10 bootstraps, each using 50% of the data. To relate max MI to means vs variances, max MI was plotted against either the average Fano factor across all stimuli for each gene, or the mean squared deviation ($MSD = \frac{\sum_{i=1}^{2}(x_i-\bar{x})^2}{2}$). For estimating the maximum mutual information of the best combination of genes, we first started from a list of the top 20 genes that individually had the best max MI value. For each of these single dimension channels, we scanned every combination of two genes, and again ranked the best combination of two genes and retained the top 20. This process was repeated until the gain in max MI for each additional gene leveled off. Retaining only the top 20 sets at each dimension made the calculation more computationally feasible, while still allowing the possibility for gene combinations that are not simply additive of the previous dimension's highest max MI combination.

Relating signaling data and gene expression

NFκB signaling data was obtained from previous publications (Adelaja et al., 2021; Cheng et al., 2021). Max MI between any two stimuli was calculated using the SLEMI algorithm, using the vector of data points up to the indicated timepoint. The max MI value at 3hrs, for instance, thus utilizes all data points from 0 to 3hrs. Max MI between pairs of stimuli was also calculated for each gene at 3hrs using the single cell RNAseq data. Max MI was plotted against gene regulatory clusters (Cheng et al, 2017) between the two stimuli.

Response Specificity Score

Response Specificity Score was calculated using two different approaches that led to similar results: 1) identification of smaller genesets by mutual information that captured stimulus-specific response diversity, and calculation of maximum mutual information between ligand and output using the reduced geneset, or 2) principal component analysis of all stimuli, and calculation of maximum mutual information using the PC scores as output.

Machine learning testing different genesets: An information theoretic approach was used to identify combinations of genes providing the highest maximum mutual information between ligand identity and gene expression. For machine learning performed on these smaller subsets of genes, the data was again partitioned into training and testing sets. After the data was partitioned, the training group was preprocessed by knn imputation (using KnnImpute), centered, and scaled. For the cross validation we used the out of box method and repeated it 40 times. To show the variable importance, we used the unscaled output of varImp (*caret*) so that the data would not be forced into a 0-100 scale. Ensemble modeling was performed by taking the predicted stimuli from each of the individual models and choosing the most common stimuli for each cell among the different models.

Projection onto dimensionality reduced space: PCA was performed centered and unscaled on all M0, M1 and M2 cells, representing the landscape of physiological macrophage responses.

Cells from disease models were projected onto this PCA space by multiplying by the PC loadings. Scores from any new projected data now sit in the same reduced space and can be compared. The maximum mutual information between ligand and PC scores was calculated for different subsets of ligands. The top 3 PCs were selected for the calculation to reduce overfitting; otherwise all samples would have a Response Specificity Score measured at the theoretical max. Three PCs were selected because they maintained the known differences in Response Specificity among M0, M1 and M2 macrophages.

REFERENCES

Adelaja, A., and Hoffmann, A. (2019). Signaling Crosstalk Mechanisms That May Fine-Tune Pathogen-Responsive NFκB. Front Immunol 10, 433.

Adelaja, A., Taylor, B., Sheu, K.M., Liu, Y., Luecke, S., and Hoffmann, A. (2021). Six distinct NFκB signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. Immunity 54, 916-930.e7.

Alleva, D.G., Pavlovich, R.P., Grant, C., Kaser, S.B., and Beller, D.I. (2000). Aberrant macrophage cytokine production is a conserved feature among autoimmune-prone mouse strains: elevated interleukin (IL)-12 and an imbalance in tumor necrosis factor-alpha and IL-10 define a unique cytokine profile in macrophages from young nonobese diabetic mice. Diabetes 49, 1106–1115.

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.

Brodin, P. (2021). Immune determinants of COVID-19 disease presentation and severity. Nature Medicine 27, 28–33.

Chaturvedi, S., Klein, J., Vardi, N., Bolovan-Fritts, C., Wolf, M., Du, K., Mlera, L., Calvert, M., Moorman, N.J., Goodrum, F., et al. (2020). A molecular mechanism for probabilistic bet hedging and its role in viral latency. PNAS 117, 17240–17248.

Cheng, C.S., Behar, M.S., Suryawanshi, G.W., Feldman, K.E., Spreafico, R., and Hoffmann, A. (2017). Iterative Modeling Reveals Evidence of Sequential Transcriptional Control Mechanisms. Cell Systems 4, 330-343.e5.

Cheng, Q., Behzadi, F., Sen, S., Ohta, S., Spreafico, R., Teles, R., Modlin, R.L., and Hoffmann, A. (2019). Sequential conditioning-stimulation reveals distinct gene- and stimulus-specific effects of Type I and II IFN on human macrophage functions. Scientific Reports 9, 5288.

Cheng, Q.J., Ohta, S., Sheu, K.M., Spreafico, R., Adelaja, A., Taylor, B., and Hoffmann, A. (2020). NFκB dynamics determine the stimulus-specificity of epigenomic reprogramming in macrophages. BioRxiv 2020.02.18.954602.

Cover, T.M., and Thomas, J.A. (2012). Elements of Information Theory (John Wiley & Sons).

Dragotakes, Q., Stouffer, K.M., Fu, M.S., Sella, Y., Youn, C., Yoon, O.I., Leon-Rodriguez, C.M.D., Freij, J.B., Bergman, A., and Casadevall, A. (2020). Macrophages use a bet-hedging strategy for antimicrobial activity in phagolysosomal acidification. J Clin Invest 130, 3805–3819.

Galani, I.-E., Rovina, N., Lampropoulou, V., Triantafyllia, V., Manioudaki, M., Pavlos, E., Koukaki, E., Fragkou, P.C., Panou, V., Rapti, V., et al. (2021). Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. Nature Immunology 22, 32–40.

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

Jetka, T., Nienałtowski, K., Winarski, T., Błoński, S., and Komorowski, M. (2019). Information-theoretic analysis of multivariate single-cell signaling responses. PLOS Computational Biology 15, e1007132.

Kolda, T.G., and Bader, B.W. (2009). Tensor Decompositions and Applications. SIAM Rev. 51, 455–500.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software 28, 1–26.

Lin, L., Song, M., Jiang, Y., Zhao, X., Wang, H., and Zhang, L. (2020). Normalizing single-cell RNA sequencing data with internal spike-in-like genes. NAR Genomics and Bioinformatics 2.

Matsushita, K., Takeuchi, O., Standley, D.M., Kumagai, Y., Kawagoe, T., Miyake, T., Satoh, T., Kato, H., Tsujimura, T., Nakamura, H., et al. (2009). Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. Nature 458, 1185–1190.

McKechnie, J.L., and Blish, C.A. (2020). The Innate Immune System: Fighting on the Front Lines or Fanning the Flames of COVID-19? Cell Host & Microbe 27, 863–869.

Mitchell, S., Mercado, E.L., Adelaja, A., Ho, J.Q., Cheng, Q.J., Ghosh, G., and Hoffmann, A. (2019). An NFκB Activity Calculator to Delineate Signaling Crosstalk: Type I and II Interferons Enhance NFκB via Distinct Mechanisms. Front. Immunol. 10.

Moore, J.B., and June, C.H. (2020). Cytokine release syndrome in severe COVID-19. Science 368, 473–474.

Muldoon, J.J., Chuang, Y., Bagheri, N., and Leonard, J.N. (2020). Macrophages employ quorum licensing to regulate collective activation. Nat Commun 11, 1–14.

Muñoz-Rojas, A.R., Kelsey, I., Pappalardo, J.L., Chen, M., and Miller-Jensen, K. (2021). Co-stimulation with opposing macrophage polarization cues leads to orthogonal secretion programs in individual cells. Nature Communications 12, 301.

Murray, P.J. (2017). Macrophage Polarization. Annual Review of Physiology 79, 541–566.

Nau, G.J., Richmond, J.F.L., Schlesinger, A., Jennings, E.G., Lander, E.S., and Young, R.A. (2002). Human macrophage activation programs induced by bacterial pathogens. Proceedings of the National Academy of Sciences 99, 1503–1508.

Rivera, A., Siracusa, M.C., Yap, G.S., and Gause, W.C. (2016). Innate cell communication kick-starts pathogen-specific immunity. Nat Immunol 17, 356–363.

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.

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al. (2020). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell 0.

Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. (2014). Accurate information transmission through dynamic biochemical signaling networks. Science 346, 1370–1373.

Sen, S., Cheng, Z., Sheu, K.M., Chen, Y.H., and Hoffmann, A. (2020). Gene Regulatory Strategies that Decode the Duration of NFκB Dynamics Contribute to LPS- versus TNF-Specific Gene Expression. Cell Syst 10, 169-182.e5.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature 510, 363–369.

Shannon, C.E. (1948). A mathematical theory of communication. The Bell System Technical Journal 27, 379–423.

Shum, E.Y., Walczak, E.M., Chang, C., and Christina Fan, H. (2019). Quantitation of mRNA Transcripts and Proteins Using the BD RhapsodyTM Single-Cell Analysis System. Adv. Exp. Med. Biol. 1129, 63–79.

Suderman, R., Bachman, J.A., Smith, A., Sorger, P.K., and Deeds, E.J. (2017). Fundamental trade-offs between information flow in single cells and cellular populations. PNAS 114, 5755–5760.

Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A., and Ng, L.F.P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol 1–12.

Tkačik, G., Callan, C.G., and Bialek, W. (2008). Information flow and optimization in transcriptional regulation. PNAS 105, 12265–12270.

Yamamoto, M., Yamazaki, S., Uematsu, S., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Kuwata, H., Takeuchi, O., Takeshige, K., et al. (2004). Regulation of Toll/IL-1-receptor-mediated gene expression by the inducible nuclear protein IkappaBzeta. Nature 430, 218–222.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16, 284–287.

CHAPTER 3

Epigenomic Programmability

3.1. A mathematical model of nucleosomal DNA interactions predicts epigenetic response to distinct dynamic features of NFκB.

ABSTRACT

Nuclear factor kB (NFκB) is a transcription factor that is activated by all pathogen-associated stimuli and can reprogram the epigenome through its ability to bind DNA. However, it is unclear how different stimuli alter NFκB-regulated regions of the epigenome to different extents. Here, we develop a mathematical model to show why NFκB remodels chromatin heavily only in response to a subset of NFκB-activating stimuli. We found that this stimulus specificity depends on the temporal dynamics of NFκB activity, most importantly, whether the dynamics are oscillatory or non-oscillatory. Non-oscillatory NFκB opens chromatin by maintaining a longer continuous residence time in the nucleus, promoting an unrelenting disruption of nucleosomal histone–DNA interactions. The model thus provides a mechanistic explanation for how the temporal dynamics of transcription factors are linked to their capacity to reprogram the epigenome.

INTRODUCTION

The precise positioning of nucleosomes is central in governing the activity of DNA regulatory regions. Nucleosomes occupying regulatory element regions, such as enhancers and promoters, prevent transcription factors from accessing DNA. Although certain DNA sequences have an affinity for being occupied by nucleosomes (Kaplan et al., 2009; Segal et al., 2006), nucleosome positioning and nucleosome occupancy is highly dynamic. Nucleosomal DNA wraps around the histone through a collection of weak, multivalent interactions (Polach and Widom, 1995). Approximately 147bp of DNA wrap two and a half times around the core histone octamer, resulting in 14 main non-covalent DNA-histone contact points (Luger et al., 1997a). In order for enhancer or promoter regions to be exposed, histones binding nucleosomal DNA must

be displaced. The removal of histones from nucleosomal DNA to expose functional regions of the genome has been shown to be mediated by multiple factors, including ATP-dependent chromatin remodelers such as SWI/SNF, and the binding of the binding of sequence-specific DNA binding proteins, referred to as “pioneer factors”, which cooperate with other transcription factors. The dynamic aspects of nucleosomes are also attributed to spontaneous wrapping and unwrapping of the DNA from histone at the locations furthest from the dyad (Tims et al., 2011a). These short periods of DNA accessibility may be stabilized by the binding of transcription factors if their cognate binding sequence is present in that stretch of DNA and they are present at sufficiently high concentrations (Segal and Widom, 2009).

Multiple previous models of spontaneous nucleosome dynamics have been proposed (Mobius et al., 2013; Polach and Widom, 1995). Mobius et al. proposed a step-wise stochastic model of 14 steps, assuming equivalent DNA-histone interaction contacts at each of the steps, resulting in an exponential relationship between site exposure time and distance from the edge of the nucleosome. Multistep models of nucleosome remodeling have also been developed to describe the process by which ATP-dependent SWI/SNF remodelers reposition nucleosomes (Morgan et al., 2018).

Experimental studies have indicated the importance of the location of TF binding sites in relation to the nucleosome dyad. Two TF binding sites on the same side of the dyad exhibit positive cooperativity against a nucleosome, while two TF sites on opposite sides of the dyad do not (Moyle-Heyrman et al., 2011). This suggests two features: 1) that the presence of TFs competes with nucleosome rewrapping, and 2) that the unwrapping of the nucleosome may be primarily unidirectional. Additional studies performed on single nucleosomes in vitro, with the linker segments of nucleosomal DNA being pulled by a small force, support the asymmetric unwrapping of DNA (Ngo et al., 2015). Nucleosome models have thus been constructed accordingly, with unwrapping beginning from one end, which

is also in agreement with previous spontaneous nucleosome models, and proceeding unidirectionally when influenced by the binding of TFs. However, none of these models consider the effect of transcription factors binding to regulatory elements, and their influence on spontaneous nucleosome wrapping and unwrapping.

Additionally, TFs known as signal dependent transcription factors (SDTFs) have been shown to exhibit various temporal patterns of nuclear activity in response to different stimuli or in disease states (Hao and O'Shea, 2012; Hoffmann et al., 2002). Incorporating such behavior can further the understanding of how chromatin responds to aberrant signaling. Although the mechanisms by which chromatin decodes distinct signaling dynamics remain unclear, analyzing these models may suggest possible mechanisms that can explain the observed response of the epigenome to SDTF dynamics. Presumably, persistent as opposed to oscillatory SDTF signaling dynamics may allow longer uninterrupted residence time in the nucleus, promoting displacement of the nucleosome from any particular location and the stabilization of nucleosome-free DNA at a previously occupied position. In addition, chromatin accessibility has been linked to increased deposition of H3K4me1 (Lara-Astiaso et al., 2014), and thus more permanent changes to epigenetic state through *de novo* enhancer formation.

Here we develop a mathematical model of DNA-histone interactions to show that the stimulus-specificity of epigenetic reprogramming depends on the temporal dynamics of NFκB activity, in particular, whether the dynamics are oscillatory or non-oscillatory. In the model we allow a dynamic SDTF input to affect the step-wise opening of the nucleosome model, to study how features in the temporal dynamics of transcription factors can control epigenetic plasticity. Non-oscillatory NFκB opens chromatin by sustained disruption of nucleosomal histone-DNA interactions, enabling activation of latent enhancers that modulate expression of immune response genes. Thus, the stimulus-specific presence of

oscillatory or non-oscillatory dynamics can determine a transcription factor's capacity to reprogram the epigenome.

