

---

# Analysis of the epigenetic landscape in murine macrophages

Verena Maria Link

---



München 2018



---

# **Analysis of the epigenetic landscape in murine macrophages**

**Verena Maria Link**

---

Dissertation  
an der Fakultät für Biologie  
der Ludwig-Maximilians-Universität  
München

vorgelegt von  
Verena Maria Link  
aus München

München, den 15.03.2018

Erstgutachter: Prof. Dr. Dirk Metzler

Zweitgutachter: Prof. Dr. Wolfgang Enard

Tag der Abgabe: 15.03.2018

Tag der mündlichen Prüfung: 20.06.2018

**ERKLÄRUNG**

Diese Dissertation wurde im Sinne von §12 der Promotionsordnung von Prof. Dr. Dirk Metzler betreut. Ich erkläre hiermit, dass die Dissertation nicht einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung unterzogen habe.

**EIDESSTATTLICHE VERSICHERUNG**

Ich versichere ferner hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist.

---

Greifenberg, 15.03.2018

Ort, Datum

---

Verena Link



## DECLARATION OF CO–AUTHOR CONTRIBUTIONS

The study in Chapter 1 (Gosselin, D., Link, V. M., Romanoski, C. E. et al. (2014), appeared in *Cell*) was designed by David Gosselin and Christopher Glass. In this project I am an equally contributing co–first author. I analyzed and interpreted the ChIP–seq and RNA–seq data in close collaboration with Casey Romanoski and with additional input from David Gosselin and Christopher Glass. I wrote the manuscript together with David Gosselin, Casey Romanoski and Christopher Glass.

The study in Chapter 2 (Eichenfield, D. Z., Troutman, D. T., Link, V. M. et al. (2016), appeared in *eLife*) was designed by Dawn Eichenfield, Ty Troutman, Christopher Glass and myself with additional input from Hanna Lesch and Richard Gallo. I am an equally contributing co–first author in this project. I analyzed and interpreted the data with help from Dawn Eichenfield, Ty Troutman, and Christopher Glass, with additional input from Michael Lam, Nathanael Spann, Jenhan Tao, Jun Muto, Richard Gallo and Ronald Evans. I performed experiments together with Dawn Eichenfield and Ty Troutman with additional help from Michael Lam, Han Cho, David Gosselin, Nathanael Spann, and Hanna Lesch. I wrote the manuscript together with Dawn Eichenfield, Ty Troutman and Christopher Glass.

The method in Chapter 3 (Link et al. (2018b) appeared in *bioRxiv*) was developed by myself with advice from Christopher Glass and Dirk Metzler. I extended the algorithm previously implemented by Casey Romanoski for pairwise comparisons. I also generalized this approach by developing a linear mixed model. Furthermore, I implemented the software including the aforementioned algorithms, as well as all the data handling in perl and R. I also wrote the manuscript with input from Christopher Glass, Dirk Metzler and Casey Romanoski.

The study in Chapter 4 (Link et al. (2018a) under review in *Cell*) was designed by Christopher Glass and myself. Dirk Metzler advised me throughout the project. I performed experiments with help from Sascha Duttke, Billy Chun, Emma Westin, Marten Hoeksema, Yohei Abe, Gregory Fonseca, Ty Troutman, Nathanael Spann, Tobias Strid, Mashito Sakai, Miao Yu, and Rong Hu. I performed the analysis with help from Inge Holtman and Dylan Skola, as well as Casey Romanoski, Jenhan Tao and Rongxin Fang. I interpreted the results and wrote the manuscript with Christopher Glass as well as feedback from Dirk Metzler and Bing Ren.

In Chapter 5 I listed further contributions during my Ph.D. studies. In the study Pirzgalska et al. (2017) (appeared in *Nature Medicine*), I developed and implemented a new method to normalize low–input RNA–seq experiments and helped the lead author analyzing the RNA–seq data after normalization. In the other study I contributed to (Oishi et al. (2017) appeared in *Cell Metabolism*) I helped with the analysis of the ChIP–seq and RNA–seq data.

The co–first authors from Gosselin, D., Link, V. M., Romanoski, C. E. et al. (2014) agree that I correctly stated my contributions to this project:

---

David Gosselin

---

Verena M. Link

---

Casey E. Romanoski

The co–first authors from Eichenfield, D. Z., Troutman, D. T., Link, V. M. et al. (2016) agree that I correctly stated my contributions to this project:

---

Dawn Z. Eichenfield

---

Ty D. Troutman

---

Verena M. Link

Furthermore, Prof. Dr. Dirk Metzler, as well as my supervisor at the University of California, San Diego, Prof. Dr. Christopher K. Glass confirm that I stated all contributions correctly.

---

Prof. Dr. Dirk Metzler

---

Prof. Dr. Christopher K. Glass

# Contents

|   |     |
|---|-----|
| Acknowledgements  | 7   |
| Summary   | 10  |
| Introduction  | 1   |
| Chapter 1   |     |
| Environment Drives Selection and Function of Enhancers Controlling Tissue-Specific Macrophage Identities                              | 13  |
| Chapter 2   |     |
| Tissue damage drives co-localization of NF- $\kappa$ B, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages | 29  |
| Chapter 3   |     |
| MARGE: Mutation Analysis of Regulatory Genomic Elements   | 61  |
| Chapter 4   |     |
| Transcription Factor Landscapes in Macrophages from Genetically Diverse Mice Reveal Connected Regulatory Domains                      | 101 |
| Chapter 5   |     |
| Additional contributions  | 173 |
| Discussion  | 207 |
| Bibliography  | 220 |
| Glossary  | 243 |



## List of Publications

- 2018**      **MARGE: Mutation Analysis for Regulatory Genomic Elements**  
VM Link, CE Romanoski, D Metzler, CK Glass  
*bioRxiv*, 268839
- Transcription Factor Landscapes in Macrophages from Genetically Diverse Mice Reveal Connected Regulatory Domains**  
VM Link, SH Duttke, HB Chun, IR Holtman, E Westin, MA Hoeksema, Y Abe, D Skola, CE Romanoski, J Tao, GJ Fonseca, TD Troutman, NJ Spann, T Strid, M Sakai, M Yu, R Hu, R Fang, D Metzler, B Ren, CK Glass  
*Cell* - under review
- 2017**      **Sympathetic neuron-associated macrophages contribute to obesity by importing and metabolizing norepinephrine**  
RM Pirzgalska, E Seixas, J Seidman, VM Link, NM Sánchez, R Mendes, V Gres, N Kubasova, I Morris, BA Arús, CM Larabee, M Vasques, F Tortosa, AL Sousa, S Anandan, E Tranfield, MK Hahn, M Iannacone, NJ Spann, CK Glass, AI Domingos  
*Nature medicine*
- SREBP1 contributes to resolution of pro-inflammatory TLR4 signaling by reprogramming fatty acid metabolism**  
Y Oishi, NJ Spann, VM Link, ED Muse, T Strid, C Edillor, MJ Kolar, T Matsuzaka, S Hayakawa, J Tao, MU Kaikkonen, AF Carlin, MT Lam, I Manabe, H Shimano, A Saghatelian, CK Glass  
*Cell metabolism*
- 2016**      **Tissue damage drives co-localization of NF- $\kappa$ B, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages**  
DZ Eichenfield\*, TD Troutman\*, VM Link\*, MT Lam, H Cho, D Gosselin, NJ Spann, HP Lesch, J Tao, J Muto, RL Gallo, RM Evans, CK Glass  
*Elife*  
\* These authors contributed equally to this work.

- 2015**      **Exploiting genomics and natural genetic variation to decode macrophage enhancers**  
CE Romanoski, **VM Link**, S Heinz, CK Glass  
*Trends in immunology*
- Mechanisms underlying the selection and function of macrophage-specific enhancers**  
**VM Link**, D Gosselin, CK Glass  
*Cold Spring Harbor symposia on quantitative biology*
- 2014**      **Environment drives selection and function of enhancers controlling tissue-specific macrophage identities**  
D Gosselin\*, **VM Link\***, CE Romanoski\*, GJ Fonseca, DZ Eichenfield, NJ Spann, JD Stender, HB Chun, H Garner, F Geissmann, CK Glass  
*Cell*  
\* These authors contributed equally to this work.

# Acknowledgements

Five years ago, when I started this journey, which eventually led to the existence of this dissertation, I did not know what it means to be a scientist or how to spearhead a scientific project. Here, I would like to thank the many people who trained me, listened, gave advice, and lifted me up.

First of all, I would like to thank my Ph.D. supervisor Dirk Metzler, who gave me the opportunity to be a Ph.D. student. He gave me the freedom to go wherever my research took me and was supportive in every step of the way. My deep thank also goes to Christopher Glass, who took me into his laboratory and allowed me to grow and thrive as a scientist, as well as a person. His guidance made this dissertation possible and his deep devotion to science made me want to pursue an academic career.

I thank all colleagues from the statistical group and department of evolutionary biology for supporting me. I want to thank every member of the Glass laboratory for the past five years, for teaching me, for working together to a common goal, but most of all for creating such a fun and productive environment, where going to work felt like a privilege every day. I would like to express my sincere thank to Nathanael Spann and Sven Heinz for teaching me everything I know about molecular biology and how to analyze the data. I would like to thank Chris Benner for being a great computational mentor and role model. I also would like to express my thanks to Tobias Strid, who spent long nights with me in the laboratory to teach me molecular biology and was a great friend. Special thanks goes to Ty Troutman, Casey Romanoski and David Gosselin, who were great scientific mentors and even better friends. My deep thank also goes to Jana Collier who can not be recognized enough for her ability to keep the laboratory running, as well as Leslie van Ael, who has magical skills in preparing figures and manuscripts. I want to thank Jason Seidman, Chantle Edillor, Emma Westin, Inge Holtman, Marten Hoeksema, Claudia Han, Alexi Nott and Johannes Schlaechtzi for being great friends all the years and making early mornings and late nights fun.

I want to thank all my co-first authors who worked through long nights to finish the projects, mainly David Gosselin, Casey Romanoski, Dawn Eichenfield and Ty Troutman. Without these people I would not be the scientist I am today.

I can not thank enough all my co authors who helped me on my main project – the so called strains project. Thank you for showing up every week for our project meeting, for working tirelessly on getting the experiments done and for spending hours thinking about good ways to analyze and present data. Without your support this project would not have

happened.

Lastly, I want to thank my family. My parents Max and Hedi Link, who supported me from the beginning and made me believe I can become whatever I choose to be. My deep appreciation goes to my sister Sabrina Link for being there whenever I needed a friend. I want to thank Carina Demel for being the best friend possible. Finally I want to express my deepest gratitude to Cyrus Afrasiabi for his support, encouragement, his believe in me and his quiet but steady love that carried me over all bumps on the way.

## Summary

Macrophages are cells of the innate immune system and play essential roles in the regulation of inflammatory responses in all parts of the body. Furthermore, macrophages are also involved in different tissue-specific functions and maintenance of the tissue homeostasis. These functions are controlled by the epigenetic landscape, consisting of promoters and enhancers that together regulate gene expression. Enhancers are stretches of regulatory genomic sequences in the non-coding regions of the genome that can be bound by lineage-determining transcription factors. These enhancers can loop in three-dimensional space to be in close proximity to promoters and contribute to the regulation of gene expression. Previous studies suggest that there are about 1 million enhancers in the mammalian genome, of which only about 30,000 – 40,000 are selected in each specific cell type. This dissertation studies the regulation of the epigenetic landscape of murine macrophages by utilizing different tissue macrophages, different complex and simple stimuli, as well as natural genetic variation as a mutagenesis screen.

The overarching research question of this dissertation is to understand how the enhancer landscape in macrophages gets selected and regulated in order to control gene expression. In more detail, the main questions answered in this dissertation are: What are the epigenetic mechanisms that are responsible for tissue-specific functions? How do complex stimuli change the epigenetic landscape of macrophages in comparison to simple stimuli? How does natural genetic variation influence the epigenetic landscape and gene expression in murine macrophages?

In Chapter 1 (Gosselin, D., Link, V. M., Romanoski, C. E. et al. (2014) appeared in *Cell*) we investigate the influence of the tissue environment on the epigenetic landscape in mouse macrophages. We compare macrophages residing in the brain (microglia) with macrophages from the peritoneal cavity by measuring mRNA expression, as well as enhancer activation (H3K4me2, H3K27ac, and PU.1). We find highly expressed genes unique to one population of macrophages, which correlates well with the activity signature at enhancers in the corresponding cells. By analyzing the enhancer landscape, we find that the macrophage lineage-determining transcription factor PU.1 plays a key role in establishing the enhancer repertoire, creating a common, macrophage-specific enhancer landscape. Furthermore, expression of tissue-specific transcription factors in collaboration with PU.1 drives a subset of tissue-specific enhancers regulating the differences in gene expression between different tissue-specific macrophage populations.

In Chapter 2 (Eichenfield, D. Z., Troutman, D. T., Link, V. M. et al. (2016) appeared in *eLife*) we investigate the effect of complex stimuli onto the epigenetic landscape in macrophages on the example of wounds. Stimulation of macrophages with homogenated tissue to mimic a wound environment shows a unique pattern of gene expression, which is different from gene expression patterns found after single stimuli (e.g. LPS, IL-4 etc.). To gain insight into the regulation of the enhancer landscape after complex stimuli, we compare the epigenome after single stimuli and tissue homogenate and find substantial differences in enhancer selection and activation. We find that the complex damage signal promotes co-localization of several signal-dependent transcription factors to enhancers not

observed under the single stimuli. Therefore, more complex polarizations of cells lead to new combinations of signal-dependent transcription factors and an epigenetic landscape different than observed with single stimuli.

In Chapter 3 (Link et al. (2018b) appeared in *bioRxiv*) MARGE (Mutation Analysis for Regulatory Genomic Elements) is presented, a new method to analyze the effect of natural genetic variation on transcription factor binding and open chromatin. MARGE provides a suite of software tools that integrates genome-wide genetic variation data (including insertions and deletions) with epigenetic data. It provides software to create custom genomes based on a reference genome and variation data, to shift coordinates between different custom genomes, as well as do downstream ChIP-seq analysis. The main algorithm in MARGE analyzes if mutations in transcription factor binding motifs are significantly affecting transcription factor binding or open chromatin. MARGE provides a pairwise comparison, in which the significance of each motif is calculated with a student's t-test. It compares the transcription factor binding distribution of each mutated motif in individual one with the distribution in individual two. For a more general approach that allows comparisons of many individuals MARGE implements a linear mixed model, modeling transcription factor binding with fixed effects motif existence and random effects locus and genotype. The development of this software allows in depth analysis of genetic variation data in combination with epigenetic data.

In Chapter 4 (Link et al. (2018a) under review in *Cell*) we analyze the effect of natural genetic variation in five diverse strains of mice on the epigenetic landscape. We choose three well-known laboratory inbred mouse strains, as well as two very diverse wild-derived inbred mouse strains. We investigate the enhancer landscape, open chromatin and binding of the most important macrophage lineage-determining transcription factors. We observe substantial strain-specific differences in gene expression of which the majority can be explained by *cis*-regulatory elements. Application of MARGE onto the transcription factor binding data reveals roles of about 100 transcription factors in establishing the enhancer repertoire in macrophages. Unexpectedly, we find that a substantial fraction of strain-specific DNA binding of transcription factors cannot be explained by local mutations. Investigation of this phenomenon in more detail shows highly interconnected clusters of transcription factors that reside within topologically associating domains. These interconnected clusters are highly correlated with activation of enhancers and gene expression of the nearest gene, uncovering a new layer of transcriptional regulation.

In Chapter 5, I briefly discuss additional contributions to the field of macrophage biology I made during my Ph.D. Namely, I was involved in two additional projects. In the first project (Pirzgalska et al. (2017) appeared in *Nature Medicine*) we identify sympathetic neuron-associated macrophages (SAM) that import and degrade norepinephrine via expression of solute carrier family 6 member 2 (Slc6a2) and monoamine oxidase A (MAOa). We demonstrate that SAM-mediated clearance of extracellular norepinephrine contributes to obesity and we show the relevance of this finding in humans, as we found that SAMs are also present in human tissues. The second project (Oishi et al. (2017) appeared in *Cell Metabolism*) studies the role of nuclear receptors (LXR and SREBP) in induction of anti-inflammatory fatty acids. We find that right after stimulation of TLR4 (during the

induction phase) NF- $\kappa$ B dependent genes are upregulated, whereas LXR dependent genes are repressed. This leads to activation of SREBP1, which drives the expression of enzymes involved in mono-unsaturated and omega-3 polyunsaturated fatty acid biosynthesis. The fatty acids produced by these enzymes repress inflammatory genes under the control of NF- $\kappa$ B and the inflammatory signal gets resolved.

In summary, my studies used a combination of experimental and computational approaches to investigate the effect of tissue-environment and factors, complex stimuli and natural genetic variation on the epigenetic landscape in macrophages. These studies broadened our understanding of the regulation of gene expression by the epigenetic landscape substantially. We showed that there is a core set of lineage-determining transcription factors in macrophages, which require diverse signal-dependent transcription factors to establish the enhancer landscape. Not only did we show that transcription factors regulated by the local environment play essential roles in establishing and maintaining tissue-specific functions of macrophages, but also that more complex stimuli can re-direct and combine signal-dependent transcription factors to establish new enhancers, not observed under the single stimuli. Using natural genetic variation as a mutagenesis screen allowed us to estimate the involvement of about 100 transcription factors in shaping the enhancer landscape, as well as to uncover a new layer of transcription regulation due to highly interconnected clusters of concordantly bound transcription factors.



# General Introduction

## Macrophages

Macrophages are important cells in the innate immune system where they play an important role in responding to infection, inflammation and tissue injury (Geissmann et al. (2010), Wynn et al. (2013) Lucas et al. (2010)). Macrophages can be found in every tissue of the body where they perform their general immune functions, as well as contribute to maintaining homeostasis of the host tissue (Wynn et al. (2013), Gordon et al. (2014)). One of the first described *in vivo* functions of macrophages was surfactant recycling by lung alveolar macrophages (Wright (1990)). Other known examples of tissue-specific functions of macrophages include bone resorption and remodeling by osteoclasts (Teitelbaum (2000)) and control of insulin sensitivity and adaptive thermogenesis in adipose tissue (Odegaard et al. (2007), Qiu et al. (2014)). Furthermore, it is known that microglia (macrophages in the brain) are involved in neuronal synaptic pruning (Paolicelli et al. (2011)). These functions are essential for the development and maintenance of healthy tissue. However, dysregulation of macrophages can lead to chronic inflammatory diseases. A prominent example is the important role of macrophages in the development of atherosclerosis (Moore and Tabas (2011), Pollard (2004)). Dysregulation of macrophages in cancer can lead to tumor growth and metastasis (Noy and Pollard (2014)). This makes macrophages an interesting cell type to target for therapeutic purposes, which requires a detailed understanding of the mechanisms by which macrophage phenotypes are controlled. This, however, remains a largely unmet goal so far.

Macrophages have been studied for a long time with special focus on their response to various ligands *in vitro* (Takeuchi and Akira (2010)). This led to the characterization of two different polarization programs. Macrophages can be classically activated with a proinflammatory phenotype (M1 macrophages) or alternately activated and show anti-inflammatory traits (M2 macrophages) (Gorden (2003), Mosser and Edwards (2008)). M1 polarization can be mainly achieved by stimulation of macrophages with lipopolysaccharide (LPS), a component of Gram-negative bacteria. Stimulation with LPS leads to activation of toll-like receptor (TLR)4, which induces members of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF- $\kappa$ B) complex and activator protein 1 (AP-1) (Figure 1) (Medzhitov and Horng (2009), Smale (2012)). M1 phenotypes can also be observed after stimulation of cytokine receptors through tumor necrosis factor (TNF) or interleukin (IL)-1/IL-6, as well as activation of signal transducer and activator of transcription (STAT)1 after exposure to

interferon gamma ( $\text{IFN}\gamma$ ) and activation of its receptor (Liu et al. (2014)). Activation of any of these factors induces hundreds of pro-inflammatory genes in macrophages leading to the initiation of an adaptive macrophage immune response. In order to activate the alternate M2 phenotype, macrophages can be treated with IL-4, which in turn induces STAT6 and interferon regulatory factor (IRF)4. These transcription factors activate a gene expression profile against parasitic infection (Gorden (2003), Takeda and Akira (2000)) (Figure 1). Furthermore, free fatty acids can induce peroxisome proliferator-activated receptor gamma ( $\text{PPAR}\gamma$ ), which also drives M2 polarization.

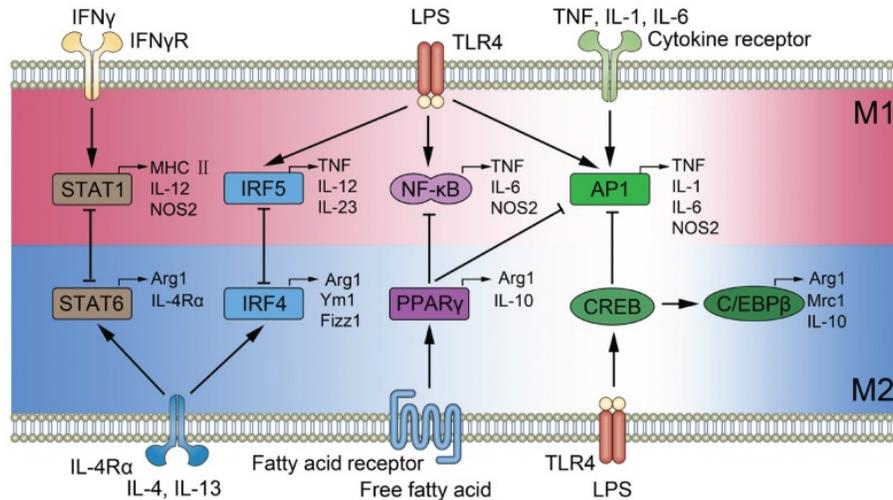


Figure 1: Pathways that lead to classically activated/proinflammatory M1 macrophages (upper part – red), as well as to the alternately activated/anti-inflammatory M2 macrophage phenotype (bottom part – blue). Figure is taken from Liu et al. (2014).

## Enhancers

Much effort has been put into characterizing enhancer-like regions in different cell types and organisms. There are about one million enhancers, of which about 30,000 to 40,000 are activated in any specific cell type. Enhancers were initially identified as discrete regions of deoxyribonucleic acid (DNA) that increase transcriptional activity of promoters from a distance (Banerji et al. (1981)). DNA is not openly accessible *in vivo* but assembled in chromatin. DNA is wrapped around proteins, called histones. These histones build octamers (consisting of the subunits H2A, H2B, H3, and H4) and can be modified post-translationally. The assembly of the histone subunits including DNA is called nucleosome, which can be further packed into chromatin. Enhancer elements are established by the binding of sequence-specific transcription factors (TFs), which compete with

nucleosomes and push them to the side to generate nucleosome-free regions. Several experimental methods have been developed to measure open chromatin (see experimental methods). Most TFs expressed in each cell type are not able to recognize their DNA-binding motifs in closed chromatin. However, so-called pioneer factors (e.g. PU.1 in macrophages) possess this unique ability and are therefore able to establish enhancers. Systematic analysis of chromatin led to the recognition that enhancers are marked by high abundance of mono- and dimethylation at histone H3 lysine 4 and concomitantly low levels of trimethylation (i.e., H3K4me1<sup>high</sup>/H3K4me2<sup>high</sup>/H3K4me3<sup>low</sup>) (Heintzman et al. (2007)). Promoters, on the other hand, display an opposite molecular phenotype (i.e., H3K4me1<sup>low</sup>/H3K4me2<sup>high</sup>/H3K4me3<sup>high</sup>).

Enhancers can be marked by different histone modifications, which are used to categorize them as primed, poised or active (Heintzman et al. (2007), Barski et al. (2007), Ernst et al. (2011)). Primed enhancers are usually marked by H3K4me1 or H3K4me2 (He et al. (2010)) without any additional active marks. Poised enhancers can be marked with H3K4me1 or H3K4me2, but they are additionally marked with the repressive mark H3K27me3 (trimethylation of histone H3 on lysine 27), which is mutually exclusive with acetylation on the same residue (Rada-Iglesias et al. (2011)). On the other hand, active enhancers are marked with acetylation of histone H3 lysine 27 (H3K27ac) additionally to H3K4me1 and H3K4me2 (Rada-Iglesias et al. (2011), Creighton et al. (2010)). Furthermore, studies showed that active enhancers are often actively transcribed by RNA polymerase II (Pol II), giving rise to enhancer RNAs (eRNAs) (Ernst et al. (2011), Hah et al. (2011), Kaikkonen et al. (2013), Step et al. (2014)). All or a combination of these features can be used to identify enhancers in different cells and tissues by chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq) and other sequencing-based methods. Using these approaches, the mouse and human genomes have been estimated to contain several hundred thousand enhancers, the great majority of which are present in cell-restricted patterns (ENCODE Project Consortium (2012)). Some studies have demonstrated that chromatin looping is facilitated by eRNAs (Schaukowitch et al. (2014), Hsieh et al. (2014), Li et al. (2013b)). Consistent with this, changes in eRNA levels correlate with changes in target gene expression (Li et al. (2013b), Kaikkonen et al. (2013), Bonn et al. (2012), Kieffer-Kwon et al. (2013), Wang et al. (2011)).