RESULTS

A mathematical model of nucleosomal DNA interactions predicts epigenetic response to distinct dynamic features of NFκB.

In response to various stimuli, NFκB enters the nucleus with distinct speeds, amplitudes, and durations, and may oscillate between the nucleus and cytoplasm. To determine whether stimulus-specific NFκB dynamics play a role in stimulus-specific activation of latent enhancers, we used live-cell microscopy of BMDMs expressing NFκB-RelA fused with the mVenus fluorophore (mVenus-RelA) (Adelaja et al., 2021) to measure the single-cell dynamics of NFκB-RelA in response to each of the five ligands (Fig. S1E). We quantified the six NFκB dynamic features that function to encode ligand identity and dose (Adelaja et al., 2021) and correlated them to mean H3K4me1 counts in the NFκB- activated latent enhancers (Fig. S3). Oscillatory power (r = -0.95), total activity (r = 0.77), and peak amplitude (r = 0.78) were highly correlated with the capacity of a given stimulus to activate latent enhancers (Fig. S1F).

We hypothesized that the temporal dynamics of NFκB activity might affect its interaction with chromatin. Crystallographic studies imply that stable NFκB-DNA binding requires the DNA to be nucleosome-free because NFκB dimers embrace the DNA double helix circumferentially (Chen et al., 1998; Luger et al., 2000) (Fig. 1A). However, NFκB can interact with nucleosomal DNA, particularly when its binding site is distal to the nucleosome dyad (Lone et al., 2013b). Indeed, the DNA-histone interface is composed of low-affinity interactions that allow spontaneous disassociation or "breathing" (Li et al., 2005b). Thus, successive disruptions of DNA-histone contacts by NFκB, in collaboration with remodeling complexes such as SWI/SNF (Kobayashi et al., 2017), chaperone proteins such as FACT

(Gasparian et al., 2011; Winkler and Luger, 2011), and/or pioneer factors such as Pu.1 or CEB/Pα (Jin et al., 2011), may displace the nucleosome (Fig. 1B). This may be followed by the deposition of histone modifications on neighboring nucleosomes, resulting in a poised or active enhancer (Ostuni et al., 2013b).

We created a multi-step model describing how dynamical NFκB activity might interact with nucleosomal DNA. A series of 14 Hill equations described the competition between NFκB and histone for interacting with DNA (Fig. 1C), reflecting the number of contact points in the histone octamer-DNA crystal structure (Davey et al., 2002). Relative rates of nucleosome wrapping and unwrapping were based on available biophysical data (Tims et al., 2011a). With measured single-cell NFκB activities (Fig. S1E) as inputs, the model simulations reproduced the differences in experimental H3K4me1 ChIP-seq data (Fig. 1D-E, S3A-B).

We used the model to investigate which features of NFκB dynamics affect chromatin accessibility. We examined the three features most highly correlated with the H3K4me1 ChIP-seq data (Fig. S1F): non-oscillatory, amplitude, and total activity. The model indicated that a non-oscillatory dynamic produces a two-fold greater chromatin accessibility than an oscillatory dynamic (Fig. 1F). The model also indicated that NFκB activity must have a minimal amplitude (Fig. 1G, S3C) and extend for a minimal duration (Fig. 1H, S3D) to open chromatin; but above these thresholds, non-oscillatory NFκB always has greater capacity to open chromatin than oscillatory NFκB. This was consistent across a range of parameter values (Fig. S4). These simulations led to the striking prediction that the presence or absence of oscillations, not the maximum amplitude or duration of activity, is the key determinant of whether NFκB preserves or alters the chromatin state.

DISCUSSION

We here described a simple model of nucleosome dynamics in response to NFκB activation, which predicted the stimulus-specific epigenetic plasticity of macrophages. Non-oscillatory as opposed to oscillatory NFκB signaling created larger continuous nuclear residence time that prevented nucleosome rewrapping and led to greater chromatin accessibility. Extending this model to include molecular components of nucleosome dynamics that increasingly reflect cellular processes involved in epigenetic reprogramming may provide further insight. Furthermore, the model can potentially be generalized to other cell types and other signal-dependent transcription factors. While here in this model, we considered only a single nucleosome at a time, data-guided models in which arrays of nucleosomes are represented could further the concordance of models to experimental data (Mobius et al., 2013). In these nucleosome array models, models of single nucleosome behavior could be linked together, where each nucleosome would be influenced by neighboring nucleosomes (Dodd et al., 2007).

Other models of the epigenome that build on models of chromatin accessibility include those that consider the deposition of histone modifiers by enzymes recruited to regions of open chromatin. Combining nucleosome arrays and histone marks have been previously been theoretically modeled in the context of propagation of histone marks along strings several neighboring nucleosomes (Dodd et al., 2007). Although the mechanisms are less clear, modeling could provide an abstraction of this process through which multiple possible mechanisms could be tested against experimental measurements of both chromatin accessibility and histone marks. The incorporation of chromatin accessibility is made further interesting by the step-wise nature of nucleosome remodeling, as opposed to single-step histone deposition reactions. Ultimately, models that include the interactions between chromatin accessibility and histone marks will be further informative.

Although increasing amounts of both structural and sequencing data have emerged for nucleosomes (Zhou et al., 2019), it is unclear how their structure affects their function and response to other dynamically regulated proteins in vivo. In our model, persistent, as opposed to oscillatory, transcription factor signaling dynamics allows transcription factor molecules to reside longer at high concentration in the nucleus, which promote increased displacement of the nucleosome due to the ability of the nucleosome to progress through more intermediate steps. Extensions of this model to include the deposition of histone marks could further support the complete repositioning of the nucleosome as an absorbing state without a backward reaction, as open chromatin regions marked as enhancers through covalently bonded histone marks more permanently alter the state of that region of DNA.

Acknowledgements: We would like to thank Kensei Kishimoto, Diane Lefaudeux, Xiao-Fei Lin, and Anup Mazumber for their experimental and analytical contributions, and Siavash Kurdistani, Michael Carey, Eason Lin and Ying Tang for their insights and critical reading of the manuscript. **Funding:** This work was supported by NIH grants R01-AI127864, R01-GM117134, F31-AI138450, T32-GM008042, and T32-AI089398, as well as the Specialty Training and Advanced Research (STAR) program of the UCLA Department of Medicine. Sequencing was performed at the UCLA Broad Stem Cell Center Sequencing Core. **Author contributions:** QC, SO, AA, and BT performed the experiments. QC, SO, KS, RS, and AA analyzed the data. KS and BT developed the mathematical model. QC, KS, and AH wrote the manuscript. All authors reviewed the manuscript. AH coordinated and funded the work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw data and count tables for ChIP-seq, ATAC-seq, and RNA-seq data have been submitted to the NCBI GEO repository under accession number GSE146068. Code to reproduce final figures are available on Zenodo (DOI: 10.5281/zenodo.4698447).

FIGURES AND FIGURE LEGENDS

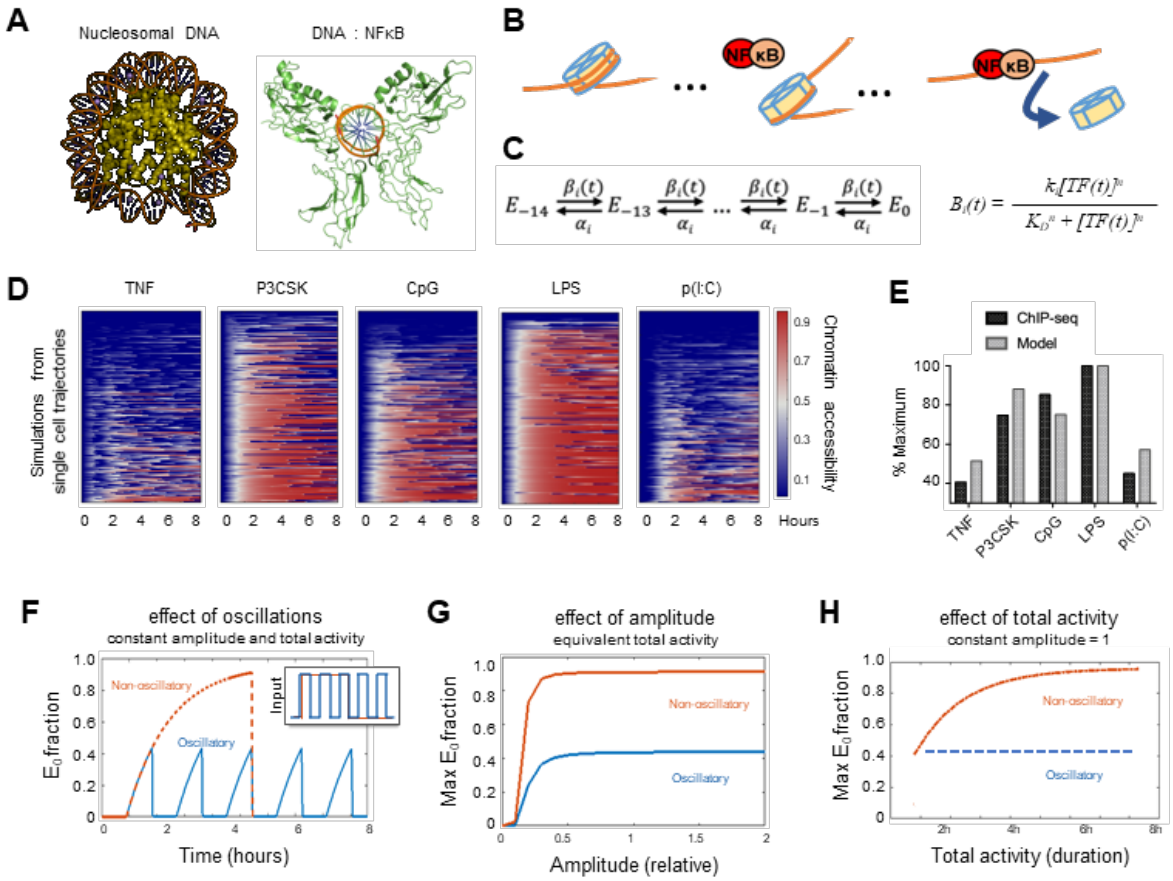


Figure 3.1.1: Mathematical model predicts epigenetic response to distinct dynamic features of NFκB.

A) Crystal structures of nucleosomal DNA (PDB 1F66) *vs.* NFκB-bound DNA (PDB 1VKX), where p65:p50 NFκB dimer is in green. **B)** Schematic of model illustrating NFκB-driven displacement of nucleosome. **C)** Multi-step model with 14 steps to complete nucleosome unwrapping, each expressed as a Hill function. **D)** Heat maps of simulations of chromatin opening in response to different stimuli, using single cell trajectories from NFκB microscopy data as input. **E)** Model simulation *vs.* ChIP-seq data. Mean ChIP-seq counts from 1071 latent NFκB enhancers (Fig. 1D), background-subtracted and scaled to maximum signal (LPS stimulation). Model simulations are mean of maximum E_0 fraction per cell (cf. Fig. S3a), scaled to LPS condition. **F)** Model simulation of predicted chromatin accessibility comparing oscillatory *vs.* non-oscillatory input activities. **G-H)** Model simulation of predicted chromatin opening across a range of amplitudes and durations.

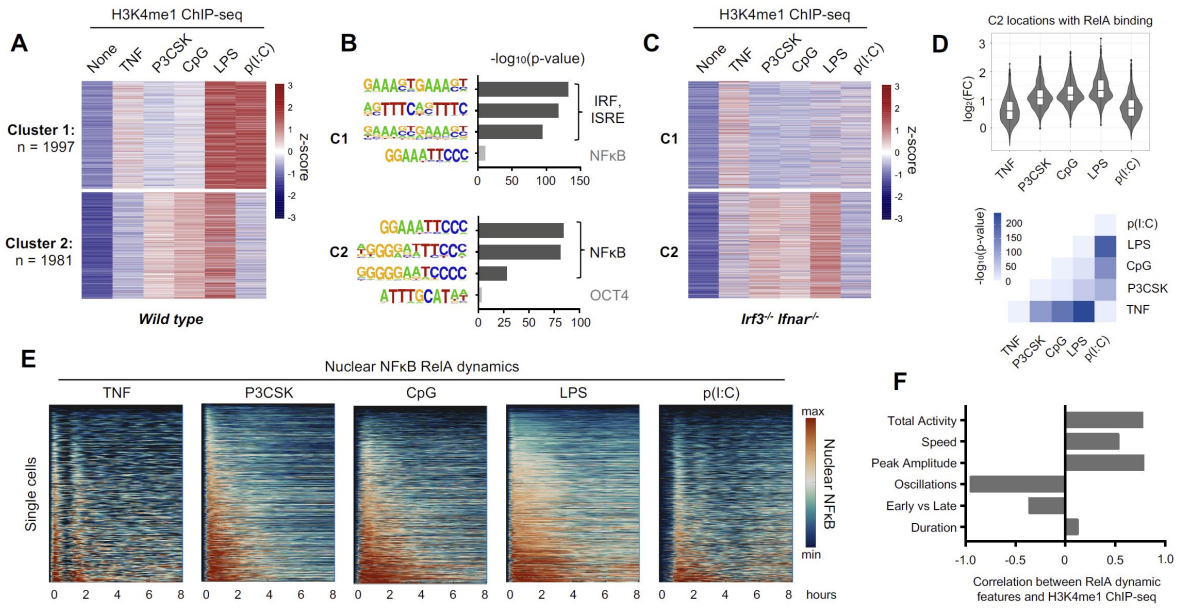


Figure S3.1.1: NFκB-driven *de novo* enhancers are stimulus-specific and correlate to dynamic features of NFκB activity.

A) Heat map of H3K4me1 ChIP-seq inducible peaks from BMDMs stimulated with five ligands for eight hours, unsupervised K-means clustering. Average of two replicates. **B)** Known transcription factor motifs with greatest enrichment in Cluster 1 and Cluster 2 peaks. **C)** Heat map of H3K4me1 ChIP-seq in *Irf3*$^{-/-}$*Ifnar*$^{-/-}$ BMDMs, using same clusters as panel (A). **D)** Violin and box plots of $\log_2$(fold-change) in H3K4me1 signal of 1071 NFκB enhancers from Cluster 2 that also contain an NFκB-RelA binding event. Corresponding matrix of *p*-values of H3K4me1 ChIP-seq fold-change, by two-tailed t-test between pairs of conditions. **E)** Heat maps of NFκB activity in single cells by live cell microscopy of *mVenus-RelA* BMDMs, showing nuclear abundance of NFκB in response to five stimuli. **F)** Bar graph of correlation coefficients (absolute value) between mean H3K4me1 ChIP-seq counts of NFκB enhancers and the six key features of NFκB dynamics[3] (see also Fig. S2).