The overall function of enhancer transcription, as well as the regulatory importance of eRNAs remain mostly elusive. Some studies in macrophages suggest potential roles in at least some regional control of gene expression, however, the overall genome-wide significance is still under debate. First, enhancer transcription at newly selected enhancers was linked to the deposition of H3K4me1/H3K4me2 at these locations (Kaikkonen et al. (2013), Ostuni et al. (2013)). Therefore, enhancer transcription might be potentially important in initiating and/or maintaining the histone signature characteristic of enhancers. Second, at least some eRNAs appear to contribute to enhancer function. For example REV-ERB nuclear receptors actively repress gene transcription in mouse macrophages by inhibiting eRNA transcription of target enhancers (Lam et al. (2013)). In particular, binding of REV-ERB at enhancers regulating matrix metalloproteinase 9 (Mmp9) and CX3C chemokine receptor 1 (Cx3cr1) gene expression represses enhancer activity and eRNA transcription, causing

low levels of *Mmp9* and *Cx3cr1* messenger RNAs (mRNAs). This likely occurs through the recruitment of the NCoR–HDAC3 repressor complex (Zamir et al. (1996), Yin, L. and Lazar, M. A. (2005)). In the absence of REV–ERB factors, these enhancers are de-repressed and transcribe high levels of eRNAs, which translates into aberrant increase of *Mmp9* and *Cx3cr1* gene expression.

More recently, the concept of stretch or super enhancers (SEs) was introduced (Hnisz et al. (2013), Whyte et al. (2013)). SEs are defined as large genomic regions, (about one order of magnitude larger than traditional enhancers) with an unusually strong enrichment of active histone marks (e.g. H3K27ac), as well as binding of transcriptional coactivators. Most cells show between 300 and 800 SEs, many of which are associated with genes encoding cell type-specific TFs and other genes important for the identity and function of that particular cell (Hnisz et al. (2013), Adam et al. (2015)).

Several studies estimate that there are between 35,000 and 45,000 primed and active enhancers in mouse macrophages (Ghisletti et al. (2010), Heinz et al. (2010)). Therefore, enhancers substantially outnumber the 12,000 active promoters in these cells. De-novo motif analysis showed that macrophage enhancers are enriched for motifs that are recognized by TFs relevant to macrophage ontogeny and functions. For example, one of the most enriched binding motifs is an ETS motif, which is recognized by PU.1. Furthermore, experimentally obtained PU.1 binding sites show a strong enrichment for H3K4me1<sup>high</sup>/H3K4me3<sup>low</sup>, a combination of marks commonly found on primed and active enhancers. Motifs associated with CCAAT-enhancer-binding protein (C/EBP), AP-1, IRF, NF- $\kappa$ B, and liver X receptor (LXR) TFs are also prevalent within enhancers in macrophages (Barish et al. (2010), Ghisletti et al. (2010), Heinz et al. (2010)).

## A collaborative/hierarchical model for enhancer selection and activation

To select enhancers in closed chromatin the binding of so-called pioneer factors is required. A subset of these factors functions as lineage-determining transcription factors (LDTFs) (McPherson et al. (1993), Bossard and Zaret (1998), Lee et al. (2005), Heinz et al. (2010)). Various examples of LDTFs include PU.1 in macrophages (Heinz et al. (2010)), forkhead box protein A1 (FOXA1) in breast cancer cells (Lupien et al. (2008)), octamer transcription factor (OCT)-4 and sex determining region Y-box (SOX)2 in cell reprogramming (Soufi et al. (2012)), and paired box protein (PAX)7 in reprogramming a corticotrope cell line into a melanotrope-like cell line (Budry et al. (2012)). A defining feature of pioneering TFs is their ability to recognize and bind their DNA recognition motifs in closed chromatin, thus effectively competing with nucleosomes to create nucleosome-free regions. TF binding is then followed by modifications of the histone tails of the enhancer-associated nucleosomes located in the vicinity of the nucleosome-free regions.

All TFs recognize short DNA sequences (6-12 bp) and their motifs show varying levels of degeneracy (D'haeseleer (2006)). Binding of TF can be measured by ChIP-seq experi-

ments (Johnson et al. (2007)). Analyses of a wide variety of different TFs have shown that only a small subset of all possible binding sites is occupied by each TF (Carr and Biggin (1999), Iyer et al. (2001), Yang et al. (2006)) and that the binding sites for the same factor can vary between cell types. For example, given the size of the mouse genome, there are potentially between 650,000 and 1.4 million sites where PU.1 can bind DNA (Heinz et al. (2013), Barozzi et al. (2014)), yet only about up to 45,000 of those are selected in differentiated macrophages (Heinz et al. (2010)).

There are likely a multitude of factors contributing to these restricted, yet functional, binding events of LDTFs. Among these factors, the collaborative activity between LDTFs and other TFs appear to be fundamental to a significant proportion of actual LDTF binding (collaborative enhancer selection) (Figure 2A). For example, the TF PU.1 is an important LDTF in macrophages and B cells. ChIP-seq experiments showed that macrophage-specific enhancers are enriched for binding motifs of PU.1 and macrophage-specific collaborative factors C/EBP and AP-1, whereas in B cells enhancers are enriched for PU.1 and B cell-specific factors E2A, early B-cell factor 1 (EBF1), and OCT-2 (Ghisletti et al. (2010), Heinz et al. (2010)).

Further support of the model was gathered by analyzing TF binding in different strains of mice. The natural genetic variation between mouse strains can be used to study selection of actual LDTF binding in macrophages (Heinz et al. (2013)). These studies showed that mutations in a C/EBP consensus motif not only eliminate C/EBP binding, but also abrogate the neighboring PU.1 binding. Importantly, mutations also affect associated enhancer and decrease abundance of H3K4me2 and H3K27ac marks at mutated enhancers. Using human white adipose tissue cells from several different individual, it was possible to show that mutations in C/EBP motifs disrupt nearby PPAR $\gamma$  binding (Soccio et al. (2015)). Many cells possess the ability to quickly adapt their gene expression program to changes in the environment. Further, the same signal can lead to the induction of the same signal-dependent transcription factor (SDTF) but different transcriptional outcomes and enhancer landscapes in different cell types. Studies showed that in more than 60%, SDTFs bind to enhancers already established by LDTFs (hierarchical enhancer selection), explaining how the same broadly expressed SDTF can exert cell-specific functions and responses (Figure 2B). Investigation of the role of natural genetic variation between inbred mouse strains on binding of the NF-kB subunit p65, showed that 34% of mouse strain-specific p65 binding was due to mutations in LDTF motifs, whereas only 9% was due to mutations in the p65 binding motif itself (Heinz et al. (2013)).

The exact biochemical mechanisms involved in the collaborative binding of LDTFs is not well understood. The motifs of the collaborative factors are in close proximity to each other in sequence (<100 bp), but most of these are not found at a distance of < 20 bp (Kazemian et al. (2013)). This implies a collaborative model that is not strictly dependent on protein-protein interactions and allows for a limited flexibility in spacing requirements. Consistent with this flexibility, the precise genomic location of cell-specific enhancers relative to target genes are largely not conserved between mice and humans (Cheng, Y., Ma, Z. et al. (2014)), however, the cell type-specific combinations of LDTF motifs appear to be conserved between species (Stergachis et al. (2014)). This suggests that while the spacing

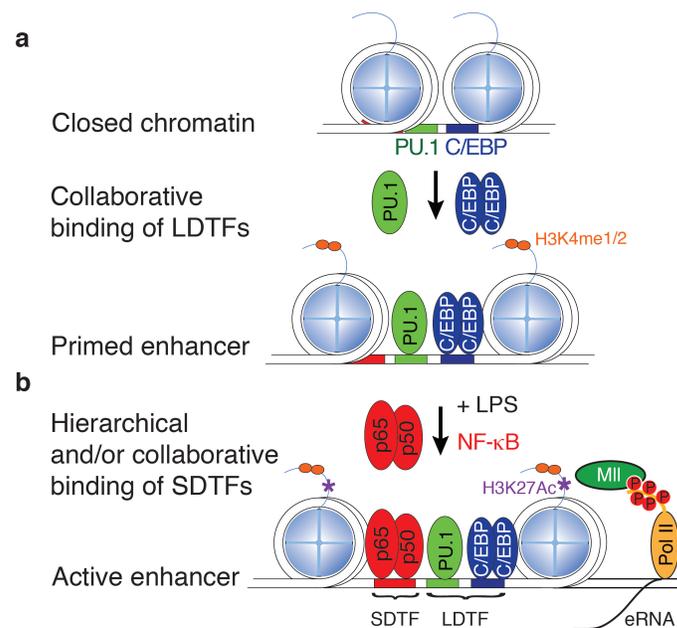


Figure 2: Collaborative/hierarchical model for enhancer selection and activation: **A** Macrophage lineage-determining transcription factors (LDTFs) PU.1 and C/EBP bind collaboratively to regions of closed chromatin where both of their binding motifs are present. Upon binding they open the chromatin and mono- and di-methyl groups are added to histone 3 lysine 4 (H3K4me1/H3K4me2) creating a primed enhancer (collaborative enhancer selection). **B** After macrophages are stimulated with LSP, the signal-dependent transcription factor NF- $\kappa$ B gets transported into the nucleus and can bind to the already established enhancers, thus activating them. This leads to acetylation of histone 3 lysine 27 (H3K27ac) and the production of enhancer RNA (eRNA) (hierarchical enhancer selection). Figure adapted from Link et al. (2015).

has changed since mice and humans diverged, the meaningful combinations of TFs that drive specific functions have largely remained the same. This is confirmed by experiments with transgenes between species. For example, when the human globin locus is inserted in mice, it is expressed with the same fetal-to-adult switch as it is in humans (McConnell et al. (2011), Peterson et al. (1993)).

The collaborative/hierarchical model addresses three key issues with regard to enhancer selection. First, it explains how relatively low numbers of binding sites are occupied by a factor in comparison to all potential binding sites. In addition, it provides a molecular and epigenetic mechanism whereby the same LDTF can set up very different enhancer repertoires in different cell types, as is the case of PU.1 in macrophages and B cells. Finally it explains how broadly expressed SDTF can result in different cell-specific transcriptional outcomes.

## Experimental Methods

This dissertation took advantage of many next-generation sequencing (NGS) methods. This section gives a short introduction into the different techniques used and their advantages and disadvantages, as well as the most common analysis strategies.

### RNA-seq and nascent transcription

Measuring the transcriptome of a cell has been a long-standing goal in the research community. It is an essential step in order to understand cell function, development and disease. The earliest methods to quantify ribonucleic acid (RNA) in a large-scale approach go back to the 1990s. The expressed sequence tag (EST) method (Adams et al. (1991)) measures gene expression by partially sequencing complementary DNA (cDNA) clones in order to detect the sequence, as well as the abundance of the transcripts. However, this method is based on Sanger-sequencing and therefore relatively low throughput, as well as very expensive. Subsequently, tag-based methods were developed that cut down on sequencing cost by only sequencing small tags of the mRNA (e.g. serial analysis of gene expression (SAGE) (Velculescu et al. (1995)), massively parallel signature sequencing (MPSS) (Brenner et al. (2000)), and cap analysis of gene expression (CAGE) (Kodzius et al. (2006))). Although cheaper than the EST assay, many of the tag-based methods were still based on Sanger-sequencing resulting in low throughput. Furthermore, many of the tags created in these methods could not be mapped to the reference genome. Another approach was the usage of hybridization methods in which known oligonucleotides complementary to known mRNAs were fluorescently labeled and attached to a microarray (Schena et al. (1995), Lockhart et al. (1996)). After adding mRNA to the array, oligonucleotides that bound to the microarray started to fluoresce and the presence of mRNAs could be detected. These methods are high throughput and inexpensive, however they require previous knowledge of existing transcripts and produce a high level of background noise due to cross-hybridization. Furthermore, these assays are not quantitative and comparison of gene expression levels between different microarrays requires sophisticated normalization methods.

The development of RNA sequencing (RNA-seq) (Nagalakshmi et al. (2008), Ozsolak et al. (2009)) allowed to directly sequence mRNA molecules in a high throughput manner, substantially reducing the amount of background noise, as well as eliminating the need of *a priori* knowledge of the transcriptome. Furthermore, RNA-seq provides a quantitative measurement of mRNA expression, which has been demonstrated by using spike-in RNA controls (Mortazavi et al. (2008)). Due to the higher stability of DNA in comparison to RNA, as well as the limitations provided by most high-throughput sequencing machines, mRNA is usually converted to cDNA. In order to eliminate ribosomal RNA (rRNA) (which makes up approximately 80% of total RNA (Lodish et al. (2002))), two different methods are commonly used (Figure 3A and B). For eukaryotic organisms, enrichment of polyadenylated RNA is one of the most commonly used methods. Almost all protein-coding RNAs, as well as many long non-coding RNAs contain a poly(A) tail, which can be easily targeted

by oligo-dT molecules that are attached to magnetic beads (Figure 3A). Another approach is the removal of rRNA by hybridizing cDNA probes and subsequently digestion of the DNA-RNA hybrid with ribonuclease H (RNase H) (Figure 3B).

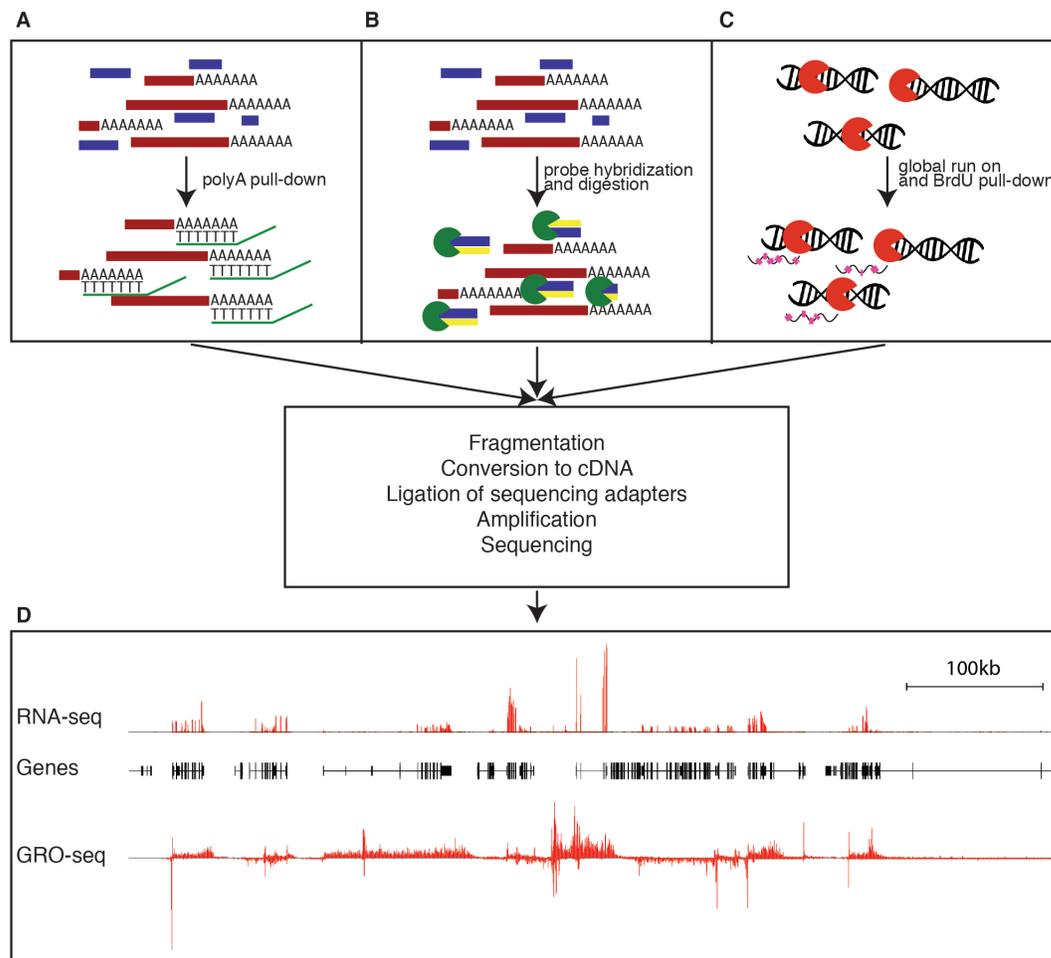


Figure 3: Overview of RNA-seq and GRO-seq method. **A** polyA RNA-seq: genes containing a polyA tail are pulled down using oligo-dT molecules. After library preparation and sequencing, data can be analyzed. **B** Ribo-zero RNA-seq: ribosomal RNA (rRNA) is hybridized with specific probes and subsequently digested, removing all ribosomal RNA in the total RNA. After library preparation and sequencing, data can be analyzed. **C** GRO-seq: transcription is temporarily interrupted and labeled nucleotides are added (run-on) which can subsequently be pulled down. Transcription is started again and biotin labeled RNAs are pulled down, measuring transcription of polymerase directly. After library preparation and sequencing, data can be analyzed. **D**. Example of RNA-seq and GRO-seq data in the mouse genome. RNA-seq signal is only observed over exons, whereas GRO-seq signal can be observed over the whole gene body. Furthermore, bidirectional transcription can be seen in GRO-seq data.

Measurement of the transcriptome gives valuable information about the expressed mRNA,

but does not give any information about pre-spliced RNAs, a more direct read-out of polymerase activity. Furthermore, many active enhancers produce eRNAs, which can not be measured by RNA-seq due to their short half-life time. Therefore, more sensitive methods were developed, that measure nascent transcription (e.g. global run-on sequencing (GRO-seq) (Figure 3C) (Core et al. (2010)) and transient transcriptome sequencing (TT-seq) (Schwalb et al. (2016))). Instead of measuring mRNA, these methods measure nascent transcription, convert RNA to cDNA and sequence these transcripts.

To get equal sequence coverage over the complete gene, a fragmentation step is necessary (e.g. RNA fragmentation with hydrolysis or DNA fragmentation with deoxyribonuclease I (DNase I) treatment or sonication). Each method introduces different biases that need to be accounted for in the analysis step. After library preparation one common step is to size-select for a certain fragment size to eliminate fragments that are too short (often times micro RNAs), as well as very long fragments.

In order to analyze the sequenced libraries, the transcripts are either mapped to an already known reference genome or assembled into contigs for unknown reference genomes. Short reads, as well as highly repetitive reads provide great challenges for the mapping, whereas longer reads or unique sequences can be mapped relatively easily. Most mapping software (e.g. bowtie2 (Langmead and Salzberg (2012)), TopHat (Trapnell et al. (2009)), bwa (Li and Durbin (2009)) or STAR (Dobin et al. (2013))) allow a certain number of mismatches to offset potential sequencing errors. However, a great number of short polymorphisms, as well as very long polymorphisms provide challenges for the mapping tools and might require a more precise annotation of the genome, deeper sequencing, or longer read lengths. RNA-seq data can be used for a variety of different applications (e.g. examination of splice junctions or alternative splicing events (Wang et al. (2008), Sultan et al. (2008), Mortazavi et al. (2008), Cloonan et al. (2008), Trapnell et al. (2010), Griffith et al. (2010), Colla et al. (2015)), detection of gene fusion (Maher et al. (2009), Asmann et al. (2012), Velusamy et al. (2013), Qin et al. (2015)), as well as the description of novel transcripts (Nagalakshmi et al. (2008), Mortazavi et al. (2008), Cloonan et al. (2008), Morin et al. (2008), Lister et al. (2008), Wilhelm et al. (2008)). The most common application to date is the quantification of differently expressed genes in different experiments (Wilhelm et al. (2008), Mortazavi et al. (2008), Lister et al. (2008), Cloonan et al. (2008)).

RNA-seq methods only cover exonic reads, as mRNA is already spliced. GRO-seq however covers the whole gene body, as it measures active transcription pre-splicing. Additionally, GRO-seq also show bidirectional transcription, often a sign for an active enhancer (Figure 3D).

## ChIP-seq and ATAC-seq

Studying epigenetic regulation requires information about the loci of TF binding, histone modifications and open chromatin. One of the earliest methods developed for this is chromatin immunoprecipitation coupled with microarrays (ChIP-chip) (Iyer et al. (2001), Ren et al. (2000)). Following the development of cheap high-throughput sequencing techniques chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq) was

developed (Johnson et al. (2007), Robertson et al. (2007), Barski et al. (2007)). ChIP-seq is one the most commonly used method to detect TF binding, as well as histone modifications, and has led to many important discoveries in the last years (e.g. Deardorff et al. (2012), Schaub et al. (2012), Mikkelsen et al. (2010), Ernst et al. (2011), Gerstein et al. (2010), modENCODE et al. (2010)).

In order to measure TF binding or histone modifications, chromatin is cross-linked and sonicated. The TF (Figure 4A) or histone modification (Figure 4B) is subsequently immunoprecipitated with an appropriate antibody. It is highly advised to keep some fragmented chromatin without pull-down as an input experiment. After pull-down, DNA fragments are sequenced and analyzed. One important consideration for ChIP-seq experiments is that not one single protocol will work in all cell types, under all conditions, and with all antibodies. Furthermore, the quality of the ChIP-seq experiment is highly dependent on the quality of the antibody used for the pull-down. Even different lots of the same antibody can have highly variable degrees of pull-down efficiency. In a study conducted by the model organism ENCODE project (modENCODE et al. (2010)), 25% of antibodies failed specificity tests and another 20% of antibodies failed immunoprecipitation experiments (Egelhofer et al. (2011)). Furthermore, ChIP-seq experiments only report the average binding over the complete cell population per experiment. When ChIP-seq experiments are performed for several factors and binding of these factors is observed at the same locus, it is unclear whether these factors co-bind to this locus in the same cell, or bind to this locus independently or even mutually exclusive in some cells of the population. To address these question, sequential ChIP-seq methods have been used, that use two or more antibodies in sequential order (Mendoza-Parra et al. (2012)).

ChIP-seq assays show the binding of one particular TF or histone modification. In order to get a more general understanding of loci where the chromatin is open (and therefore any TF can bind), several other methods have been developed. Some of the earliest methods were enzyme-based (e.g. MNase-seq (Schones et al. (2008)), DNase I hypersensitivity sites sequencing (DNase-seq) (Sabo et al. (2006))) or sonication-based (e.g. formaldehyde-assisted isolation of regulatory elements (FAIRE) (Giresi et al. (2007)), sonication of cross-linked chromatin sequencing (Sono-seq) (Auerbach et al. (2009))). However, all of these methods require large quantities of cells. Furthermore, the enzyme-based assays are challenging protocols to perform and give plenty of opportunity for failure. Recently, a new method was developed to measure open chromatin, termed assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Figure 4C) (Buenrostro et al. (2013)). This method requires relatively few cells (down to 10,000 cells per assay), is easy to perform (about 1 day of bench work) and gives comparable results to FAIRE and DNase-seq.