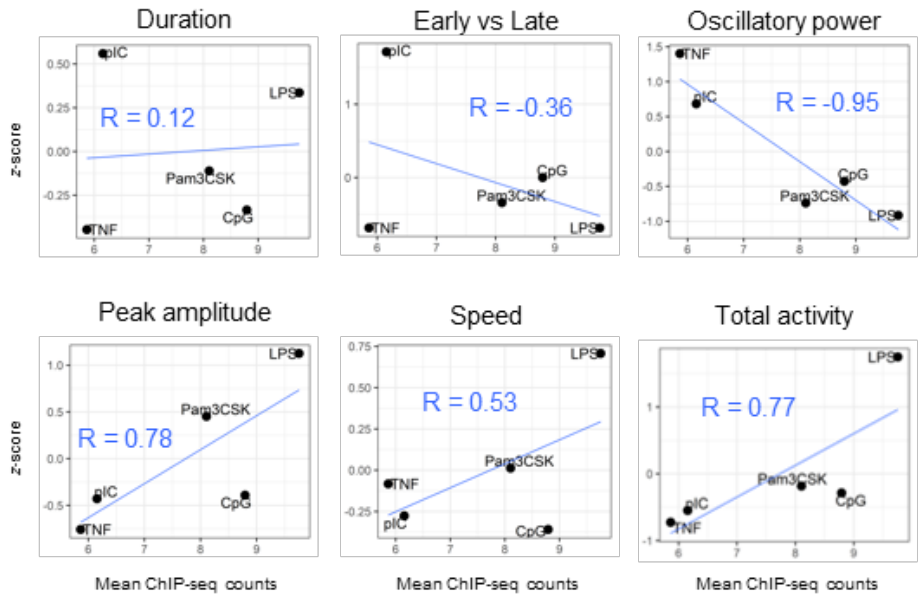


Figure S3.1.2: Correlation of NFκB dynamics to ChIP-seq data.

Scatterplots and correlations of mean NFκB cluster ChIP-seq counts (Fig 1D) *vs.* stimulus-specific z-scores for each of the six key features of NFκB signaling dynamics

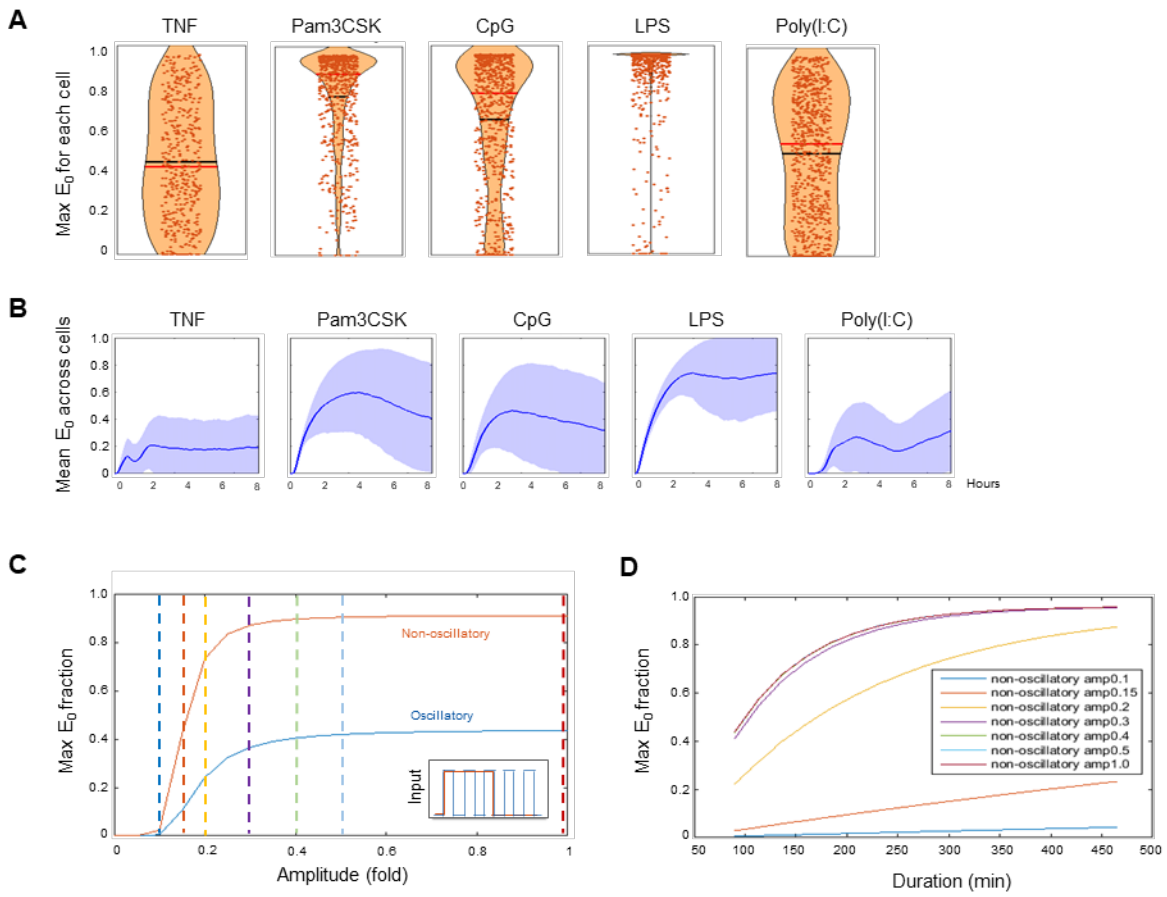


Figure S3.1.3: Summary of model simulations.

A) Violin plots of maximum chromatin opening over eight hours per single-cell stimulation, using NFκB trajectories as input to the model. Black line = mean, Red line = median. **B)** Simulated mean chromatin opening over time across all single cells. **C)** Model simulations across a range of NFκB amplitudes, comparing oscillatory and non-oscillatory trajectories. **D)** Model simulations across a range of NFκB durations, comparing a range of NFκB amplitudes marked by dotted lines in Panel (C).

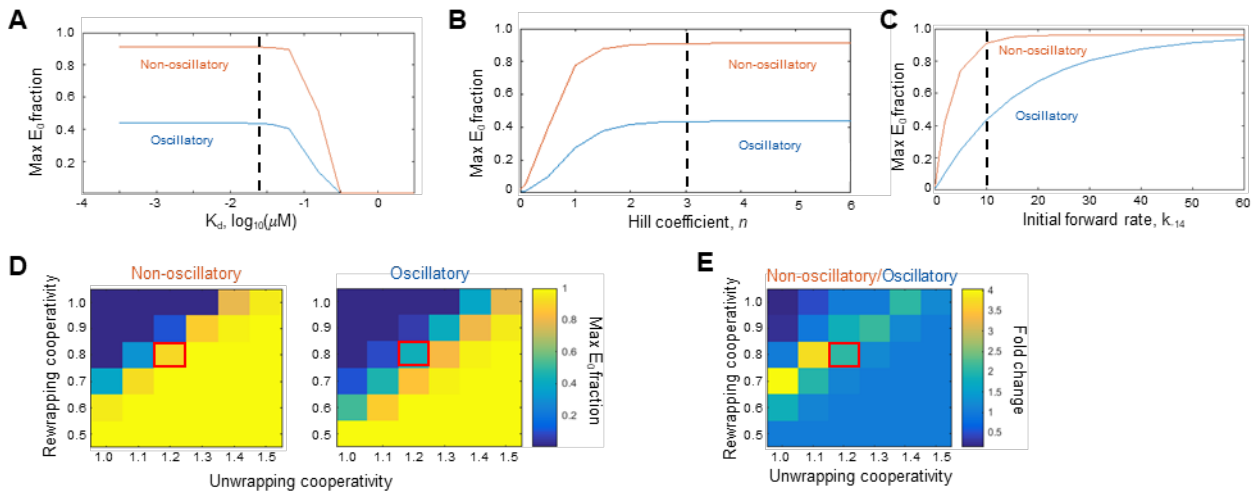


Figure S3.1.4: Parameter sensitivity analysis.

A) Chromatin opening behavior when the model is tested across a range of K_{DS}, **B)** across a range of Hill coefficients, or **C)** across a range of forward rates for the first unwrapping step, k_{-14}. For model simulations (Fig. 2D), $K_D = 0.025$, Hill = 3, and $k_{-14} = 10$ were used, marked by the dotted black line. **D-E)** Heat map of chromatin opening across a range of unwrapping and rewrapping cooperativity factors, showing maximum E_0 fraction in non-oscillatory and oscillatory conditions (D) and fold change difference between maximum non-oscillatory and oscillatory conditions (E). Red box indicates the parameter values used for model simulations.

TABLES

#	Reactants	Rxn	Products	Value	Unit
1	tf + E_0	→	E_1	10.00	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
2	E_1	→	E_0	75.00	min^{-1}
3	tf + E_1	→	E_2	12.00	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
4	E_2	→	E_1	60.00	min^{-1}
5	tf + E_2	→	E_3	14.40	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
6	E_3	→	E_2	48.00	min^{-1}
7	tf + E_3	→	E_4	17.28	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
8	E_4	→	E_3	38.40	min^{-1}
9	tf + E_4	→	E_5	20.74	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
10	E_5	→	E_4	30.72	min^{-1}
11	tf + E_5	→	E_6	24.88	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
12	E_6	→	E_5	24.58	min^{-1}
13	tf + E_6	→	E_7	29.86	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
14	E_7	→	E_6	19.66	min^{-1}
15	tf + E_7	→	E_8	35.83	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
16	E_8	→	E_7	15.73	min^{-1}
17	tf + E_8	→	E_9	43.00	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
18	E_9	→	E_8	12.58	min^{-1}
19	tf + E_9	→	E_10	51.60	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
20	E_10	→	E_9	10.07	min^{-1}
21	tf + E_10	→	E_11	61.92	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
22	E_11	→	E_10	8.05	min^{-1}
23	tf + E_11	→	E_12	74.30	$uM^{-1}min^{-1}$

		→		3	Hill coef.
		→		0.025	uM (K_D)
24	E_12	→	E_11	6.44	min^{-1}
25	tf + E_12	→	E_13	89.16	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
26	E_13	→	E_12	5.15	min^{-1}
27	tf + E_13	→	E_14	106.99	$uM^{-1}min^{-1}$
		→		3	Hill coef.
		→		0.025	uM (K_D)
28	E_14	→	E_13	4.12	min^{-1}

Table 3.1.1. Table of parameter values and reactions for the nucleosome model.

METHODS

Model of nucleosome opening

Model formulation: A multistep model of DNA unwrapping from the histone octamer was formulated based on structural studies that identified 14 contacts between the histone octamer and double helical DNA and biophysical studies of single nucleosomes *in vitro* that revealed multiple, step-wise transitions in DNA unwrapping (Luger et al., 1997a; Möbius et al., 2006a; Tims et al., 2011a). The model describes the population average or probability of many stochastic events, with each species representing the fraction of NFκB-responsive latent enhancers in a cell (E) at a given state of accessibility. E_{-14} describes the most closed state in which all 14 contact points are engaged, and E_0 the most open state in which the histone octamer is entirely evicted.

$$E_{-14} \underset{\alpha_{-14}}{\overset{\beta_{-14}(t)}{\rightleftarrows}} E_{-13} \cdots \underset{\alpha_i}{\overset{\beta_i(t)}{\rightleftarrows}} E_{i+1} \cdots \underset{\alpha_{-1}}{\overset{\beta_{-1}(t)}{\rightleftarrows}} E_0$$

A system of ordinary differential equations was formulated to describe this model:

$$\frac{dE_{-14}}{dt} = \alpha_{-14}E_{-13} - \beta_{-14}(t)E_{-14} \tag{1}$$

$$\frac{dE_{-13}}{dt} = \beta_{-14}(t)E_{-14} - \alpha_{-14}E_{-13} - \beta_{-13}(t)E_{-13} + \alpha_{-13}E_{-12} \tag{2}$$

...

$$\frac{dE_i}{dt} = \beta_{i-1}(t)E_{i-1} - \alpha_{i-1}E_i - \beta_i(t)E_i + \alpha_i E_{i+1} \tag{3}$$

...

$$\frac{dE_0}{dt} = \beta_{-1}(t)E_{-1} - \alpha_{-1}E_0 \tag{4}$$

The rate constant α_i describes the closing transition. The term $\beta_i(t)$ represents an NFκB-dependent opening transition that varies with time as nuclear NFκB concentration varies with time. Experimental studies have demonstrated that NFκB can interact with nucleosomal DNA

(Lone et al., 2013b) while the lowest energy structure of the NFκB -DNA complex is sterically incompatible with DNA-histone octamer interactions (Chen et al., 1998; Suto et al., 2000), suggesting that NFκB can promote nucleosome unwrapping. The opening transition is thus formulated as:

$$\beta_i(t) = \frac{k_i[NF\kappa B(t)]^n}{K_D^n + [NF\kappa B(t)]^n}$$

where K_D is the dissociation constant of the NFκB -DNA interaction, n is the Hill coefficient, k_i is the transition rate constant, and [NFκB*(t)*] represents the time-dependent nuclear NFκB concentration.

Model parameters: Parameters were selected to preserve the ratio of biophysically determined nucleosome unwrapping and rewrapping rates. As the initial first step rewrapping rate is approximately 5-10 times faster than the unwrapping rate, we set the ratio of rewrapping to unwrapping to 7.5 (Tims et al., 2011a). As *in vivo* nucleosomes are stabilized by linker histones and cooperative binding within the nucleosome array, we set transition rates to be 50-fold slower, with the unwrapping rate constant of the first opening step k_{-14} set to 10 $\mu M^{-1} min^{-1}$, and the rewrapping rate constant α_{-14} to 75 min^{-1}. To account for the inherent cooperativity of contact points within a nucleosome (Miller and Widom, 2003a), unwrapping rates increase by a factor of 1.2 while rewrapping rates decrease by a factor of 0.8 with each step of nucleosome unwrapping; these stepwise factors were examined and selected empirically through a parameter sweep. Varying K_D and n within reasonable ranges led to qualitatively similar results. For the shown simulations K_D was set to 0.025 μM and the Hill coefficient was set to 3.

Model simulations: Simulations were performed in MATLAB R2014b. Experimental and theoretical single cell traces (Supplemental Table 1) were used as input to the ODE system described above. Experimental values of mVenus-RelA fluorescent intensities were converted to μM concentrations of NFκB based on previously published NFκB

models, with the maximum fluorescence of the first peak for the single cell trajectories approximating the maximal nuclear NFκB concentration of 0.25 μM reported in previous studies (Shih et al., 2012). For simulations, the initial value of the E_{-14} state was set to 1, and all other states set to 0. The MATLAB function *ode*15*s* was used to solve the ODE system, and the concentration of the most open chromatin state E_0 was plotted.

3.2. Stochastic models of nucleosome dynamics reveal regulatory rules of stimulus-induced epigenome remodeling

ABSTRACT

The genomic positions of nucleosomes are a defining feature of the epigenomic state and hence of cell identity. Signal-dependent transcription factors (SDTFs), upon activation, modify the positioning of nucleosomes and cause epigenome remodeling. Here, we developed Markov models of nucleosome wrapping and unwrapping, and fit them to high resolution deep sequencing data of DNA accessibility to reveal biophysical principles of nucleosome dynamics. We found that 1) the dynamics of DNA unwrapping are significantly slower *in vivo* than reported from *in vitro* experimental data, 2) there is clear evidence for cooperativity in wrapping and unwrapping, 3) SDTF activity produced highest eviction probability when its binding site is close to but not on top of the nucleosome dyad, and 4) oscillatory SDTFs produce more variability than constant SDTF activities. Our work uncovers the regulatory rules governing nucleosome dynamics *in vivo,* which can predict epigenomic alterations during inflammation at single nucleosome resolution.