Many computational pipelines have been implemented in recent years, however the most fundamental steps are very similar between the different pipelines. Sequencing reads are mapped to a reference genome and subsequently genomic regions are identified that contain an enrichment of sequencing reads greater than the background noise. For ChIP-seq experiments, the usage of an input sequencing experiment is recommended. However, it is not possible to generate an input sequencing experiment for ATAC-seq, requiring more stringent enrichment calls. There are several different biases that can influence the data

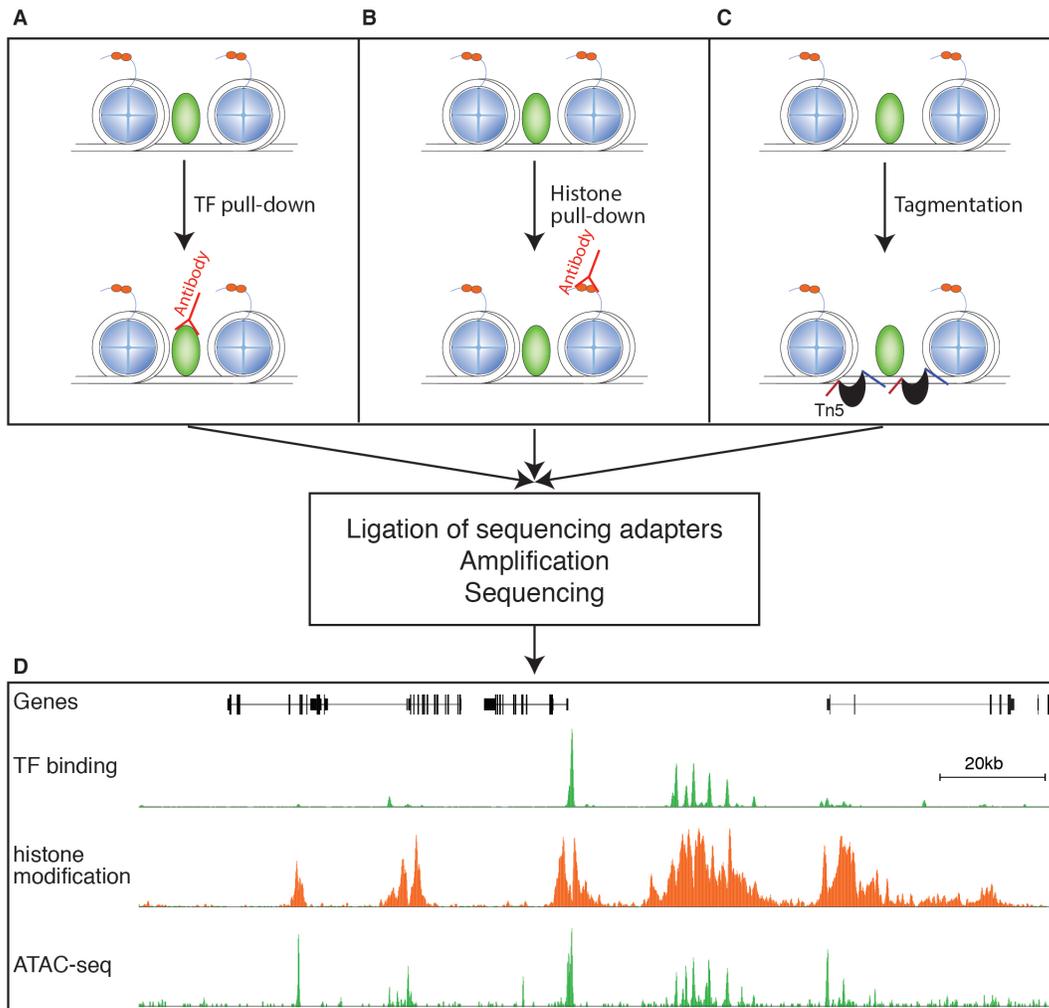


Figure 4: Overview of ChIP-seq and ATAC-seq methods. **A** Transcription factor binding ChIP-seq: a specific antibody for a transcription factor is used for pull-down. After ligation of sequencing adapters, amplification and sequencing the library, data can be analyzed. **B** Histone modification ChIP-seq: a specific antibody for a histone modification is used for pull-down. After ligation of sequencing adapters, amplification and sequencing the library, data can be analyzed. **C** ATAC-seq: Transposase Tn5 with specific barcodes is used to fragment open chromatin. After ligation of sequencing adapters, amplification and sequencing the library, data can be analyzed. **D** Visualization of ChIP-seq data for a transcription factor, showing a very local enrichment of signal and a histone modification, showing a broad signal that spans large domains. ATAC-seq shows similar signals as ChIP-seq, but enrichment is found at more loci, as it measures all open chromatin, not only the binding of one specific transcription factor.

and have been studied extensively (Chen et al. (2010), Khrameeva and Gelfand (2012), Schwartz et al. (2011), Cheung et al. (2011), Minoche et al. (2011), Benjamini and Speed

(2012), Nakamura et al. (2011)). As previously discussed the quality of the antibody for ChIP-seq experiments is crucial for a good enrichment. Furthermore, the machine used for sequencing can introduce biases. Other types of biases include uneven nucleotide distributions across reads, GC content, the distribution of sequencing errors and mappability of the reads. However, commonly used mapping software (e.g. TopHat (Trapnell et al. (2009)), bowtie2 (Langmead and Salzberg (2012)), bwa (Li and Durbin (2009)), STAR (Dobin et al. (2013))), as well as analysis pipelines (e.g. Model-based Analysis of ChIP-Seq 2 (MACS2) (Zhang et al. (2008)) or Hypergeometric Optimization of Motif EnRichment (HOMER) (Heinz et al. (2010))) are accounting for most of these biases.

Genomic regions with enriched signal over noise are called 'peaks'. Peaks can be highly localized signals, showing a sharp peak (Figure 4D), usually observed for TFs and open chromatin. ATAC-seq signal can be found on more loci than ChIP-seq signal, as ATAC-seq measures all regions of open chromatin, not only the binding of one specific TF. Furthermore, peaks can also be signals that span large domains (then often called 'regions') (Figure 4D) usually observed for histone modifications, as well as Pol II. Each of these shapes require different detection strategies. Some tools are able to detect both shapes (e.g. MACS2 (Zhang et al. (2008)) and HOMER (Heinz et al. (2010))), whereas others are specialized for one (e.g. SICER (Xu et al. (2014))).

# Chapter 1

## Environment Drives Selection and Function of Enhancers Controlling Tissue-Specific Macrophage Identities

David Gosselin\*, **Verena M. Link\***, Casey E. Romanoski\*, Gregory J. Fonseca, Dawn Z. Eichenfield, Nathanael J. Spann, Joshua D. Stender, Hyun B. Chun, Hannah Garner, Frederic Geissmann, and Christopher K. Glass

*\* authors contribute equally*

*Cell* (2014) 159, 1327-1340



# Environment Drives Selection and Function of Enhancers Controlling Tissue-Specific Macrophage Identities

David Gosselin,<sup>1,6</sup> Verena M. Link,<sup>1,2,6</sup> Casey E. Romanoski,<sup>1,6</sup> Gregory J. Fonseca,<sup>1</sup> Dawn Z. Eichenfield,<sup>1</sup> Nathanael J. Spann,<sup>1</sup> Joshua D. Stender,<sup>1</sup> Hyun B. Chun,<sup>1</sup> Hannah Garner,<sup>3,4</sup> Frederic Geissmann,<sup>3,4</sup> and Christopher K. Glass<sup>1,5,\*</sup>

<sup>1</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0651, USA

<sup>2</sup>Faculty of Biology, Department II, Ludwig-Maximilians Universität München, Planegg-Martinsried 82152, Germany

<sup>3</sup>Centre for Molecular and Cellular Biology of Inflammation, King's College London, London SE1 1UL, UK

<sup>4</sup>Peter Gorer Department of Immunobiology, King's College London, London SE1 1UL, UK

<sup>5</sup>Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0651, USA

<sup>6</sup>Co-first author

\*Correspondence: [ckg@ucsd.edu](mailto:ckg@ucsd.edu)

<http://dx.doi.org/10.1016/j.cell.2014.11.023>

## SUMMARY

Macrophages reside in essentially all tissues of the body and play key roles in innate and adaptive immune responses. Distinct populations of tissue macrophages also acquire context-specific functions that are important for normal tissue homeostasis. To investigate mechanisms responsible for tissue-specific functions, we analyzed the transcriptomes and enhancer landscapes of brain microglia and resident macrophages of the peritoneal cavity. In addition, we exploited natural genetic variation as a genome-wide “mutagenesis” strategy to identify DNA recognition motifs for transcription factors that promote common or subset-specific binding of the macrophage lineage-determining factor PU.1. We find that distinct tissue environments drive divergent programs of gene expression by differentially activating a common enhancer repertoire and by inducing the expression of divergent secondary transcription factors that collaborate with PU.1 to establish tissue-specific enhancers. These findings provide insights into molecular mechanisms by which tissue environment influences macrophage phenotypes that are likely to be broadly applicable to other cell types.

## INTRODUCTION

Macrophages are phagocytic cells of the innate immune system that populate every organ, making key contributions to their development, functions, and protection against infections and injuries (Geissmann et al., 2010; Gordon et al., 2014; Wynn et al., 2013). Accordingly, each population of tissue macrophages must adapt to its surrounding environment and engage in tissue-specific functions to be effective auxiliary cells. In sup-

port of this, recent mRNA profiling studies revealed significant differences between distinct populations of resident tissue macrophages (Gautier et al., 2012; Okabe and Medzhitov, 2014). Thus, in spite of common elements shared across all subtypes of tissue macrophages, including dependency on the transcription factor PU.1 and signaling downstream of the CSF1 receptor for ontology and survival (Schulz et al., 2012; Wynn et al., 2013), each subset of tissue macrophage possesses its own unique gene expression profile that presumably allows it to function in synergy with the tissue in which it resides.

Accumulating evidence suggests that signaling factors derived from tissue environments play key roles in promoting the ontology and phenotype of the residing macrophage populations. For example, absence of TGF- $\beta$ 1 signaling in the mouse brain impairs the development of the microglia population (Butovsky et al., 2014; Makwana et al., 2007). In the peritoneum, omentum-derived retinoic acid (RA) promotes expression of Gata6 in a subpopulation of local macrophages (Okabe and Medzhitov, 2014). Interestingly, Gata6 expression is exclusive to this particular tissue macrophage population, and decreasing or eliminating its expression interferes with their functions and survival (Gautier et al., 2012, 2014; Okabe and Medzhitov, 2014; Rosas et al., 2014).

Precisely how these and other signals act on macrophages at the genomic level to promote specialized phenotypes and unique transcriptional signatures remains unknown. However, strong evidence suggests that enhancers, which are fundamental determinants of gene expression, may play a key role in this context (Andersson et al., 2014; Levine, 2010; Shlyueva et al., 2014). Enhancers, in comparison to promoters, exhibit significant enrichment for combinations of DNA recognition motifs that correspond to binding sites for lineage-determining transcription factors (LDTFs), which are required for the development of distinct cell types. Different patterns of LDTF expression drive the selection of cell-specific repertoires of enhancers that are considered to be central to the establishment of cell identity and regulatory potential.

Studies of primary macrophages and B cells indicated that PU.1 acts as an essential LDTF that contributes to the selection

of a large fraction of the cell-specific enhancer-like elements in each of these cell types (Barozzi et al., 2014; Ghisletti et al., 2010; Heinz et al., 2010). Macrophage-specific enhancer selection by PU.1 required collaborative interactions with additional macrophage-restricted transcription factors (TFs), including C/EBP and AP-1 factors (Heinz et al., 2013). In contrast, B-cell-specific enhancer selection by PU.1 required collaborative interactions with B-cell-restricted factors, including EBF and E2A (Heinz et al., 2010).

Pre-existing enhancer landscapes occupied by PU.1 and/or C/EBP factors were shown to be the major sites that bound signal-dependent transcription factors (SDTFs), such as NF $\kappa$ B, nuclear receptors, and STAT proteins (Ostuni et al., 2013; Heinz et al., 2010). A similar hierarchical relationship for LDTFs and SDTFs was found in regulatory T cells, embryonic stem cells, and dendritic cells (Mullen et al., 2011; Samstein et al., 2012; Garber et al., 2012). The collaborative and hierarchical relationship of LDTFs and SDTFs at pre-existing enhancers was validated at the level of the DNA template by studies of effects of natural genetic variation on enhancer selection and function (Heinz et al., 2013). Mutations in PU.1 motifs causing loss of PU.1 binding resulted in loss of the collaborative binding of C/EBP $\alpha$ . Conversely, mutations in C/EBP motifs causing loss of C/EBP $\alpha$  binding resulted in a loss of collaborative binding of PU.1. Either type of mutation abolished signal-dependent binding of NF $\kappa$ B, whereas mutations in NF $\kappa$ B motifs that abolish NF $\kappa$ B binding rarely affected the binding of PU.1 or C/EBP $\alpha$ . However, in contrast to the picture at pre-existing enhancers, NF $\kappa$ B was also shown to be capable of selecting “latent” or “de novo” enhancers by collaborating with PU.1 to bind to genomic locations lacking prior features associated with active enhancers (Kaikkonen et al., 2013; Ostuni et al., 2013). These observations provide an example of an environmentally driven modification of the enhancer repertoire by a broadly expressed SDTF that is nonetheless cell type specific due to the obligatory participation of PU.1.

Given that each tissue environment is distinguished by a unique combination of signaling factors, it is likely that gene expression in each corresponding macrophage population is under the control of distinct combinations of SDTFs that can modulate the activity of a pre-existing enhancer repertoire to achieve context-dependent gene expression. In addition, it is also possible that environmental signals control the expression and activities of TFs that result in selection of tissue-specific enhancers, analogous to the establishment of “latent” or “de novo” enhancers. Here, we sought to determine the extent to which environment shapes distinct macrophage enhancer repertoires and the underlying mechanisms.