Keywords: nucleosome dynamics, stochastic model, ATAC-seq, signal-dependent transcription factor, NFκB, histone eviction, random walk, time-dependent Markov model, cooperativity

HIGHLIGHTS

- Markov models of nucleosome-SDTF interactions fit stimulus response ATAC-seq data
- *In vivo* unwrapping may occur in minutes, rather than seconds as in *in vitro*
- Models reveal cooperativity and estimate an SDTF effect range of 30-40 base pairs
- SDTF sites close to but not on the dyad evict nucleosomes with greatest likelihood

INTRODUCTION

Nucleosomes are critical to packaging the eukaryotic genome into the nucleus: 2m of human DNA must be packed into a 1μm nucleus (Alberts et al., 2002). As a consequence of packing, access to the DNA is limited, yet selective access is important for gene expression (Allfrey et al., 1963). Hence nucleosomes have evolved to be highly dynamic. Dynamic nucleosome repositioning, including histone assembly, disassembly, and eviction, are important for generating dynamic chromatin states that are ultimately permissive or non-permissive to gene expression (Lee et al., 2004; Shivaswamy et al., 2008).

Biophysical *in vitro* studies of histone octamer-DNA interactions started from the 1970's, with high resolution studies of static interactions (culminating in x-ray or cryo-EM structures), and dynamic interactions *in vitro,* via nucleosomal DNA sequences bound to reconstituted histones. Early high-resolution structures elucidated the interaction points between the histone octamer (H2A-H2B pairs and H3-H4 pairs) and the DNA wrapped around it (Luger et al., 1997b). Experiments on nucleosomes reconstituted *in vitro* helped quantify the rates of nucleosome unwrapping and rewrapping *in vitro,* via a variety of methods such as FRET, which reveal quantities such as the average time taken for spontaneous DNA unwrapping (Li et al., 2005), and the differences in timescales of dissociation and reassociation of the different DNA-histone contact regions on the nucleosome (Tims et al. 2011). Studies on asymmetry in DNA wrapping also revealed that the unwrapping of one side helped to stabilize the other side (Ngo et al., 2015).

Mathematical models for the dynamic behavior of nucleosomes *in vitro* have been explored in the literature. Previous mathematical modeling efforts for nucleosome behavior have investigated different aspects of chromatin biology, including the effect of chromatin remodeling proteins on nucleosome sliding (Chou, 2007) and the deposition of histone marks along nucleosome arrays for epigenetic memory (Dodd et al., 2007). Both Markov models and

hidden Markov models were employed to describe nucleosome dynamic under a force (Kruithof and van Noort, 2009; van Opheusden et al., 2013) as well as DNA unzipping (Forties et al., 2011). Nucleosomes have also been modeled with biophysical accuracy by incorporating the nucleosomal structure of 14 DNA-histone contact points and describing how DNA wrapping/rewrapping depends on the rate parameters (Cheng et al., 2021; Dobrovolskaia and Arya, 2012; Möbius et al., 2006b). These theoretical approaches show that a mathematical model, especially involving Markov chains and Brownian motions, can be used to reproduce *in vitro* experimental measurements and to provide a qualitative analysis of the chromatin system such as an analytic form of the mean DNA detachment time, DNA bending angles, and bistability in histone modifications.

However, little is known about nucleosome dynamics *in vivo*, as they occur on native chromatin. *In vivo* dynamics are likely markedly different from dynamics measured *in vitro* because the interactions between DNA polymer and histone octamer are constrained, and because additional protein factors that are not present in biochemical studies may further stabilize or destabilize the nucleosome. For example, enzymatic machines such as SWI/SNF (Dechassa et al., 2010) or FACT complexes facilitate nucleosome repositioning (Chen et al., 2018; Liu et al., 2020). However, we know little about these dynamics quantitatively because there has not been a straightforward way to measure nucleosome positioning *in vivo* and no controlled way to perturb steady state positions.

Recent advances have allowed us to probe nucleosome dynamics. First, next generation sequencing (NGS) has provided ways to measure nucleosome accessibility and positioning with DNAse1, and more recently with ATAC-seq. These genome-wide measurements revealed that nucleosome positions *in vivo* are to a large degree determined by DNA sequence which provides a propensity for wrapping around the core (Segal et al., 2006). Second, the activation of proteins that may displace nucleosomes by competing with histone for DNA

contacts provide a means to perturb nucleosomes at a given timepoint. Some DNA binding proteins have higher propensity to affect nucleosome positioning, thereby increasing chromatin accessibility that is key to enhancer function or gene activation (Fernandez Garcia et al., 2019). Such factors have been termed pioneer factors, and in some specific peptide domains they have been shown to destabilize histone octamer-DNA interactions. The discovery that stimulus-induced or signal-dependent transcription factors (SDTFs) may also trigger nucleosome re-positioning now allows them to be used as a probe to study *in vivo* dynamics, as they provide a trigger to perturb DNA-histone interactions within the cell at controllable start times (Ostuni et al., 2013a; Sen et al., 2020; Weinmann et al., 1999).

The present work is prompted by recent reports that in the case of SDTF-induced nucleosome positioning, it is the dynamics of SDTF activity that determines the propensity of repositioning nucleosomes (Sen et al, 2020; Cheng et al, 2020). This suggests that SDTF activation with stimulus-specific dynamics may be used as a probe to study the interaction dynamics within the nucleosome, via NGS measurements at stimulus start and end points.

We present a stochastic model based on structural features of the nucleosome to investigate the regulatory rules behind nucleosome eviction. Using probability theory, we calculated the probability of histone eviction and mean chromatin accessibility under various dynamical SDTF signaling patterns. We found that oscillatory SDTF signals potentially induced greater variability of cell fate in heterogeneous cell environments than constant SDTF signals. Then, by experimentally tracking nucleosomes at different genomic locations and counting the number of nucleosome evictions between two time points, we found that optimal eviction takes place when the SDTF binds adjacent to the dyad, defined as the center position of nucleosomal DNA, rather than directly on top of it. This surprising experimental finding can be understood as evidence of cooperativity in nucleosome unwrapping rates, because in the absence of cooperativity the maximal eviction takes place when SDTF binds at the dyad

location. Furthermore, the model using the fitted parameters predicted that SDTF binding disrupts histone-DNA contacts about 30-40 base pairs around its binding location. We predicted that the initial unwrapping event occurs after approximately 7 minutes on average, which is significantly faster than the corresponding *in vivo* measurement. Despite the slow initial unwrapping time, cooperativity and SDTF binding allow the system to completely unwrap with at least 70% likelihood after 4 hours. Our stochastic model also reproduces nucleosome eviction timecourse data shown in previous stimulus-response ATAC-seq studies (Cheng et al., 2021; Sen et al., 2020). Our modeling approach provides a framework for understanding location specific, SDTF-induced epigenomic change in different cellular contexts, and it constitutes a tool to predict eviction probability for single *in vivo* nucleosomes responding to inflammation.

RESULTS

A stochastic model accounts for nucleosome dynamics upon SDTF binding *in vivo*

When SDTFs bind to DNA, their specific stimulus-specific temporal dynamics disrupt the resting state distribution of nucleosomes, affecting chromatin accessibility (Figure 1A). Epigenetic dynamics can be modeled as a continuous system - for example, deterministic ordinary differential equation models describing chromatin accessibility in bulk have been used to describe chromatin opening steps that result in enhancer formation (Cheng et al., 2021). However, for a single nucleosome, intrinsic noise plays a significant role as DNA unwrapping/rewrapping is noisy in the highly dynamic environment of a nucleus. Moreover, the binding of SDTFs to DNA can be regarded as a time-dependent on/off switch dramatically influencing chromatin dynamics – this binding is discrete and stochastic. To incorporate such noisy behavior and discreteness, we used a continuous time, discrete-state Markov chain to

model chromatin accessibility with time-dependent SDTF binding. This model is time-inhomogeneous as the transitions given by SDTF binding/unbinding are time-dependent (SI Section S1).

To reflect the biophysical structure of the nucleosome, we assumed that each nucleosome consists of 14 step-wise unwrapping and rewrapping transitions, consistent with structural data on the number of contact points between the histone and DNA (Luger et al., 1997b), as well as previous nucleosome unwrapping models (Figure 1B) (Cheng et al., 2021; Mobius et al., 2013). Approximately 147bp of DNA wrap one and three-quarter times around the core histone octamer (Luger et al., 2012), resulting in 14 main non-covalent DNA-histone contact points (Luger et al., 1997). In order to fully displace the nucleosome from any particular genomic location, multiple steps may be required.

Hence, based on structural and biophysical measurements performed on single nucleosomes in vitro, we used a coupled stochastic process $\big(X(t), N(t)\big)$, where $X(t)$ represents the number of disassembled DNA-histone contact regions, and N(t) takes either 0 or 1 to represent the on/off state of the SDTF binding (Figure 1C). We considered unwrapping behavior of a single nucleosome rather than slipping or sliding behavior, which requires additional ATP-dependent chromatin remodeling complexes that have also been previously modeled (Chou, 2007; Mueller-Planitz et al., 2013). The spontaneous wrapping and unwrapping of the DNA from the histone occur at the locations furthest from the dyad (state 7). Regarding the symmetry of the model, we assumed DNA unpacks from state 0. Although multiple studies have indicated that the unwrapping and wrapping of the nucleosome is primarily unidirectional (Bilokapic et al., 2018; Li et al., 2005a; Ngo et al., 2015), we also considered the possibility that this unwrapping and wrapping takes place simultaneously at both ends of the DNA. We analyzed a model of this phenomenon using two random walkers, corresponding to each edge of DNA (SI Section S1.1), and we found that the qualitative

behavior of both 1-sided and 2-sided stochastic models were similar. Hence in the main results of this paper, we use the 1-sided model.

The amount of energy released by re-establishing hydrogen bonds between histone and DNA is greater than the energy released by the straightening of the DNA polymer during unwrapping, so the rates of rewrapping exceed that of unwrapping, which in our model corresponds to setting $a_i < b_i$ (Tims et al., 2011b). We set the unwrapping/rewrapping parameters as $a_n = a_1 h^{n-1}[min^{-1}]$ and $b_n = b_1 h^{-n+1}[min^{-1}]$ with a cooperativity constant h so that DNA unwraps more easily around the right edge (state 14). Biophysical and structural measurements on single nucleosomes support the cooperative and multistep transitions in DNA unwrapping from the histone (Li et al., 2005a; Polach and Widom, 1995; Tims et al., 2011b), but the extent of such cooperativity remains a free parameter that can be later fit to data. We note that evidence for cooperativity in the literature is measured in isolated nucleosomes *in vitro*, whereas our measurements below were carried out within the full cellular chromatin environment.

We then considered the effect of a dynamic signaling protein that competes for DNA binding with the histone core octamer. Short periods of DNA accessibility may be stabilized by the binding of transcription factors if their cognate binding sequence is present in that stretch of DNA and they are present at sufficiently high concentrations (Klemm et al., 2019). Spontaneous nucleosome dynamics allow transient exposure of nucleosomal DNA, and the binding of SDTFs provide steric hindrance that occludes the rewrapping of DNA-histone contacts within the nucleosome. The on-state of SDTF makes the nucleosome rewrapping rate d_n's much less than b_n around the SDTF binding state (Figure 1C), while c_n is set to be identical to a_n. Once a histone is fully evicted it detaches entirely from the DNA and might not dock to the same genomic location again. Thus, we assumed that $b_{14} = d_{14} = 0$ so that

state 14 is an absorbing state of $X(t)$. That is, if $X(s) = 14$ for some s, then $X(t) = 14$ for all $t > s$.

It is known that transcription factor binding operates at a faster timescale than DNA wrapping or unwrapping (Callegari et al., 2019). Hence for a given SDTF concentration $f(t)$, we used the SDTF binding rate $\kappa_{on}(t) = cf(t)$ with a large constant c, and the unbinding rate κ_{off} is proportional to $\kappa_{on}(0)$. Indeed, the stochastic system behaves almost identically with any choice of large c, and this was proved in the Supplementary Information using a timescale decomposition argument (SI Section S1.3). For large values of *c,* the ratio $BF = \kappa_{off}/(\kappa_{on} + \kappa_{off})$ approximately determines the fraction of time that the SDTF is unbound. See Methods and Table 1 for a rigorous definition of the stochastic model, the definition of the parameters, and the choice of parameter values.

In further difference from previous models, we take into account the SDTF binding position in relation to the original nucleosome dyad. As the nucleosome encompasses ~147 base pairs of DNA, and SDTF binding motifs typically stretch 8-10 base pairs (Stewart et al., 2012), the stochastic binding and unbinding of the SDTF from DNA at the site of its motif is modeled with genome-specific resolution by incorporating the relative location of binding motifs from the nucleosome dyad. When the SDTF binds to its cognate motif, it tends to disrupt DNA-histone contacts in its vicinity. The effect of SDTF binding on the rewrapping rate is highest near the SDTF binding site and decreases with distance. We write the new rewrapping rate of each step as $d_n = b_n(1 - exp\left(-\frac{(n-s)^2}{r}\right))$, where s is the position of the binding motif and r represents the range of SDTF effect. This similarity between the structure of the stochastic model and what can be measured in biological experiments will allow us to evaluate theoretical predictions using genomic sequencing data of *in vivo* cellular immune response systems.

The stochastic system behaves qualitatively differently depending on the temporal patterns of SDTF activity, the unwrapping/rewrapping parameters, and biophysical conditions such as the locations of SDTF binding sites or the range of SDTF influence (Figure 1D). Since $X(t)$ represents the accessibility state at any time t, the mean nucleosome accessibility can be computed as $\sum_{n=0}^{14} n\text{Prob}(X(t) = n)$. Also, since full eviction is reached at state 14, one can compute the probability of eviction by time T by calculating $\text{Prob}(X(T) = 14)$. See SI Section S1.2 for a mathematical derivation of these quantities from the stochastic process $\big(X(t), N(t)\big)$.

Innate immune cells alter nucleosome positioning when responding to immune threats

When innate immune cells respond to immune threats, activation of SDTFs may remodel chromatin to varying degrees across genomic locations. To identify predictions of the nucleosome model that could potentially be tested with experimental data, we analyzed ATAC-seq data for both chromatin accessibility and nucleosome positioning (Buenrostro et al., 2013). Since a central SDTF activated during immune responses is NFκB, we used ATAC-seq data at a baseline time point when no stimulus has been introduced and no NFκB activation has yet occurred, and at a second time point 4 hours later, after NFκB has been activated, which is evident in both single cell traces and populations of cells (Figure 2A). We investigated two previously published experimental systems, mouse bone-marrow derived macrophages (BMDMs) with oscillatory SDTF activity (WT) or mutant cells with non-oscillatory activity (Mut) (Adelaja et al., 2021; Cheng et al., 2021). Evidence from this experimental system allows us to evaluate which analytical results from the model may be true, which in turn suggest mechanisms operating to control nucleosome dynamics and the associated parameter regimes.

In order to compare the model with experimental data we examined several properties of the chromatin locations. First, the two experimental systems displayed differences in variance in the amount of post-stimulation chromatin accessibility among genomic locations (Figure 2B). Second, comparing the two conditions, WT and Mut, allowed us to assess the distribution of stimulus-induced fold changes for each genomic location that are attributable to differences in signaling dynamics (Figure 2C). We hypothesized that these calculated experimental distributions across genomic locations would allow us to analyze the stochastic model to find kinetic rules that allowed the model to match the distributions seen in the data. Indeed, under fitted parameter values (Table 1), simulations of the stochastic model are able to reproduce the distributions observed in the experimental measurements (Figure 2D-E). It has recently been established that nucleosome eviction is likely to take place under a long NFκB signal pulse of approximately 120 minutes, but that it rarely occurs under a shorter NFκB signal pulse of less than 45 minutes (Cheng et al., 2021). Similar observations were made for another experimental system after 60 minutes and 150 minutes, respectively (Sen et al., 2020). These observations can be reproduced with our stochastic model under the fitted parameters (Figure 2F). This comparison with experimental data constitutes a form of validation of the dynamic rates of DNA wrapping and unwrapping in the model.

Periodicity of SDTF oscillations affects DNA accessibility

In inflammation signaling, the importance of signaling dynamics is well appreciated (Werner, 2005, Purvis & Lahav, 2013). For example, for NFκB signaling, the amplitude (Lee et al., 2014) and duration (Hoffmann et al., 2002; Sen et al., 2020) of the signal controls which genes are activated. However, only recently has the importance of oscillatory versus non-oscillatory signaling been studied (Cheng et al 2020), and interestingly, the effect of non-oscillatory

signaling is primarily in remodeling the epigenome rather than in gene expression itself (Barken, 2005). *In vivo*, it is challenging to alter the period of NFκB oscillations. Thus, we used the stochastic model to examine how the period of SDTF oscillations alters the chromatin accessibility; we analyzed the results of numerical computations with the probability distribution of the full histone eviction time.

The period of the oscillation quantitatively affects the time-course dynamics of chromatin accessibility (Figure 3A-C). We set the cooperativity constant $h = 1.1$, and we set the unwrapping/rewrapping parameters as $a_n = 0.1\, h^{n-1},\ b_n = 0.6\, h^{-n+1}$ and $d_n = 0$ for each state n in the stochastic nucleosome model. We consider two oscillatory SDTF inputs of 60 min and 180 min half-periods, respectively, that have the same aggregate signal within the time interval $[0,720]$ min. We sampled 50 timecourses of our stochastic model under each of these two oscillatory inputs, using the Extrande method (Voliotis et al, 2016), which is a stochastic simulation algorithm for Markov chains with time-dependent transition rates. The rapid oscillatory SDTF signal of half-period 60 min unpacks the nucleosome completely in 13 out of 50 samples within 720 min, while 30 samples are fully unpacked by 720 min under the slow oscillatory SDTF of half-period 180 min. This result is consistent with the experimental results of oscillatory NFκB dynamics stimulated with TNF.

In order to investigate the discrepancy between the system modeled under the two different oscillatory SDTF signals, we describe the DNA wrapping process as a 'success-or-failure game', which can be analyzed with a geometric distribution. In the case of a cooperative system with $h = 1.1$, when $X(t)$ reaches state 10 or above, the unwrapping rates a_{n+1} are comparable to the rewrapping rates b_n so that $X(t)$ can easily reach state 14 (state of full eviction) even without the support of SDTF binding. Hence success of $X(t)$ is reaching state 10, and we use the probability of the success to analyze the distinct behaviors of DNA under two oscillatory inputs.

If the nucleosome is exposed to an SDTF signal at level 10 for 60 min, then only about 5% of DNA samples reach state 10 (Figure S3B). Hence during the on phase (i.e. SDTF signal is at level 10), nearly 5% of DNA segments can successfully unwrap the entire histone octamer under this rapid oscillation. Moreover after 60 min, when the SDTF signal is turned off, then most remaining DNA, which failed to reach state 10 during the previous on-phases, rapidly rewraps around the histone because the rewrapping rate b_n is much greater than the unwrapping rate a_n for $n < 10$, likely returning back to state 0. Therefore on the next on-phase, about 5% of the remaining free DNA can be fully wrapped, and DNA undergoes this process 6 times by 720 min. This success-or-failure game under the oscillatory SDTF signal can be described using the geometric distribution $Geo(0.05)$ with the success probability 0.05. The full eviction probability by 720 min under the rapid oscillation is approximately

$$Prob(X(720) = 14) \approx Prob(Geo(0.05) \leq 6) = \sum_{i=1}^{6} 0.05(1 - 0.05)^{i-1} = 0.2649.$$

Hence after the SDTF signal turns on and off repeatedly every 60 min, about 26% of chromatin is accessible at 720 min (Figure 3B-C).

On the other hand, if the SDTF signal is maintained at the level 10 for 180 min, about 35% of DNA can reach state 10 (Figure S3C). Although DNA rapidly rewraps the histone over the next turned off phase of the SDTF signal, once the SDTF is turned on again for the next 180 min, about 35% of remaining DNA will be unpacked. Through the success-or-failure game within [0,720] min twice, about $58\% = Prob(Geo(0.36) \leq 2) \times 100\% = \big(0.35 + 0.35(1 - 0.35)\big) \times 100\%$ of DNA is fully unpacked (Figure 3B-C).

Faster oscillation of the SDTF signal does not necessarily make the DNA less accessible. We consider an extremely fast oscillation with half-period 0.6 sec (Figure 3A bottom). Since each phase of the SDTF signal is extremely short, the SDTF dynamics is

averaged out in time so that it is decoded by the nucleosome as a constant signal at level 5, which is the half of the amplitude. Therefore, despite the extremely short on-phase of the oscillation, about 33% of DNA temporal trajectories are fully unwrapped by 720 min (Figure 3B bottom), which is higher proportion than the case of half-period 60 min. The time evolutions of probability of histone eviction under the three different SDTF signals are displayed in Figure 3C. See SI Section S2.1 and Figure S3 for more detailed mathematical analysis about the full eviction probability under different frequencies of the SDTF signal.

Oscillatory SDTF inputs can lead to heterogeneous chromatin accessibility responses

Although above we investigated the effect of different SDTF oscillatory frequencies, in cellular conditions, the measured period of oscillation of a certain SDTF such as NFκB is actually consistent across cell types and experimental systems, and it is also hardwired by the IκBα-NFκB negative feedback loop (Longo et al., 2013). However, given the same oscillatory signal, each nucleosome within a cell is exposed to a potentially different environment. Hence different regions of chromatin might have different model parameters.

To abstract these molecular mechanisms affecting the differential responses of various chromatin regions to the same dynamic signal, we scanned the nucleosome unwrapping/rewrapping rate parameters and computed the probability of histone eviction using the stochastic model under non-oscillatory input or oscillatory input with a fixed period. We then highlighted that the system can be more sensitive to the unwrapping/rewrapping rates under an oscillatory SDTF. For the computation of the full eviction probability, we set the parameters to $h = 1.1$, $a_n = 0.05\, h^{n-1}$ and $b_n = 0.25\, h^{-n+1}$ in the stochastic nucleosome model. For simplicity, we further assumed that the rewrapping rate d_n with SDTF bound is 0 for each n.

Under oscillatory and constant SDTF dynamics (Figure 3D), we calculated the probability of histone eviction at T=360 minutes after multiplying each of the unwrapping/rewrapping rates a_n and b_n by a fold-change parameter m (Figure 3E left). The full DNA eviction probability under oscillatory SDTF dynamics rapidly grows for $m \in [2,4]$. In fact this graph has a sigmoidal shape, while it grows more evenly under the constant SDTF input. A more sigmoidal shape is indicative of a higher sensitivity with respect to fold change increases, so that the same oscillatory input can lead to widely different responses for different parameter values. The full eviction probability grows more evenly when the SDTF dynamics is constant, or for higher values of the cooperativity parameter (Figure 3E right). In SI Section S2.2, using simple matrix exponentials, we explored sensitivity analysis with our stochastic model under both constant transition rates and time-dependent oscillatory transition rates.

This sensitivity potentially implies that the greater variability of chromatin accessibility under an oscillatory SDTF input allows more variability of cell fates. For instance, if the cell type is determined by a threshold mean accessibility, then cells with oscillatory SDTF input can be distributed between type A and type B (Figure 3F). However, cells with constant SDTF input may more consistently convert to a single cell type B (through changes in chromatin state) because of the narrower range of induced chromatin accessibility changes.

Eviction probability profile using SDTF binding location determines cooperativity

In addition to the relation between signaling patterns and chromatin accessibility, we next sought to use the nucleosome model to investigate how the location of the SDTF binding in relation to the nucleosome might control nucleosome eviction in accessible regions. We thus re-sequenced the ATAC-seq data using paired-end sequencing to separate nucleosomal read fragments from nucleosome-free read fragments, and calculated nucleosome positions across

the genome (Schep et al., 2015) for the IκBα knockout mutant experimental system that displayed significant changes in chromatin accessibility upon TNF stimulation (Cheng et al., 2021). We focused our analysis on nucleosomes near NFκB binding motifs, which are potentially affected by NFκB binding. We found that nucleosome eviction, estimated by assigning nucleosome dyads to the closest transcription start site (TSS), and reassessing the locations of dyads assigned to the same TSS at a later timepoint, appeared to have some dependence on distance between NFκB binding motif location and the nucleosome dyad (Figure 4A).

To understand the mechanisms behind how the location of the binding motif in relation to the nucleosome position affects nucleosome eviction, we next added additional detail to the model**:** We explored the probability for complete DNA unwrapping as a function of the SDTF binding site location along the 147 base pair stretch of DNA that encompasses the nucleosome. We assumed that SDTF binding locally alters the rewrapping rates. Hence only the sites lying within a certain range around the SDTF binding site have a rewrapping rate d_n that differs from b_n as $d\_n = b\_n(1 - exp(1 - (s - n)^2/2\sigma^2)$, where s is the SDTF binding location and σ represents the standard deviation of the SDTF effect range (Figure 4B). That is, the SDTF binding effect propagates from the SDTF binding location, and the effect range is determined with the parameter σ indicating how many DNA-histone contact sites could be disrupted by SDTF binding (Figure 4C). We then tested which binding location is optimal in order to open the nucleosome under a constant temporal SDTF signal. SDTF binding motifs are distributed across the DNA strand in the range [-100 bps, 100 bps] centered at histone dyad (state 7 in Figure 4B). We assume that a SDTF binds at one of the states s in {-3,-2,…,16,17}, which is an extended range from the original state space {0,1,…,14} (Figure 1C), in order to consider SDTF binding motifs lying slightly outside the nucleosome. Then for each

distance relative to the histone dyad, computed as $|7 - s| \times 10$ (bps), we measure the full DNA eviction probability.

We assumed that either the unwrapping rates of the nucleosome are cooperative with differences in the degree of cooperativity, or non-cooperative. In other words, for these two main hypotheses, each step may have the same unwrapping and rewrapping rates, or the unwrapping/rewrapping rates may change as the nucleosome becomes more open/more closed. Prior evidence suggests that the parameters of the epigenome model measured *in vitro* are highly cooperative with the open rate increasing and the close rate decreasing as the DNA unwraps (Figure 4E left) (Miller and Widom, 2003b). To incorporate cooperativity of the rate, we have set $a_n = a_1 h^{n-1}$ and $b_n = b_1 h^{-n+1}$ for each n with the cooperativity constant h. A special case of $h = 1$ indicates that the parameters are not cooperative.

The resulting behaviors under various levels of cooperativity of the rates a_n and b_n are distinct as the optimal binding site is either in the center of the nucleosome or toward the extremes. Under non-cooperative rates ($h = 1$), the optimal binding site is at the histone dyad so that the full eviction probability is symmetric about the relative distance (Figure 4D). On the contrary, in case the rates are cooperative ($h > 1$), the optimal site is closer to the unwrapping edge, and hence the full eviction graph has a peak close to this edge (Figure 4E). This is because once the first few contacts between DNA and histone are unwrapped, the cooperativity of the system facilitates the unwrapping of the rest of the histone. After averaging multiple cells, due to the symmetry of nucleosomes unwrapping from either end, the probability-binding site plot has a center dip (Figure 4E right). In SI Section S3.1, we provide a mathematical analysis of the SDTF binding site effect for the probability of DNA full eviction.

SDTF binding site location provides evidence of cooperativity and estimates range of SDTF effect

The special pattern in the plot of the full eviction probability as a function of SDTF binding site (Figure 4E) is altered not only by nucleosome cooperativity (parameter h) but also by the range of the SDTF binding effect (parameter σ), which is likely to be SDTF-specific. If the range is wider, all the rewrapping rates d_n are equally affected so that the probability plot becomes more flattened. By using this relation between the model parameters and the special pattern in the full DNA eviction probability plot, we can help predict the model parameters using experimental data on the DNA eviction probabilities for each SDTF binding site. The parameter σ corresponds to the standard deviation of the Gaussian curve of influence of the SDTF binding, in units of number of nucleosome binding sites. Since these binding sites are approximately 10 base pairs away from each other, a value of $\sigma = 2$ would correspond to a standard deviation of around 20 base pairs, or a range of 40 base pairs around the SDTF binding site.

We first examined how different values of the cooperativity parameter h and the SDTF range parameter σ could alter the chromatin accessibility (Figure 5A). We tested multiple potential values of this range parameter, as well as cooperativity parameters $h = 1$ (non-cooperative), 1.1 and 1.2 (high cooperativity). It is notable that for higher cooperativity and narrower SDTF binding effect range, the plot of full DNA eviction probability displays a clearer center valley.

To better compare these computational results with the experimental measurements, we returned to the time-course experimental data from cells responding to an inflammatory cytokine stimulus (Figure 4A). As ATAC-seq data can provide an estimate of the nucleosome positions, we assigned nucleosomes to their nearest TSS at the baseline time point, and tracked

whether the nucleosomes matching to the same TSS changed in position or disappeared at the later time point – quantified by having fewer or no nucleosomes mapping to that TSS. By using this approach to match the nucleosomes across 0hr and 4hr time points (see further detail in Methods), we calculated how the motif location in relation to the nucleosome dyad affects the counts of remaining nucleosomes after 4 hours. Using the difference of the nucleosome counts between two time points, we computed experimental full DNA eviction probability for each relative distance by 240 min as

$$\text{Prob}(X(240)=14) = 1 - \text{Prob}(X(240) < 14) \approx 1 - \frac{\#\ of\ nucleosome\ at\ 4hr}{\#\ of\ nucleosome\ at\ 0hr}.$$

We used this data to fit values of the different parameters in our model using a deductive approach together with the gradient descent search algorithm, including the cooperativity parameter, SDTF range parameter, unwrapping/rewrapping rates and SDTF binding/unbinding rates as parameters to be fit. Fitting these parameters resulted in a probability profile relative to distance to dyad and associated parameters that best matched the data (Figure 5BC). We tested this approach on both the macrophage system, in which IκBα knockout macrophages were treated with TNF for 4 hours, and in a melanoma system, in which melanoma cells were treated with IFNγ for 2 hours. See SI Section S3.2 and Figure S5 for additional details, and Table 1 for the resulting parameter values. We find the nucleosome unwrapping/rewrapping rates are likely cooperative, and that the range of effect of the SDTF is specific to the SDTF activated, as the analysis is restricted to nucleosomes close to the appropriate SDTF motif. Also, in this Table we show that the SDTF range parameters are measured at 1.6 and 2.0 for each of the experimental data sets, respectively. This corresponds to a radius of 16 and 20 base pairs from the SDTF binding site, or a range to 32 and 40 base pairs around the binding site, respectively.