## RESULTS

### Environment-Specific Gene Expression

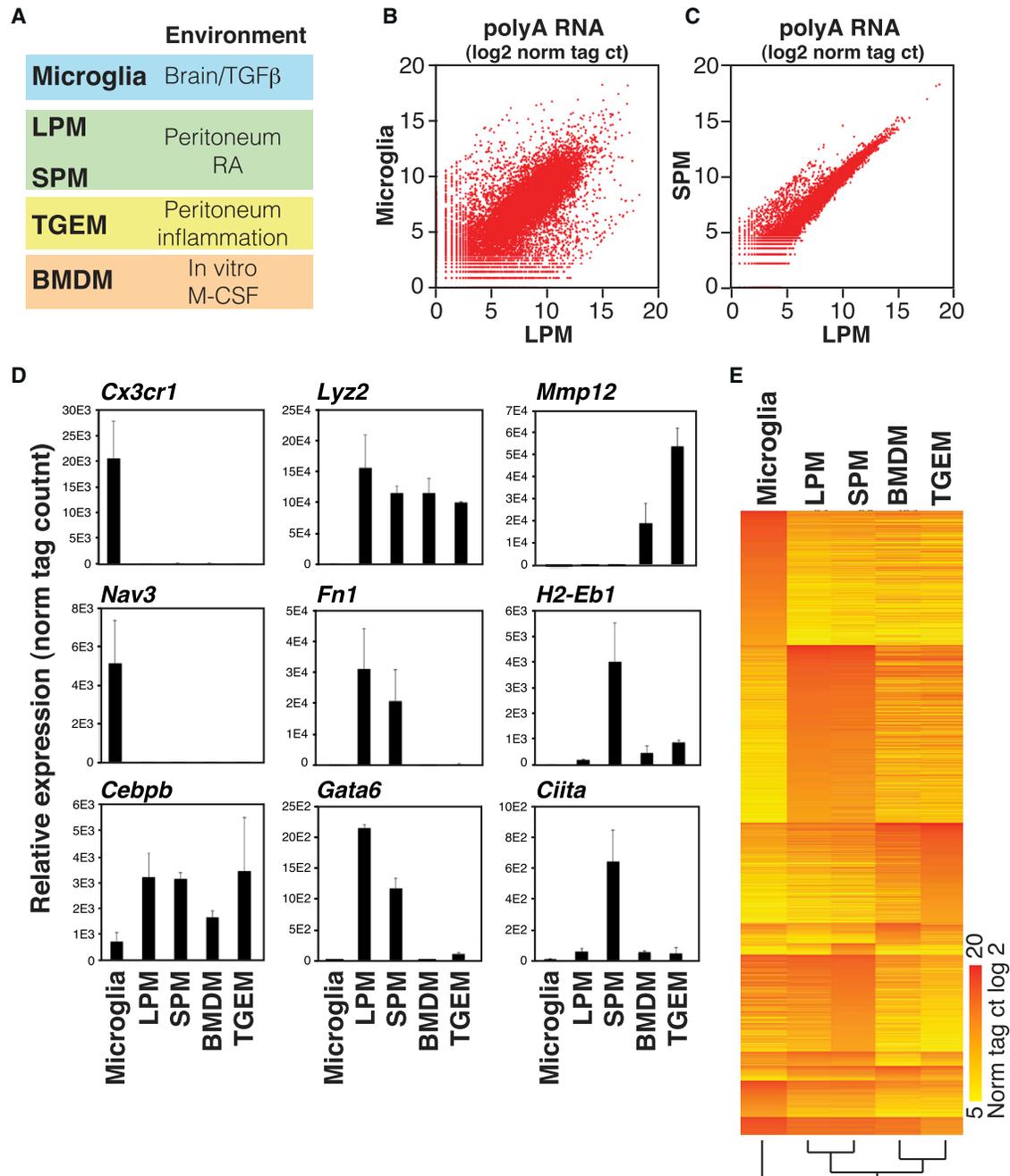
To investigate mechanisms responsible for tissue-specific macrophage phenotypes, we isolated microglia (MG; brain macrophages) and two distinct populations of resident peritoneal macrophages (RPMs) that are discriminated by cell-surface expression levels of MHCII—large peritoneal macrophages (LPMs, low MHCII) and small peritoneal macrophages (SPMs,

high MHCII)—by flow cytometry (Figures S1A and S1B available online) (Ghosn et al., 2010; Okabe and Medzhitov, 2014) (Figure 1A). These three populations of macrophages allow comparisons of gene expression and epigenetic landscapes in distinct macrophage populations residing in the same environment (i.e., LPMs versus SPMs), as well as different environments (i.e., LPM versus MG). In addition, we included thioglycollate-elicited peritoneal macrophages (TGEMs) and bone-marrow-derived macrophages (BMDMs) for comparison, as these macrophages, although maintained in culture conditions, are widely used models of macrophage biology that are derived from different sources (Figure 1A).

Gene expression profiles determined by RNA sequencing (RNA-seq) from independent biological replicates revealed substantial differences in the patterns of gene expression across the different macrophage populations examined (Figures 1B, 1C, and S1C and Table S1), in agreement with previous studies (Gautier et al., 2012; Okabe and Medzhitov, 2014). In particular, ~7,000 genes are differently expressed in MG compared to LPMs (p value < 0.01), with >500 genes being >16-fold more highly expressed in MG and >600 genes being >16-fold more highly expressed in LPMs. On the other hand, LPMs and SPMs share strong similarities (Figure 1C), with SPMs expressing only 108 genes > 16-fold higher than LPMs, and LPMs expressing only 5 genes > 16-fold higher than SPMs. These results corroborate many previous findings, including the highest level of expression of *Cx3cr1* in MG and the selective expression of *Gata6* in RPMs (Figure 1D) (Cardona et al., 2006; Gautier et al., 2012; Jung et al., 2000; Okabe and Medzhitov, 2014). Interestingly, *Ciita*, a transcription factor that regulates MHCII expression (Steimle et al., 1993), is preferably expressed in the SPM population (Figure 1D). Finally, gene clustering analyses confirmed that, whereas LPMs and SPMs show highly similar gene expression, MG differ substantially from the other macrophage subsets (Figure 1E). TGEMs and BMDMs are also more similar to one another than either one is to any of the three in vivo subsets, potentially reflecting the similarity of the cell culture environment. Overall, these findings suggest a strong role of environment in determining macrophage gene expression.

### Common and Distinct Macrophage Enhancer Repertoires

The dissimilarities in gene expression between different macrophage subsets revealed by RNA-seq analysis imply important differences in how these cells organize and/or use their enhancer repertoires. To examine this, we analyzed dimethylation status of lysine 4 of histone 3 (H3K4me2) and acetylation status of lysine 27 of histone H3 (H3K27ac) by chromatin immunoprecipitation sequencing (ChIP-seq) in these cells (Figure S2 and Tables S2, S3, and S4). H3K4me2 marks promoters and enhancers (He et al., 2010; Kaikkonen et al., 2013), whereas H3K27ac correlates positively with transcriptional activity at these elements (Creighton et al., 2010). Deposition of H3K4me2 results from the binding of LDTFs and other TFs but is not necessarily associated with enhancer activity. We therefore use a heuristic of defining H3K4me2-positive/H3K27ac-negative regions as “primed” and regions positive for both marks as “active.” Genomic annotation enabled segregation of these regions into



### Figure 1. Variation in Gene Expression in Different Macrophage Subsets

(A) Macrophage subsets used for analysis and corresponding environmental factors (see Figures S1A and S1B for sorting protocols).

(B and C) Scatterplots illustrating relative gene expression of polyA-selected RNA transcripts in MG compared to LPMs (B) and SPMs compared to LPMs (C). Values are log<sub>2</sub> of tag counts normalized to 10<sup>7</sup> uniquely mapped tags. See Figure S1C for a representative replicate.

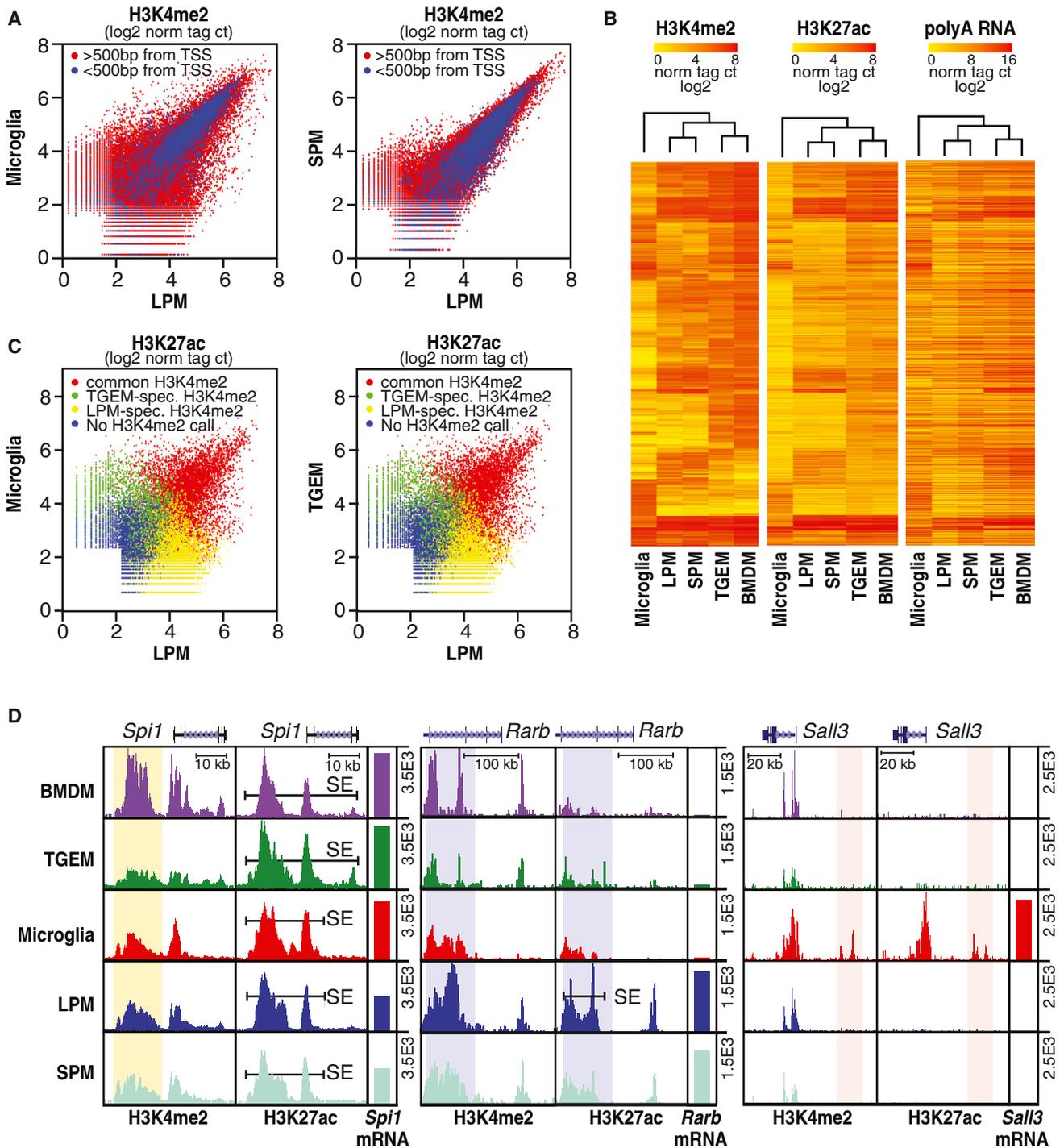
(D) Relative gene expression means for the indicated genes are shown from replicate RNA-seq experiments (error bars represent SD).

(E) Heat map of transcripts exhibiting an expression value of at least 64 normalized tags in at least one subset and differing in expression by at least 16-fold in at least one of the indicated subsets.

See also Table S1.

promoters or enhancers by proximity to gene transcriptional start sites (TSS). Notably, the pattern of H3K4me<sub>2</sub> deposition in MG substantially differs from that of LPMs (Figure 2A), indicating selection of distinct regulatory landscapes. Of 7,937

promoters marked by H3K4me<sub>2</sub> in one or both subsets, 275 exhibit >4-fold differences (3%), far fewer than the ~1,700 mRNAs exhibiting >16-fold differences in expression. Of 36,607 regions > 500 bp from TSS marked by H3K4me<sub>2</sub> in one



**Figure 2. Variation in Enhancer Landscapes in Different Macrophage Subsets**

(A) Scatterplots of normalized H3K4me2 tag counts at genomic regions marked by significant H3K4me2 tags in LPMs and/or MG (left) or LPMs and/or SPMs (right). Points colored in blue are within 500 bp of a TSS. See Figure S2 for representative replicates.