DISCUSSION

Our study pairs stochastic modeling and epigenomic measurements from primary cells *in vivo* to investigate the biophysical regulatory rules of histone-octamer-DNA interactions that determine nucleosome positioning. Using probability theory, we described nucleosome eviction as a success-or-failure game scheme, as DNA has a chance of full eviction only under the on-phase of the SDTF signal. This scheme revealed the role of oscillatory inputs in nucleosome eviction, and heterogeneity in DNA accessibility under oscillatory SDTF dynamics. Nucleosome positioning data provided the nucleosome eviction probability profile as a function of SDTF motif location. By analyzing the probability profile with our stochastic model, we found evidence of cooperativity in the DNA unwrapping steps for two separate *in vivo* experimental systems, which is consistent with previous findings for nucleosome unwrapping *in vitro* (Tims et al., 2011b).

We matched the stochastic model to the experimental eviction probability profile and fitted the nucleosome parameters. Our stochastic model under the fitted parameters revealed quantitative aspects of the nucleosome dynamics: 1) 30-40 base pairs of DNA-histone contacts around the SDTF binding site are disrupted and 2) the expected initial DNA unwrapping time from the fully wrapped state is about 7 minutes. The minute-scale unwrapping rate combined with cooperativity leads to a 70% full eviction probability after 240 minutes of steady NFκB signal, and less than 25% full eviction probability under a 45 minutes NFκB signal. Supportive of this model, these quantitative features of our model are consistent with previous experimental observations (Cheng et al., 2021).

Immune responses are central to multiple physiological and pathological processes, for example as seen in the dysregulated hyper-immune activation in the innate immune responses of some subsets of COVID-19 patients. Here we focus on the epigenome of innate immune macrophages, and show that the modeling approach can be applied to other contexts as well

where cells encounter an inflammatory signal that produces stimulus-induced epigenomic changes (including but not limited to cancer cell plasticity during immunotherapy). For innate immune responses particularly, the variation in the baseline epigenome that results from a prior exposure, rather than variation in genetically-encoded receptors like for T- and B-cells, may be a critical component of innate immune memory and response to future inflammatory threats (Netea et al., 2016). Thus, a predictive mechanistic understanding of how SDTF activity can evict nucleosomes can guide further investigation into epigenomic reprogramming events induced by inflammation.

Naturally, as with all mathematical models, the *in vivo* system is more complex than the model describes, and our model is necessarily an abstraction describing a single piece of the dynamic epigenome that results when mammalian cells encounter an inflammatory threat. However, our model is able to assess several characteristics of nucleosome dynamics that may govern the rules and parameter rates at which nucleosomes are evicted across the epigenome. These predictions help formulate hypotheses that are compared to time-course epigenomic sequencing data, which allows the selection of one of the hypotheses or the establishment of parameter ranges. Notably, the model can be used to evaluate numerous different stimulus-response systems, including those with different SDTFs activated (Calderon et al., 2019), or different cell types and genomic locations that may have different kinetic rates governing the unwrapping and rewrapping of the nucleosome.

Nucleosome dynamics at each location along the genome is influenced by multiple factors, including but not limited to the stiffness of the DNA at each nucleosome position, the histone marks or histone variants that are present, the density of nucleosomes at that region, and the binding motif location in relation to the position of the nucleosome (Brahma and Henikoff, 2020). In addition, both the unwrapping behavior and nucleosome sliding behaviors are possible mechanisms for changes in nucleosome occupancy and positioning. One

interesting question is to determine to what extent nucleosomes can slide along the DNA upon SDTF binding rather than unbinding from the DNA.

The development and parameterization of this mechanistic model has several implications. First, although we used macrophages stimulated with TNF as data to compare the model against, the model may also allow the prediction of how strongly nucleosomes are evicted across the epigenome in response to other inflammatory stimuli that activate different SDTFs with different dynamics. Second, because the relationship of the motif location and nucleosome dyad correlates with eviction probability, given a baseline epigenomic starting state, the model can make a prediction on the probability for nucleosome eviction in a location-dependent manner. Third, the model arrives at biological insights related to the nucleosome parameters themselves: by comparing pre- and post-stimulation nucleosome distributions, we can calculate experimental nucleosome eviction probabilities and fit the model to estimate the degree of cooperativity within the nucleosome and the range of effect of SDTF binding on disrupting nucleosomal contact-points.

This stochastic model describes the nucleosome, which is the fundamental unit of chromatin containing multi-step dynamic processes, and serves as a starting point for describing other epigenomic features (Bilokapic et al., 2018; Eslami-Mossallam et al., 2016; Hall et al., 2009; Henikoff, 2016). Future work incorporating other key elements of nucleosome dynamics, such as the structure of nucleosome arrays and the effect of histone modifications, or behaviors such as nucleosome sliding or rolling, which we have not yet considered here, may reveal further insights. In addition, although here we use an optimization approach to analyze this model topology and initial conditions with respect to the data, model parameters can also be further trained with machine learning approaches that incorporate additional layers of epigenomic data as training data in order for the parameters to incorporate more elements of the epigenomic complexity that exists *in vivo*. Our modeling framework and these further

possibilities support the feasibility of combining biophysically detailed mechanistic models of epigenetic processes, with next generation sequencing epigenome-wide measurements, to characterize kinetic rules controlling cellular responses to inflammation.

ACKNOWLEDGEMENTS

JK and GE were partially supported by NSF grants DMS1763272, DMS1616233 and Simons Foundation grant 594598 (Qing Nie). KMS is supported by the UCLA Medical Scientist Training Program (NIH NIGMS T32 GM008042) and Systems and Integrative Biology Training Grant (T32-GM008185). AH was funded by R01AI127864. We thank Sho Ohta for early access to the macrophage sequencing data and resequencing of the macrophage ATAC-seq libraries. We thank Ashley Kim and Antoni Ribas for early access to the melanoma sequencing data. Sequencing data from this paper was performed with the services of the UCLA Technology Center for Genomics and Bioinformatics Sequencing Core. We also thank the NSF-Simons Center for Multiscale Cell Fate for an Interdisciplinary Opportunity Award that initiated this project.

AUTHOR CONTRIBUTIONS

JK, KMS, AH, and GE conceived the project. JK and KMS developed the model and analyzed the data. JK analyzed the model and performed model computations. KMS performed bioinformatic analyses of the data. QJC provided experimental data and helpful input on the manuscript. JK, KMS, AH, and GE wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURES AND FIGURE LEGENDS

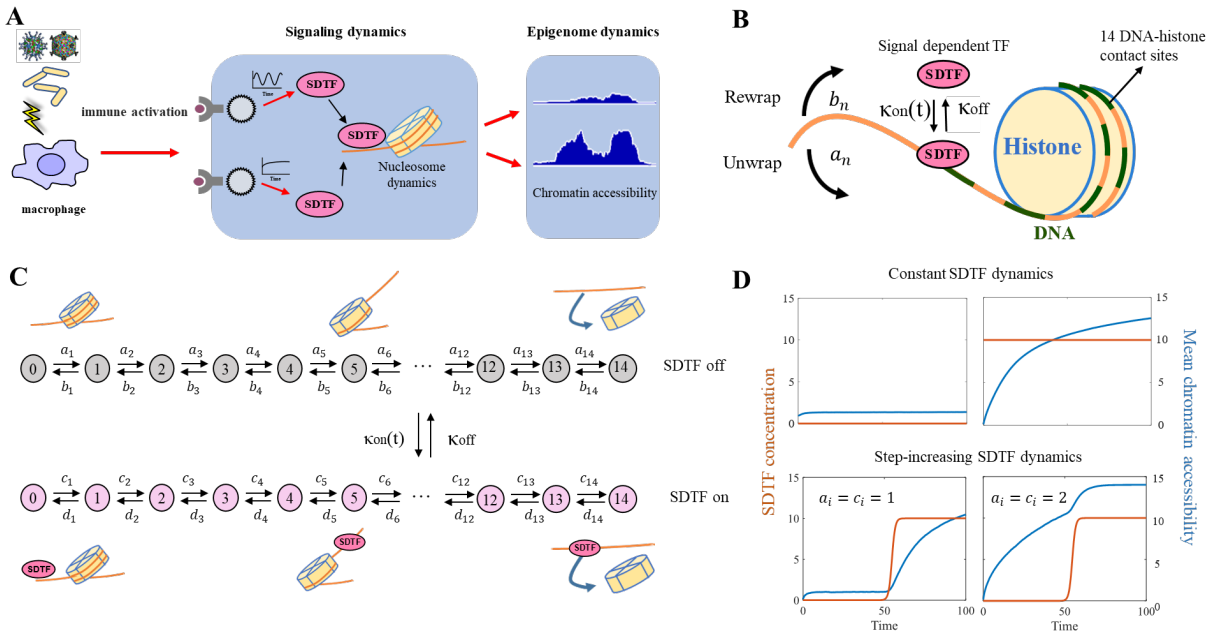


Figure 3.2.1. A stochastic model for chromatin accessibility through nucleosome eviction by dynamic SDTF activity.

A) Immune responses activate SDTFs with different temporal dynamics, ultimately affecting chromatin accessibility. B) Schematic for the unwrapping/rewrapping model for nucleosome dynamics under SDTF signaling dynamics. C) State configuration of the stochastic nucleosome model. D) Illustration of how spontaneous nucleosome dynamics interface with different SDTF dynamical signals, using different choices of parameters. Top: Constant SDTF dynamics at different levels and corresponding chromatin accessibility. The parameters are $c_i = a_i = 1$, $b_i = 2$ and $d_i = 0$. Bottom: Inducible SDTF dynamics and corresponding chromatin accessibility with different unwrapping rates. The unwrapping rates are $c_i = a_i = 1$ for the left panel and $c_i = a_i = 2$ for the right panel. The rewrapping rates are $b_i = 2$, and $d_i = 0$ for i=1,2,…,5 and d_i=2 otherwise, for both panels. (See also Figure S1)

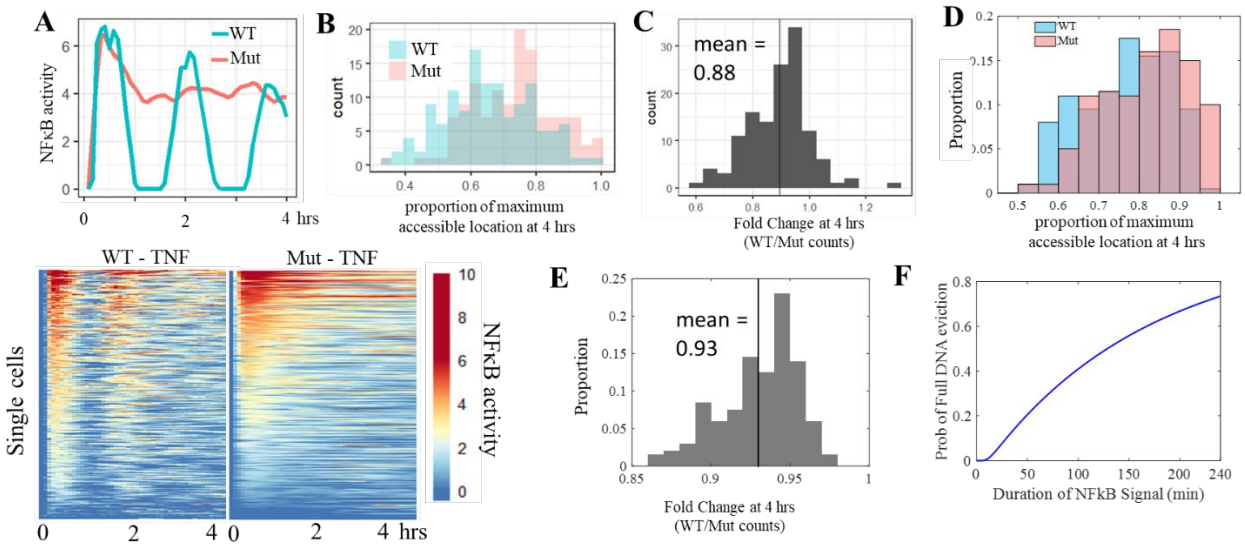


Figure 3.2.2. Cellular signaling and epigenomic measurements show how nucleosome positions change in response to an inflammatory stimulus.

A) Experimental knowledge of SDTF signaling dynamics in single cells (top: two individual single cells, bottom: hundreds of single cells). WT and Mut cells activate NFκB with different temporal dynamics (Adelaja et al., 2021). **B)** Variance in chromatin accessibility across genomic locations at 4 hours in WT and Mut cells, as measured by bulk ATACseq. **C)** Fold change (Mut/WT) (bottom) of resulting chromatin accessibility after activation of SDTFs with different dynamics. **D-E)** Reproduction of the experimental measurements shown in panels B and C, respectively, using the stochastic nucleosome model under the fitted parameters in Table 1. Panel D and E are converted from counts to proportion due to simulation of a different number of nucleosome locations. **F)** The full DNA eviction probability under a steady NFκB input signal using the fitted parameters in Table1.

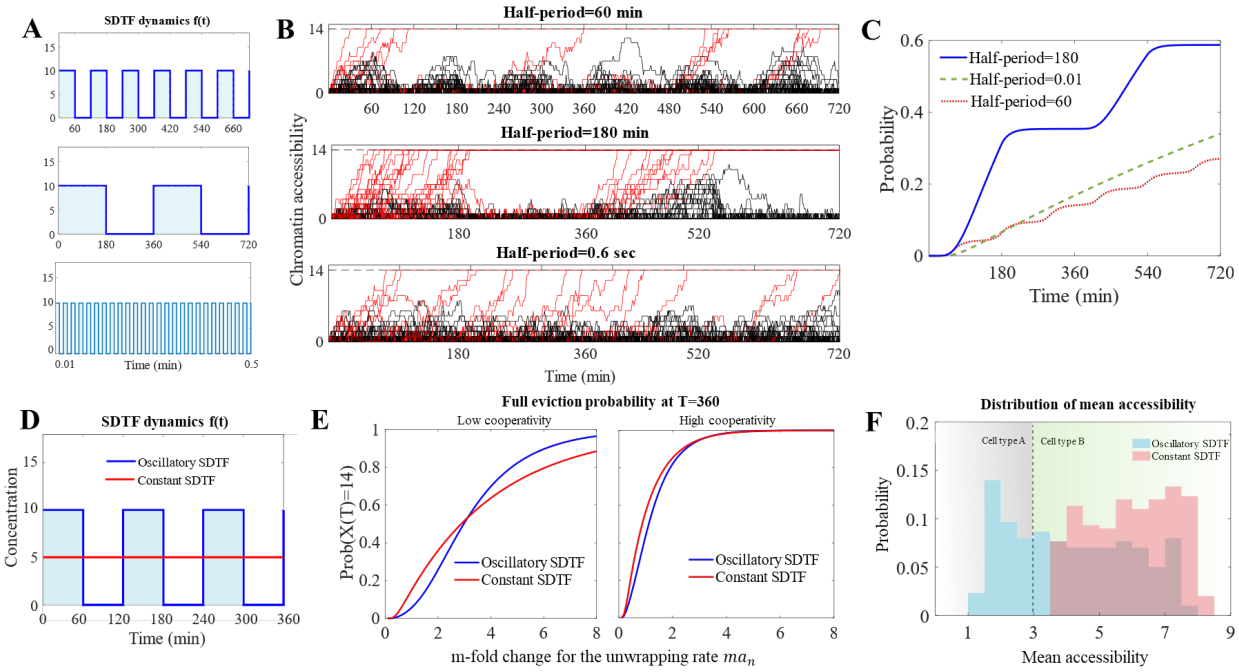


Figure 3.2.3. Chromatin response to oscillatory SDTF dynamics of different frequency.

A) SDTF dynamics with rapid (top) or slow oscillation (bottom). **B)** 50 sample traces of DNA dynamics under the oscillatory SDTF inputs of half-period=60(top), 180(center) and 0.1(bottom). Red traces are simulations that reach the fully evicted state, black traces are those that do not. **C)** Time evolution of histone eviction probability. **D-F) Parameter sensitivity of chromatin accessibility with a fixed SDTF signal. D)** Oscillatory and constant SDTF signal inputs. **E)** Full eviction probability vs unwrap parameter with low cooperativity (h =1.1) and high cooperativity (h =1.2). m represents the fold-change increase in unwrapping/rewrapping rates. **F)** Mean chromatin accessibility distribution at $t = 360$ min with the oscillatory or constant SDTF dynamics. To model heterogeneous cell environment, we randomly perturb the system parameters. Coefficient variation (=mean/standard deviation) of the distributions under oscillatory SDTF and constant SDTF are 0.44 and 0.22, respectively.

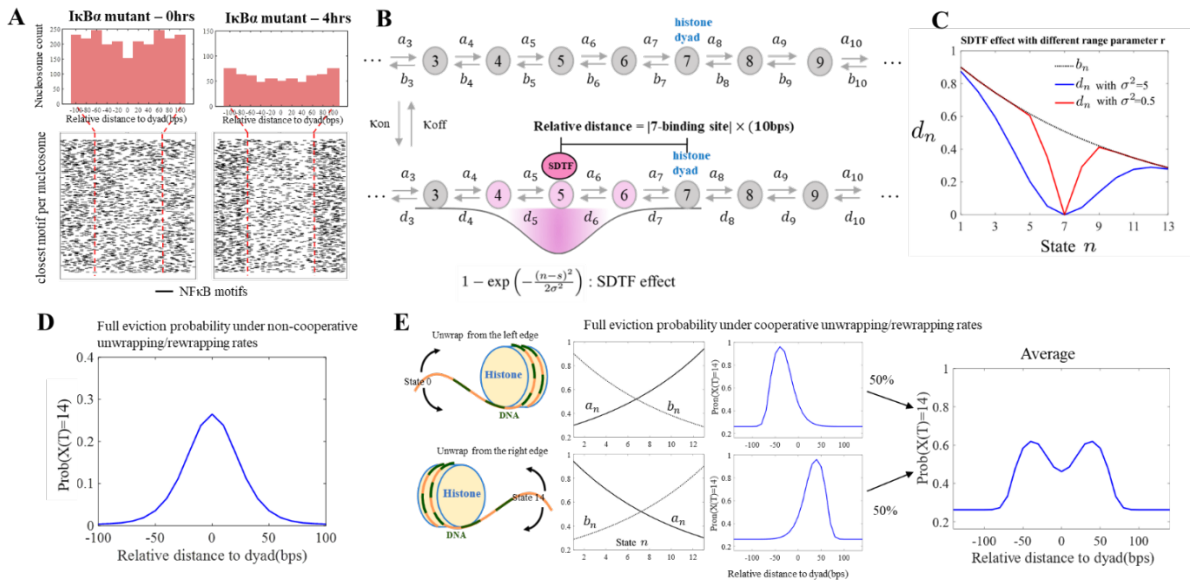


Figure 3.2.4. Modeling SDTF binding sites, range of SDTF effect, and cooperativity in unwrapping steps reveals potential eviction probability profiles.

A. Summary of NFκB motifs adjacent to nucleosome dyads. Mapping of SDTF binding motifs on the nucleosome suggests that SDTF binding location in the relation to the nucleosome, and the range of effect of the binding footprint, may be important for predicting nucleosome eviction. Shown are NFκB motifs in relation to each nucleosome dyad called by NucleoATAC (Schep et al., 2015) at 0hours, and 4hours after TNF treatment, in mouse BMDMs. Locations shown have an NFκB motif +/- 100bp of the nucleosome dyad. **B.** SDTF locally affects the DNA-histone contact regions near by the SDTF binding location. **C.** The range parameter σ determines how widely the SDTF affects the rewrapping rates. **D.** Computation of the full eviction probability via the stochastic model shows that motifs at the dyad promote greater nucleosome unwrapping probability under a non-oscillatory SDTF signal and non-cooperative open/close parameters. **E.** The full eviction probability is maximal at the SDTF binding location between the edge and dyad under cooperative open/close parameters. Assuming 50% of right edge-unpacking and 50% of left edge-unpacking, the average full eviction probability displays a center valley.

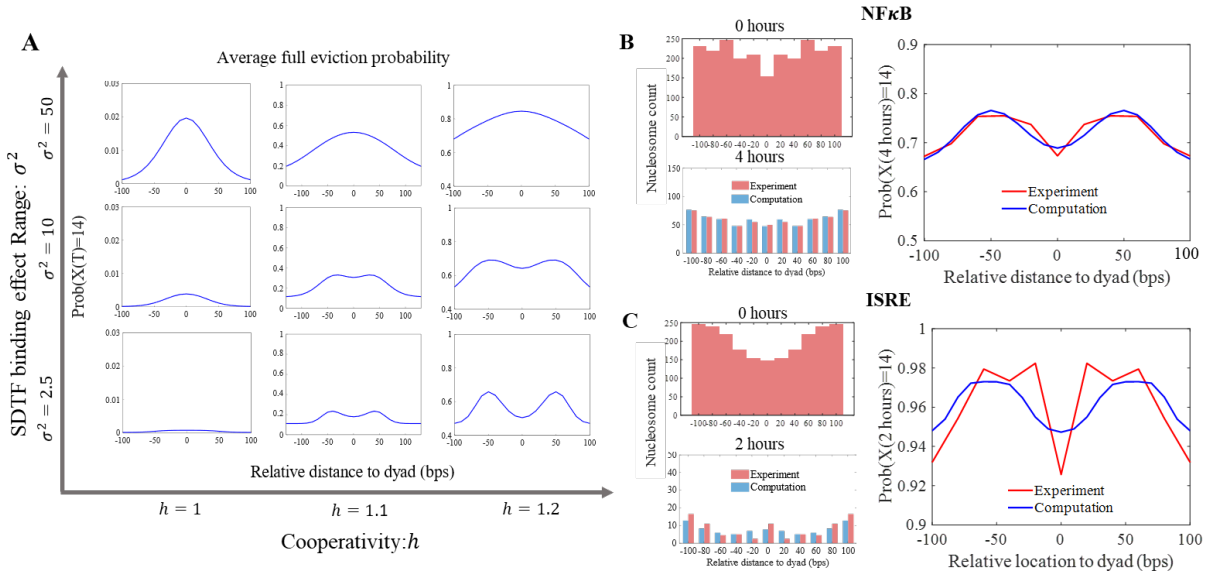


Figure 3.2.5: Fitting the model eviction probability profiles to SDTF binding location data provides evidence of cooperativity and estimates model parameters.

A. Probabilities of full eviction with respect to relative motif position from the nucleosome dyad, and SDTF binding effect range for the macrophage system under non-oscillatory NFκB signal. Three different ranges ($\sigma^2 = 1.5, 10,$ and 50) and cooperativity parameters ($h = 1, 1.1$ and 1.2) are chosen. **B and C.** Experimental measurement of nucleosome counts for macrophage or melanoma system under non-oscillatory TNF-induced NFκB (B left) activity at NFκB motifs or IFNγ-induced IRF1 (C left) activity at ISRE motifs, at 0 hours and 4 or 2 hours, respectively. Full eviction probability-SDTF binding locations, the red curve is calculated by converting the nucleosome counts in the experimental measurement (red curves in the right panel). We identify the relation between system parameters and the distinctive pattern of the full eviction probabilities. We then fit the parameters to reproduce the best-matched probability plot (blue curves in the right panel).

TABLES

<table>
<tr><td rowspan="2">Object</td><td rowspan="2">Symbol</td><td colspan="4">Values</td></tr>
<tr><td>Figure 3B</td><td>Figure 3EF</td><td>Figure 5B (fitted parameters)</td><td>Figure 5C (fitted parameters)</td></tr>
<tr><td>Unwrapping rate when $X = n$</td><td>$a_n = c_n = a_1 h^{(n-1)}$ $[min^{-1}]$</td><td>$a_1 = 0.1$</td><td>$a_1 \in [0,0.5]$</td><td>$a_1 = 0.16$</td><td>$a_1 = 0.27$</td></tr>
<tr><td>Rewrapping rate when $(X, N) = (n, 0)$</td><td>$b_n = b_1 h^{(-n+1)}$ $[min^{-1}]$</td><td>$b_1 = 6a_1$</td><td>$b_1 = 5a_1$</td><td>$b_1 = 5.03a_1$</td><td>$b_1 = 7a_1$</td></tr>
<tr><td rowspan="2">Rewrapping rate when $(X, N) = (n, 1)$</td><td rowspan="2">d_n $[min^{-1}]$</td><td rowspan="2">0</td><td rowspan="2">0</td><td colspan="2">$d_n = b_n \times g_r(n)$</td></tr>
<tr><td>$\sigma = 2$</td><td>$\sigma = 1.6$</td></tr>
<tr><td>Cooperativity constant</td><td>h</td><td>1.1</td><td>1.1 and 1.2</td><td>1.35</td><td>1.89</td></tr>
<tr><td>SDTF binding rate at time t</td><td>c$\kappa_{on}(t)$ $[min^{-1}]$</td><td colspan="2">Binary oscillation between 0 and 100</td><td>100</td><td>100</td></tr>
<tr><td>SDTF unbinding rate</td><td>$c\kappa_{off}$ $[min^{-1}]$</td><td>42.85</td><td>42.85</td><td>245</td><td>43</td></tr>
<tr><td>Time fraction SDTF is unbound</td><td>$BF = \frac{\kappa_{off}}{\kappa_{on}(0) + \kappa_{off}}$</td><td>0.3</td><td>0.3</td><td>0.71</td><td>0.3</td></tr>
</table>

Table 3.2.1. Model parameters used for the panels displayed in Results.

Function $g_r(n) = 1 - exp(-\frac{(n-s)^2}{2\sigma^2})$ determines the range of the SDTF binding effect with range parameter r.

MATERIALS AND METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Macrophage PE ATACseq		GSE156385
Macrophage SE ATACseq	Cheng et al, 2020	GSE146068
Melanoma PE ATACseq	Kim et al, 2020	GSE154483
Software and Algorithms		
NucleoATAC	(Schep et al., 2015)	https://github.com/GreenleafLab/NucleoATAC
HOMER	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/ngs/peakMotifs.html
ENCODE-DCC ATACseq Pipeline		https://github.com/ENCODE-DCC/atac-seq-pipeline

LEAD CONTACT

Further information and requests for resources and materials should be directed to and will be fulfilled by the Lead Contact, German Enciso (enciso@uci.edu).

QUANTIFICATION AND STATISTICAL ANALYSIS

Model simulations

Model implementation and simulations were performed in MATLAB 2016b. Further detailed description of the model can be found in Supplementary Information.

Nucleosome data analysis

ATAC-seq data processing

Macrophage ATAC-seq samples were generated as previously described (Buenrostro et al., 2015), and single-end data was obtained from (Cheng et al., 2021). Macrophage ATAC-seq libraries of the IκBα knockout mouse from Cheng et al, 2020 were re-sequenced paired-end 2x150 on HiSeq4000. Melanoma paired-end ATAC-seq data was obtained from prior publications (Kim et al., 2021). Only paired-end sequencing allows the separation of nucleosomal fragments from non-nucleosomal fragments, as read fragments with lengths shorter than the nucleosome footprint of ~150 basepairs can be classified as nucleosome-free accessible regions, while read fragments of ~150bp, or a multiple of 150bp, can be classified as accessible nucleosomal genomic regions, with cut sites flanking nucleosome boundaries. ATAC-seq fastqs were processed through the ENCODE-DCC ATAC-seq pipeline (https://github.com/ENCODE-DCC/atac-seq-pipeline). The reads were trimmed using cutadapt, and aligned to mm10 or hg38 using bowtie2. Picard was used to de-duplicate reads, which were then filtered for high quality, paired reads using samtools. Peak calling was performed using macs2. The optimal Irreproducible Discovery Rate (IDR) thresholded peak output was used for all downstream analyses, with a threshold p-value of 0.05. Other ENCODE3 parameters were enforced with the flag --encode3. Reads that mapped to mitochondrial genes or blacklisted regions, as defined by the ENCODE pipeline, were removed. The peak files were merged using bedtools merge to create a consensus set of peaks across all samples.

ATAC-seq nucleosome analysis

Nucleosome positions were called using the merged regions, from paired-end ATAC-sequencing data, using the published software NucleoATAC (Schep et al., 2015). An example genomic location *Cxcl2*, illustrates the information obtained is orthogonal to simply chromatin accessibility (Figure S6). The output of this software provides putative nucleosomal and nucleosome-free regions of accessible chromatin, by analyzing the patterns of ATAC-seq read

fragment sizes. As described in full detail in Schep et al, 2015, nucleosome occupancy is called by maximum likelihood estimation, and nucleosome dyad positions are called by considering the local maxima of candidate nucleosome positions. Genomic locations of nucleosome positions called were annotated, and NFκB motifs were found using the tool HOMER (Heinz et al., 2010). Motif searching was done using the three NFκB motif position weight matrices within the HOMER database, for length 9, 10, 11. Motifs were listed if they occurred within +/-200 basepairs of the nucleosome dyad. Nucleosomes across timepoints were matched by assigning them to their closest transcription start site for each sample. Nucleosomes assigned to a TSS for the baseline time point, and subsequently not found at that TSS at the later time point, were considered evicted. For analyses where the model calculated a probability of nucleosome eviction, nucleosomes that appeared, and matched to a new gene at the second time point but not in the first, were ignored. Probabilities of eviction with respect to location of the binding motif and distance from nucleosome dyad were calculated by taking bins of distance from dyad, and using the following formula for each bin: (#initial - #2nd timepoint)/# initial.