(B) Heatmaps of normalized H3K4me2, H3K27ac, and nearest expressed gene RNA-seq tag counts at genomic locations showing >4-fold pairwise differences in H3K4me2 tag counts between at least two of the five macrophage subtypes. Row order is the same for all three data types.

(C) Scatterplots of normalized H3K27ac tag counts at genomic regions marked by significant H3K27ac tags in LPMs and/or MG (left) or LPMs and/or SPMs (right). Points are colored red if genomic locations are also marked by H3K4me2 (>16 tags) in both subsets, green if marked by H3K4me2 selectively in MG (left) or SPMs (right), yellow if marked by H3K4me2 selectively in LPMs, or blue if not associated with H3K4me2 in either subset.

(legend continued on next page)

or both subsets, 9,083 exhibit >4-fold differences (24%). The vast majority of differential H3K4me2-marked regions are thus distant from promoters and correspond to potential enhancers. In contrast to the comparison of LPMs and MG, both the enhancer and promoter repertoires of the two subsets of RPMs share a much higher degree of similarity (Figure 2A). Furthermore, clustering analyses of the H3K4me2 deposition pattern revealed that MG were more divergent from the other subsets than any two other macrophage subsets are from one another, which is consistent with gene expression data (Figure 2B).

H3K27ac was present at a large fraction of H3K4me2-marked regions and generally but imperfectly correlated with nearest gene expression (Figure 2B). Overlap of the H3K27ac data with H3K4me2-defined enhancers allowed the identification of common but quantitatively differently activated enhancers, as well as activation of enhancers unique to one subset. Figure 2C illustrates such comparisons for LPMs versus MG and LPMs versus TGEMs. Genomic regions marked by H3K4me2 in both subsets are color coded in red and represent activation of an enhancer landscape that is primed in both subsets. In contrast, regions exclusively marked by H3K4me2 in LPMs, shown in yellow, represent LPM-specific enhancers. Conversely, regions exclusively marked by H3K4me2 in MG or TGEMs, indicated in green, represent MG or TGEM-specific enhancers, respectively. Comparing LPMs versus MG, 60% of the active enhancers resided at common regions of H3K4me2, 30% at LPM-specific regions, and 10% at MG-specific regions. Specific examples are indicated in Figure 2D. As expected, the *Spi1* enhancer, controlling expression of PU.1, is marked by H3K4me2 and H3K27ac in all macrophage populations. Interestingly, the RA-inducible *Rarb* gene is also marked by H3K4me2 in all macrophage populations, but high H3K27ac is only observed in LPMs and SPMs, suggesting a role of local RA in enhancer activation. Finally, the *Sal3* gene, which is exclusively and highly expressed in MG, is near a genomic region that is exclusively marked by H3K4me2 and H3K27ac in MG. In sum, these analyses provide strong evidence that both differential activation of a common enhancer landscape and the selection of subtype-specific enhancers contribute to the specific transcriptional signature of each subset of macrophages.

### Tissue-Specific Super-Enhancers Emerge from Common Enhancer Landscapes

Genome-wide analysis of features of active enhancers, including the presence of Mediator and deposition of H3K27ac, indicates marked variation in their local distribution patterns. In all cell types evaluated thus far, ~400–800 regions, representing a small fraction of the genome, exhibit a disproportionately high density of active regulative marks and transcription factor binding (Hnisz et al., 2013; Lovén et al., 2013; Whyte et al., 2013). These regions, recently termed super-enhancers (SEs), are selected in a cell-specific manner and frequently occur near or encompass

genes that play essential roles in defining the identity and function of the corresponding cell type (Hnisz et al., 2013). Although LDTFs are enriched in and likely determine cell-specific SE selection, evidence also suggests that the extracellular environment can influence formation of SEs in endothelial cells (Brown et al., 2014). To investigate this relationship in tissue macrophages, we defined SEs in each macrophage subset based on H3K27ac ChIP-seq. In agreement with previous studies, we observed common and subset-specific SEs, with ~600 to 750 SEs being identified among the five cell types examined. Clustering of these SEs results in the same relationships between subsets as observed using RNA-seq, H3K4me2, or H3K27ac data (Figure 3A). This analysis also revealed a high concordance between the distribution of SEs genome wide and the expression level of the nearest genes (Figure 3A). This strong relationship is further illustrated for SEs and nearest gene expression in MG and LPMs, in which the correlation coefficient was 0.62 (Figure 3B), much higher than that observed for the individual enhancer elements not associated with SE regions in these subsets. This may be due to a more accurate assignment of SEs to their target genes than conventional enhancers.

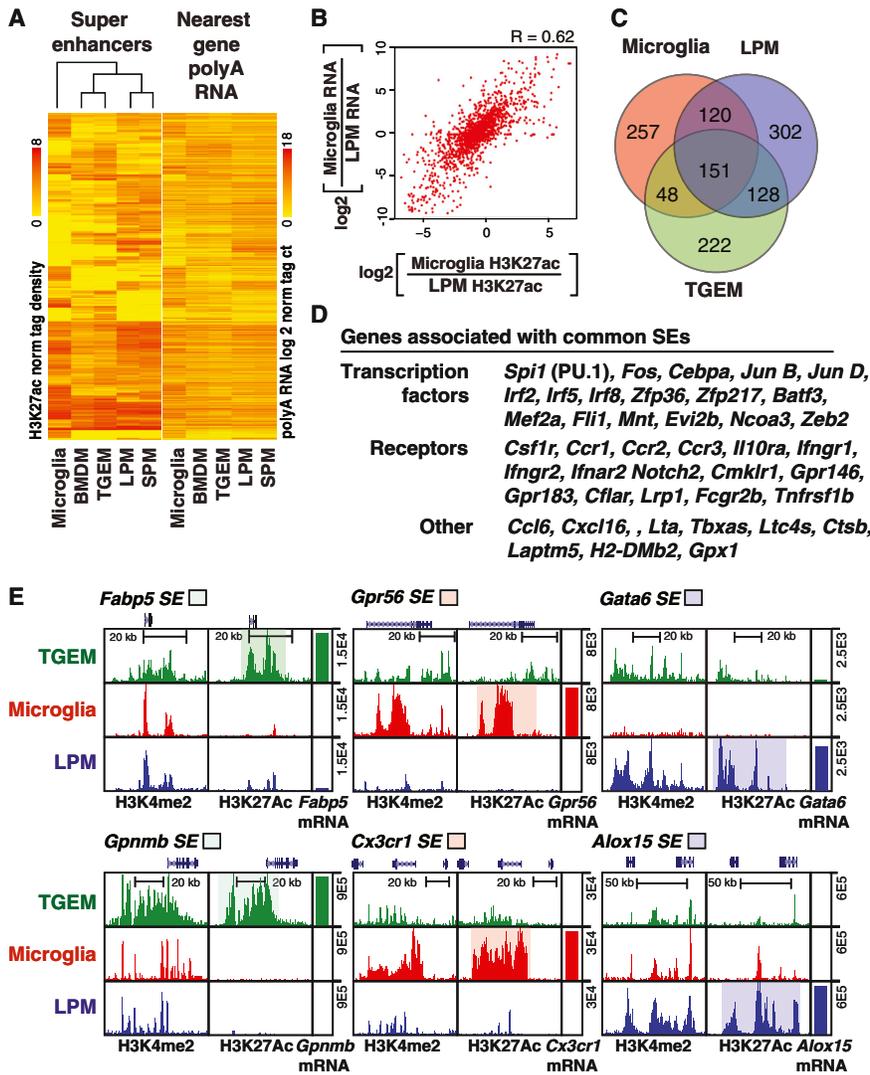
Approximately 40% to 50% of the SEs in a particular macrophage subset are unique to that subset, illustrated by the Venn diagram of LPM, MG, and TGEM in Figure 3C. In concert with previous findings (Hnisz et al., 2013; Whyte et al., 2013), common SEs are associated with numerous genes important to macrophage ontology and functions, including *Spi1*, *Cebpa*, members of the Irf family, *Csf1r*, *Fcgr2b*, *Ctsb*, etc. (Figure 3D). This pattern is exemplified by the region upstream of *Spi1*, which is scored as a SE in all five subsets (Figure 2D). In contrast, many SEs are macrophage subset specific and reside near or surround genes that are highly differentially expressed (Figure 3E). Although some SEs exhibit highly specific H3K4me2 and H3K27ac markings, such as the LPM-specific SE upstream of *Gata6* (Figure 3E), the majority of SEs are located at regions that are marked by H3K4me2 in multiple macrophage subsets but only attain SE status in one or a few subsets. For example, LPM-specific SEs reside in the vicinity of *Rarb* (Figure 2D) and *Alox15* (Figure 3E) genes, which are selectively expressed in LPMs but that also exhibit H3K4me2 in other macrophage subsets. Similar relationships are observed for the MG-specific SEs surrounding *Gpr56* and *Cx3cr1* and the TGEM-specific SEs surrounding *Fabp5* and *GpnmB* (Figure 3E). These findings suggest that environmental signals play roles in the transition of collections of primed enhancers to genomic regions exhibiting features of SEs.

### PU.1 Colocalizes with Distinct TF Motifs at Subset-Specific Enhancers

The observation that PU.1 localization to macrophage- or B-cell-specific enhancers is dependent on collaborative interactions with alternate LDTFs (Heinz et al., 2010) led us to consider the possibility that an assessment of PU.1 binding in different macrophage subsets might yield insights into the TFs that

(D) UCSC browser images of selected genomic regions with corresponding RNA-seq data plotted as bar graphs. Bars labeled SE indicate super-enhancers, and vertical highlights designate regions of interest for subset-common (*Spi1*) or subset-specific (*Rarb* and *Sal3*) loci. All data are normalized to input and library dimension.

See also Tables S2, S3, and S4.



**Figure 3. Variation in Super-Enhancer Landscapes in Different Macrophage Subsets**

(A) Heatmaps of H3K27ac tag densities at super-enhancers and RNA-seq tag densities at nearest genes. Rows are ordered the same for both plots. (B) Scatterplot of the relationship between ratio of MG to LPMs H3K27ac tag density at super-enhancers (x axis) and the ratio of nearest gene expression (y axis). (C) Venn diagram indicating overlap and specificity of super-enhancers in MG, LPMs, and TGEMs. (D) Examples of genes associated with common super-enhancers. (E) UCSC genome browser images of selected subset-specific super-enhancers and associated genes with subset-specific regions of interest highlighted.

IRF, KLF, and GATA transcription factor family members (Figure 4C). Conversely, MG-specific PU1-binding sequences were coenriched for a PU.1-IRF composite sequence and motifs corresponding to CTCFL, HIC2, MEF2, and SMAD TFs (Figure 4D). In addition, by using alternative subset-specific PU.1-binding sites as background, motifs recognized by retinoic acid receptors (e.g., NR2F2) were identified to be coenriched with PU.1-binding sites in LPMs (Figure 4C).