Stochastic model for nucleosome accessibility

The Markov process $\big(X(t), N(t)\big)$ represent the number of dissembled DNA-histone contact sites by $X(t)$ and the SDTF binding by $N(t)$. The probabilities of transition for the process are

$$Prob(N(t+dt)=1|N(t)=0) = k_{on}(t)dt + o(dt),$$

$$Prob(N(t+dt)=1|N(t)=1) = k_{off}dt + o(dt),$$

$$Prob(X(t+dt)=n+1|X(t)=n) \ = a_n dt + o(dt), \text{ for } n \leq 13$$

$$Prob(X(t+dt)=n-1|X(t)=n, \ \ N(t)=0) = b_n dt + o(dt) \text{ for } n \geq 1, \text{ and}$$

$$Prob(X(t+dt) = n-1 | X(t) = n,\ N(t) = 1) = d_n dt + o(dt) \text{ for } n \geq 1.$$

Data availability

Macrophage paired-end ATAC-seq data generated for this paper are deposited to GEO (GSE156385, reviewer token: apqvkeywhpkvhkh). Previously published single-ended sequencing of the same samples can be found at GSE146068. Model and analysis code is available at Github (https://github.com/signalingsystemslab).

References

Adelaja, A., Taylor, B., Sheu, K.M., Liu, Y., Luecke, S., and Hoffmann, A. (2021). Six distinct NFκB signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. Immunity *54*, 916-930.e7.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Chromosomal DNA and Its Packaging in the Chromatin Fiber. Molecular Biology of the Cell. 4th Edition.

Allfrey, V.G., Littau, V.C., and Mirsky, A.E. (1963). ON THE ROLE OF HISTONES IN REGULATING RIBONUCLEIC ACID SYNTHESIS IN THE CELL NUCLEUS*. Proc Natl Acad Sci U S A *49*, 414–421.

Barken, D. (2005). Comment on "Oscillations in NF- B Signaling Control the Dynamics of Gene Expression." Science *308*, 52a–52a.

Bilokapic, S., Strauss, M., and Halic, M. (2018). Histone octamer rearranges to adapt to DNA unwrapping. Nat Struct Mol Biol *25*, 101–108.

Brahma, S., and Henikoff, S. (2020). Epigenome Regulation by Dynamic Nucleosome Unwrapping. Trends in Biochemical Sciences *45*, 13–26.

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. Current Protocols in Molecular Biology *109*.

Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat Genet 1–12.

Callegari, A., Sieben, C., Benke, A., Suter, D.M., Fierz, B., Mazza, D., and Manley, S. (2019). Single-molecule dynamics and genome-wide transcriptomics reveal that NF-kB (p65)-DNA binding times can be decoupled from transcriptional activation. PLOS Genetics *15*, e1007891.

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.

Chen, P., Dong, L., Hu, M., Wang, Y.-Z., Xiao, X., Zhao, Z., Yan, J., Wang, P.-Y., Reinberg, D., Li, M., et al. (2018). Functions of FACT in Breaking the Nucleosome and Maintaining Its Integrity at the Single-Nucleosome Level. Mol Cell *71*, 284-293.e4.

Cheng, Q.J., Ohta, S., Sheu, K.M., Spreafico, R., Adelaja, A., Taylor, B., and Hoffmann, A. (2021). NF-κB dynamics determine the stimulus specificity of epigenomic reprogramming in macrophages. Science *372*, 1349–1353.

Chou, T. (2007). Peeling and Sliding in Nucleosome Repositioning. Phys. Rev. Lett. *99*, 058105.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J. Mol. Biol. *319*, 1097–1113.

Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladde, M.P., and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. Mol Cell *38*, 590–602.

Dobrovolskaia, I.V., and Arya, G. (2012). Dynamics of Forced Nucleosome Unraveling and Role of Nonuniform Histone-DNA Interactions. Biophys J *103*, 989–998.

Dodd, I.B., Micheelsen, M.A., Sneppen, K., and Thon, G. (2007). Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. Cell *129*, 813–822.

Eslami-Mossallam, B., Schiessel, H., and van Noort, J. (2016). Nucleosome dynamics: Sequence matters. Advances in Colloid and Interface Science *232*, 101–113.

Fernandez Garcia, M., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H., and Zaret, K.S. (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. Molecular Cell *75*, 921-932.e6.

Gasparian, A.V., Burkhart, C.A., Purmal, A.A., Brodsky, L., Pal, M., Saranadasa, M., Bosykh, D.A., Commane, M., Guryanova, O.A., Pal, S., et al. (2011). Curaxins: Anticancer Compounds That Simultaneously Suppress NF- B and Activate p53 by Targeting FACT. Science Translational Medicine *3*, 95ra74-95ra74.

Hall, M.A., Shundrovsky, A., Bai, L., Fulbright, R.M., Lis, J.T., and Wang, M.D. (2009). High resolution dynamic mapping of histone-DNA interactions in a nucleosome. Nat Struct Mol Biol *16*, 124–129.

Hao, N., and O'Shea, E.K. (2012). Signal-dependent dynamics of transcription factor translocation controls gene expression. Nat Struct Mol Biol *19*, 31–39.

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

Henikoff, S. (2016). Mechanisms of Nucleosome Dynamics In Vivo. Cold Spring Harb Perspect Med *6*.

Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science *298*, 1241–1245.

Jin, F., Li, Y., Ren, B., and Natarajan, R. (2011). PU.1 and C/EBP(alpha) synergistically program distinct response to NF-kappaB activation through establishing monocyte specific enhancers. Proc Natl Acad Sci U S A *108*, 5290–5295.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Kim, Y.J., Sheu, K.M., Tsoi, J., Abril-Rodriguez, G., Medina, E., Grasso, C.S., Torrejon, D.Y., Champhekar, A.S., Litchfield, K., Swanton, C., et al. (2021). Melanoma dedifferentiation induced by IFN-gamma epigenetic remodeling in response to anti–PD-1 therapy. J Clin Invest *131*.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics *20*, 207–220.

Kobayashi, K., Hiramatsu, H., Nakamura, S., Kobayashi, K., Haraguchi, T., and Iba, H. (2017). Tumor suppression via inhibition of SWI/SNF complex-dependent NF-κB activation. Sci Rep *7*, 11772.

Kruithof, M., and van Noort, J. (2009). Hidden Markov Analysis of Nucleosome Unwrapping Under Force. Biophysical Journal *96*, 3708–3715.

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

Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet *36*, 900–905.

Lee, R.E., Walker, S.R., Savery, K., Frank, D.A., and Gaudet, S. (2014). Fold change of nuclear NF-kappaB determines TNF-induced transcription in single cells. Mol Cell *53*, 867–879.

Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005a). Rapid spontaneous accessibility of nucleosomal DNA. Nat Struct Mol Biol *12*, 46–53.

Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005b). Rapid spontaneous accessibility of nucleosomal DNA. Nature Structural & Molecular Biology *12*, 46–53.

Liu, Y., Zhou, K., Zhang, N., Wei, H., Tan, Y.Z., Zhang, Z., Carragher, B., Potter, C.S., D'Arcy, S., and Luger, K. (2020). FACT caught in the act of manipulating the nucleosome. Nature *577*, 426–431.

Lone, I.N., Shukla, M.S., Charles Richard, J.L., Peshev, Z.Y., Dimitrov, S., and Angelov, D. (2013). Binding of NF-κB to Nucleosomes: Effect of Translational Positioning, Nucleosome Remodeling and Linker Histone H1. PLoS Genet *9*, e1003830.

Longo, D.M., Selimkhanov, J., Kearns, J.D., Hasty, J., Hoffmann, A., and Tsimring, L.S. (2013). Dual Delayed Feedback Provides Sensitivity and Robustness to the NF-kB Signaling Module. PLOS Computational Biology *9*, 15.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997b). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.

Luger, K., Suto, R.K., Clarkson, M.J., and Tremethick, D.J. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat. Struct Biol. *7*, 1121–1124.

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

Miller, J.A., and Widom, J. (2003a). Collaborative competition mechanism for gene activation in vivo. Mol. Cell. Biol. *23*, 1623–1632.

Miller, J.A., and Widom, J. (2003b). Collaborative Competition Mechanism for Gene Activation In Vivo. Molecular and Cellular Biology *23*, 1623–1632.

Möbius, W., Neher, R.A., and Gerland, U. (2006a). Kinetic Accessibility of Buried DNA Sites in Nucleosomes. Phys. Rev. Lett. *97*, 208102.

Möbius, W., Neher, R.A., and Gerland, U. (2006b). Kinetic Accessibility of Buried DNA Sites in Nucleosomes. Physical Review Letters *97*.

Mobius, W., Osberg, B., Tsankov, A.M., Rando, O.J., and Gerland, U. (2013). Toward a unified physical model of nucleosome patterns flanking transcription start sites. Proceedings of the National Academy of Sciences *110*, 5719–5724.

Morgan, A.M., LeGresley, S.E., Briggs, K., Al-Ani, G., and Fischer, C.J. (2018). Effects of nucleosome stability on remodeler-catalyzed repositioning. Physical Review E *97*.

Moyle-Heyrman, G., Tims, H.S., and Widom, J. (2011). Structural Constraints in Collaborative Competition of Transcription Factors against the Nucleosome. Journal of Molecular Biology *412*, 634–646.

Mueller-Planitz, F., Klinker, H., and Becker, P.B. (2013). Nucleosome sliding mechanisms: new twists in a looped history. Nat. Struct. Mol. Biol. *20*, 1026–1032.

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

Ngo, T.T.M., Zhang, Q., Zhou, R., Yodh, J.G., and Ha, T. (2015). Asymmetric Unwrapping of Nucleosomes under Tension Directed by DNA Local Flexibility. Cell *160*, 1135–1144.

van Opheusden, B., Redig, F., and Schiessel, H. (2013). A Markov model for the dynamics of the nucleosome. J. Phys. A: Math. Theor. *46*, 095005.

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

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013b). Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell *152*, 157–171.

Polach, K.J., and Widom, J. (1995). Mechanism of Protein Access to Specific DNA Sequences in Chromatin: A Dynamic Equilibrium Model for Gene Regulation. Journal of Molecular Biology *254*, 130–149.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Research *25*, 1757–1770.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends in Genetics *25*, 335–343.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.-P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772–778.

Sen, S., Cheng, Z., Sheu, K.M., Chen, Y.H., and Hoffmann, A. (2020). Gene Regulatory Strategies that Decode the Duration of NFκB Dynamics Contribute to LPS- versus TNF-Specific Gene Expression. Cell Syst *10*, 169-182.e5.

Shih, V.F.-S., Davis-Turak, J., Macal, M., Huang, J.Q., Ponomarenko, J., Kearns, J.D., Yu, T., Fagerlund, R., Asagiri, M., Zuniga, E.I., et al. (2012). Control of RelB during dendritic cell activation integrates canonical and noncanonical NF-κB pathways. Nat Immunol *13*, 1162–1170.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol *6*, e65.

Stewart, A.J., Hannenhalli, S., and Plotkin, J.B. (2012). Why Transcription Factor Binding Sites Are Ten Nucleotides Long. Genetics *192*, 973–985.

Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat. Struct. Biol. *7*, 1121–1124.

Tims, H.S., Gurunathan, K., Levitus, M., and Widom, J. (2011a). Dynamics of Nucleosome Invasion by DNA Binding Proteins. Journal of Molecular Biology *411*, 430–448.

Tims, H.S., Gurunathan, K., Levitus, M., and Widom, J. (2011b). Dynamics of Nucleosome Invasion by DNA Binding Proteins. Journal of Molecular Biology *411*, 430–448.

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.

Winkler, D.D., and Luger, K. (2011). The Histone Chaperone FACT: Structural Insights and Mechanisms for Nucleosome Reorganization. J. Biol. Chem. *286*, 18369–18374.

Zhou, K., Gaullier, G., and Luger, K. (2019). Nucleosome structure and dynamics are coming of age. Nature Structural & Molecular Biology *26*, 3–13.

CHAPTER 4

Discussion

By examining two properties of macrophage function – response-specificity and stimulus-specific memory – we have begun to delineate quantitative measures for macrophage health, which may ultimately contribute to further metrics that depict the health of the innate immune system. As the initial responders that then communicate with all downstream cells of the activated during inflammation, immune sentinel health itself may be a vanguard for the health of the entire immune system. What the recent COVID19 pandemic has made abundantly clear is that the immune system is hugely variable across people, and often, no outward sign of immune system dysfunction is evident until it is challenged by a pathogen or any inflammatory threat. In the pandemic, several strong trends stood out – those with inflammatory conditions, such as obesity or age-associated inflammaging, were statistically much more susceptible to severe disease. However, beyond statistics, more individual measures of risk for inflammatory disease would be beneficial.

Clinically, a few tests for assessing the health of particular organ systems already follow the approach of testing its function through a perturbation or a challenge. Cardiac stress tests are one example, where the steady state may show normal sinus rhythm, but only upon exertion can problems of oxygenation or arrhythmia be diagnosed. Another functional test is the glucose tolerance test, where the body is challenged with a bolus of sugar to see how well the pancreas functions to produce insulin that should move the sugar from the blood to the tissues – a common test for diabetes. Similarly, the immune system health can be measured through challenge to its function, by perturbing immune cells isolated from blood and profiling the resulting transcriptome or epigenome. Establishing the healthy ranges for response-specificity and characterizing the mechanisms that generate stimulus-specific epigenetic memory, may move us towards this goal. However, the diagnosis or prognosis for a wide variety of diseases depends on immune system function, including autoimmune diseases, cancers, or neurodegenerative diseases, to name a few. It remains to be seen how well the functional

metrics that capture immune sentinel health are reflective of risk or disease stage of each of these.

Mechanistically, a key question still open is how and how much the misregulation of the two functional properties described here contributes to disease, which may further illustrate a distinct importance for each one. For example, the first hallmark of response-specificity is a measure of the distribution overlap in sentinel responses to different ligands; single cell response distributions with abnormal spread may result in improper immune sequelae. One disease category that involves the loss of healthy response-specificity are autoimmune diseases, where symptoms are often sporadic with unknown triggers. A second disease category are diseases of conditioning. As opposed to the genetic changes that are often risk factors for autoimmune disease, diseases that depend on environmental exposures or microenvironment context, such as aging, obesity, or cancer, reflect poor regulation in context-dependent response-specificity. Furthermore, in contrast to response-specificity that measures immediate responses, diseases involving misregulation of stimulus-specific memory may also result when the primary stimulus improperly alters the epigenome, resulting in hyperinflammatory disorders upon a second encounter. These changes may also be positive, as seen with BCG vaccination that appears to have benefits against other infectious agents. It will thus of great interest to investigate how disease may result from the dysregulation of the functional hallmarks of sentinel health, and a further understanding of the mechanisms from which each hallmark arises may suggest appropriate therapeutic strategies for intervention.