Previous studies indicated that motifs for collaborative binding partners of PU.1 typically reside within ~100 bp of the PU.1 motif itself (Barozzi et al., 2014; Heinz et al., 2010). We therefore analyzed the genomic distance distribution of enriched motifs (from Figures 4C and 4D) within a 400 bp window relative to the

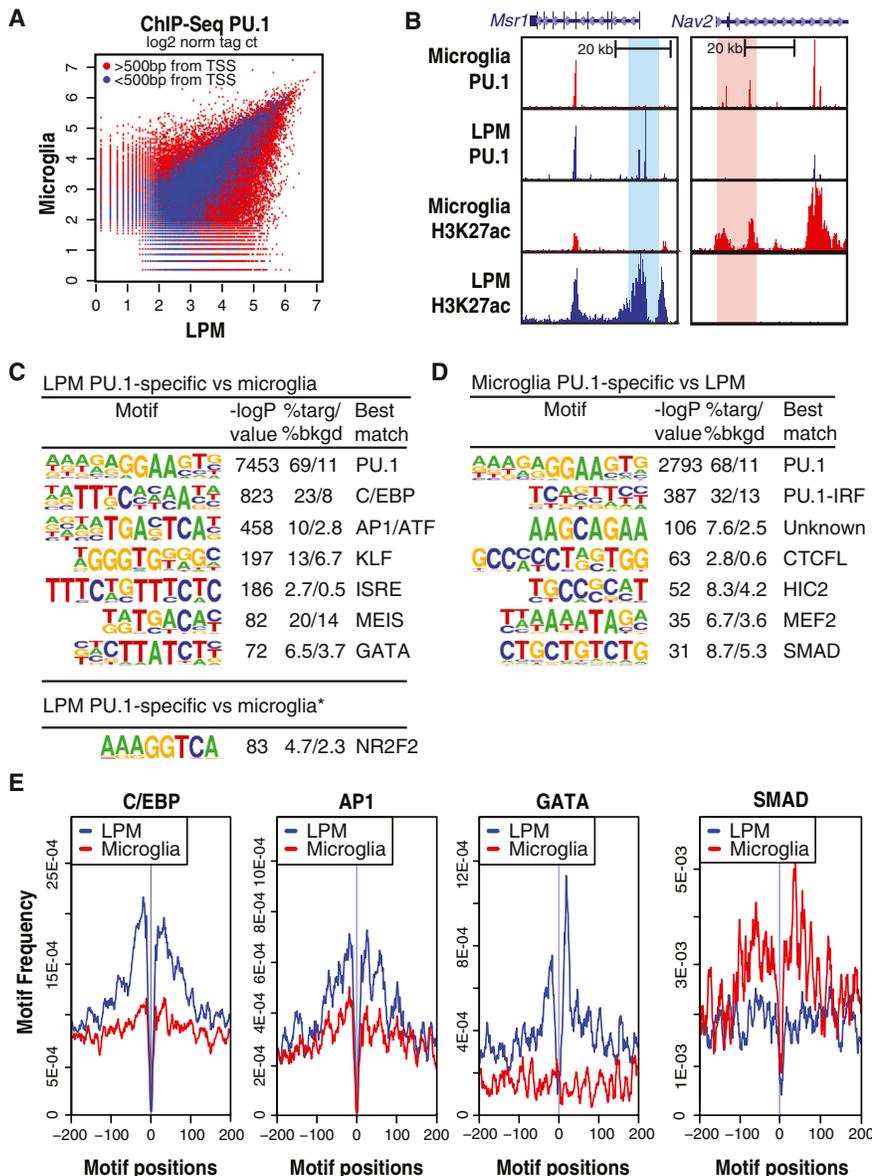
drive the selection of subset-specific enhancers. We therefore extended existing genome-wide binding profiles for PU.1 to include MG, LPMs, and SPMs. These studies indicated that PU.1 bound to both common and subset-specific genomic locations, exemplified for LPMs and MG in Figure 4A (all comparisons in Tables S2 and S3). The great majority of subset-specific binding sites were observed at distal regions (>500 bp from an mRNA TSS, Figure 4A), which is consistent with the patterns of H3K4me2 (Figure 2A). Examples of LPM-specific and MG-specific binding sites for PU.1 in enhancer-like regions vicinal to *Msr1* (expressed exclusively in LPMs) and *Nav2* (expressed exclusively in MG) genes are illustrated in Figure 4B.

De novo motif enrichment analysis of 200 bp sequences encompassing PU.1 peaks identified the identical PU.1 recognition motif in both LPMs and MG as the most enriched sequence. However, completely different motifs were coenriched within the two subsets (Figures 4C and 4D). Using GC content-matched genomic sequence as background, enriched sequences specific to LPMs corresponded to motifs known to bind C/EBP, AP-1,

bound PU.1 motif of LPM- and MG-specific PU.1 peak sets (Figure 4E). This analysis indicated that C/EBP, AP1, and GATA motifs frequently occurred near PU.1-bound motifs in LPMs, but not in MG, indicating that genomic loci containing PU.1 and closely spaced C/EBP, AP-1, or GATA motifs were more likely to become LPM-specific enhancers. The GATA motif was selectively enriched in LPMs relative to MG, suggesting a fundamental difference for the LPM resident population compared to elicited macrophages (Figure 4E). In contrast, the SMAD motif showed MG specificity (Figure 4E), which is consistent with TGFβ signaling in the brain. These findings provide evidence that selection of subset-specific enhancers is in part driven by collaborative interactions between PU.1 and alternative sets of TFs in each subset.

**Use of Natural Genetic Variation to Validate and Discover Collaborative TFs**

Although motif enrichment suggests the identities of TFs that contribute to the function of subset-specific enhancers, this



**Figure 4. PU.1 Binds to Subset-Specific Enhancers**

(A) Scatterplot of normalized tag counts for PU.1 peaks in MG versus LPMs. Points colored blue are within 500 bp of the TSS.

(B) UCSC genome browser images of PU.1 binding in the vicinity of the *Msr1* and *Nav2* genes in MG and LPMs cells and association with H3K27ac highlighting specific regions.

(C) Motifs enriched in the vicinity of PU.1-binding sites that are specific for LPMs versus MG using a random GC-corrected genomic background (top) or a background corresponding to MG-specific PU.1 peaks (bottom).

(D) Motifs enriched in the vicinity of PU.1-binding sites that are specific for MG using a random GC-corrected genomic background.

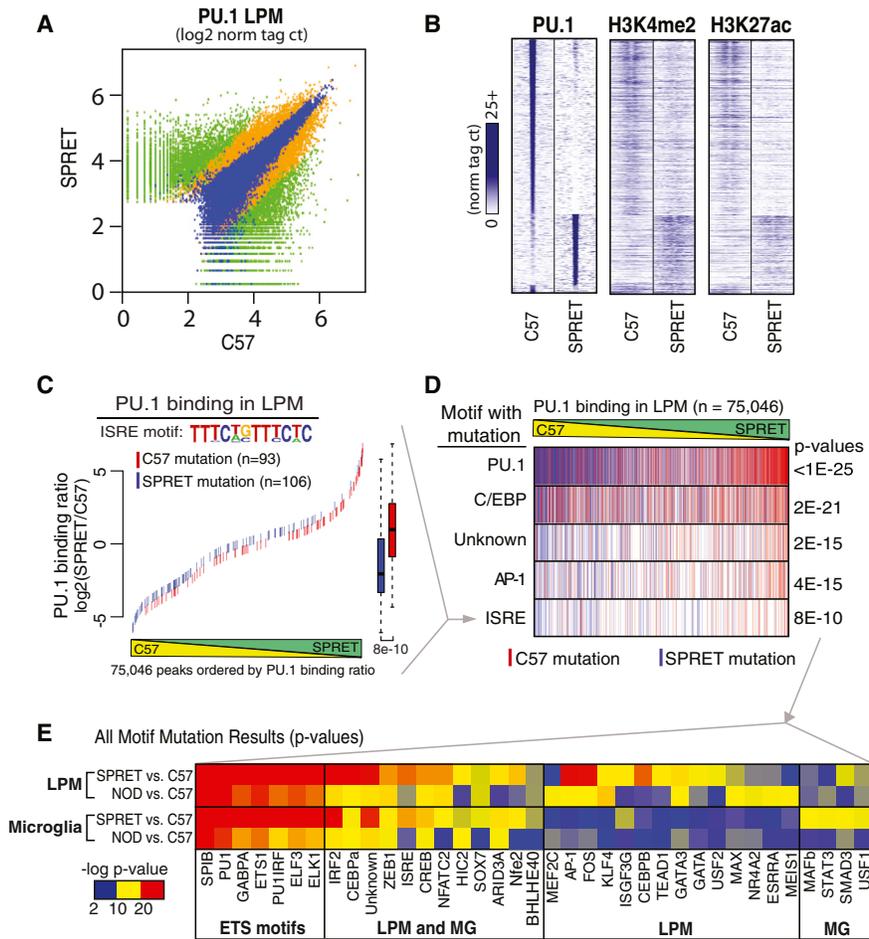
(E) Distribution plots of motif frequencies (y axis) for the indicated motifs within 400bp centered on the PU.1 motif at genomic loci bound specifically by PU.1 in LPMs (blue) or MG (red).

Compared to C57BL/6J (C57) mice, NOD mice have about 5 million SNPs and indels, whereas SPRET mice have about 40 million (Keane et al., 2011). This variation is associated with corresponding levels of strain-specific binding of PU.1, illustrated for LPMs derived from C57 and SPRET mice (Figure 5A). Similar observations are made with respect to MG (Table S6). Approximately 8-fold fewer strain-specific PU.1-binding sites were identified in LPMs and MG derived from NOD mice compared to C57, which is consistent with the lower number of variants between these two strains. Strain-specific binding of PU.1 was associated with corresponding strain-specific H3K4me2 and H3K27ac marks (Figure 5B), suggesting that many strain-specific PU.1-binding sites localize to functional enhancers.

approach does not establish whether or not they are required for collaborative binding. Loss-of-function strategies are challenging for this purpose because many of the identified motifs are recognized by multiple members of corresponding TF families. An alternative means to test for collaborative binding is to mutate motifs recognized by the TF family of interest and determine whether this results in loss of binding of a nearby factor. We considered the possibility that this could be accomplished for informative motifs on a genome-wide scale by leveraging the vast degree of natural genetic variation provided by inbred laboratory and wild strains of mice.

To explore the potential of this approach to validate and discover TFs required for collaborative binding and function of PU.1, we determined the genome-wide patterns of PU.1, H3K4me2, and H3K27ac in LPMs and MG isolated from NOD/ShiLtJ (NOD) and SPRET/EiJ (SPRET) mice (Table S5).

To search for motifs mediating DNA binding by collaborative TFs, we analyzed strain-specific binding of PU.1 that was not associated with mutations in PU.1 recognition motifs. This was accomplished by scanning a 200 bp window surrounding PU.1-binding sites lacking PU.1 motif mutations for the presence of the DNA recognition motifs of the 100 most highly expressed TFs in LPMs and MG in C57 or the alternate (NOD or SPRET) genomic sequence. Mutated loci were then queried for a corresponding decrease in PU.1 binding relative to the unmutated strain. The significant result for ISRE motif mutations affecting PU.1 binding in LPMs is exemplified in Figure 5C. The ISRE was found to be mutated in the vicinity of PU.1-binding sites 93 times in LPMs isolated from C57 mice (indicated by red hash lines in Figure 5C) and 106 times in LPMs isolated from SPRET mice (indicated by blue hash lines in Figure 5C). PU.1 binding strength is rank ordered from most C57 specific at left



**Figure 5. Motif Mutations in Potential PU.1 Collaborating Transcription Factors Confirm Cooperative Binding for Subset-Common and Subset-Specific Factor Combinations**

(A) PU.1 binding between SPRET and C57 is shown for 200 bp regions where green signifies differential binding (>4-fold,  $p < 1 \times 10^{-4}$ ,  $n = 13,199$ ), blue similar binding (<4-fold,  $p < 1 \times 10^{-4}$ ,  $n = 11,022$ ) and orange in between ( $n = 12,367$ ).

(B) Heatmap of 2 kb differentially bound PU.1 genomic regions (rows) centered on PU.1 binding for ChIP-seq tags of PU.1, H3K4me2, and H3K27ac between C57 and SPRET (columns).

(C) An example of motif mutation analysis is shown for the ISRE motif. 200 bp genomic sequence at all PU.1 bound loci (in A) were queried for genetic variants that mutated the ISRE motif matrix in either C57 or SPRET. Mutations were colored according to the genome mutated: red, C57; blue, SPRET. ISRE mutations were plotted according to the PU.1-binding strain ratio (y axis) as measured in LPMs at that locus and rank-ordered on the x axis. Boxplots of corresponding color indicate the effect of ISRE motif mutations on PU.1 binding where whiskers extend to data extremes and p value are from two-sided t test.

(D) Results from analyses described in (C) are vertically compressed and shown in rows for PU.1, C/EBP, Unknown, AP-1, and ISRE motif mutation events.

(E) Heatmap showing p values resulting from analysis described in (C) and (D) for motif mutations best matching transcription factors indicated on x axis. Each motif was tested for affecting PU.1 binding between C57 and NOD and between C57 and SPRET both in MG and LPMs (y axis). See also Figure S3 and Tables S5 and S6.

to most SPRET specific at right. Many mutations are not associated with strain-specific binding, which is consistent with prior studies indicating that the specific position of the variant (i.e., core versus periphery of motif), the distance of the motif from the peak center, and presence of additional redundant motifs affect the impact of individual mutations (Heinz et al., 2013). Overall, however, C57 mutations in the ISRE were associated with SPRET-specific binding of PU.1, whereas SPRET mutations in the ISRE associated with C57-specific binding of PU.1 ( $p = 8 \times 10^{-10}$ ). This strong genetic association implicates factors binding to the ISRE as collaborative partners of PU.1 in LPMs.

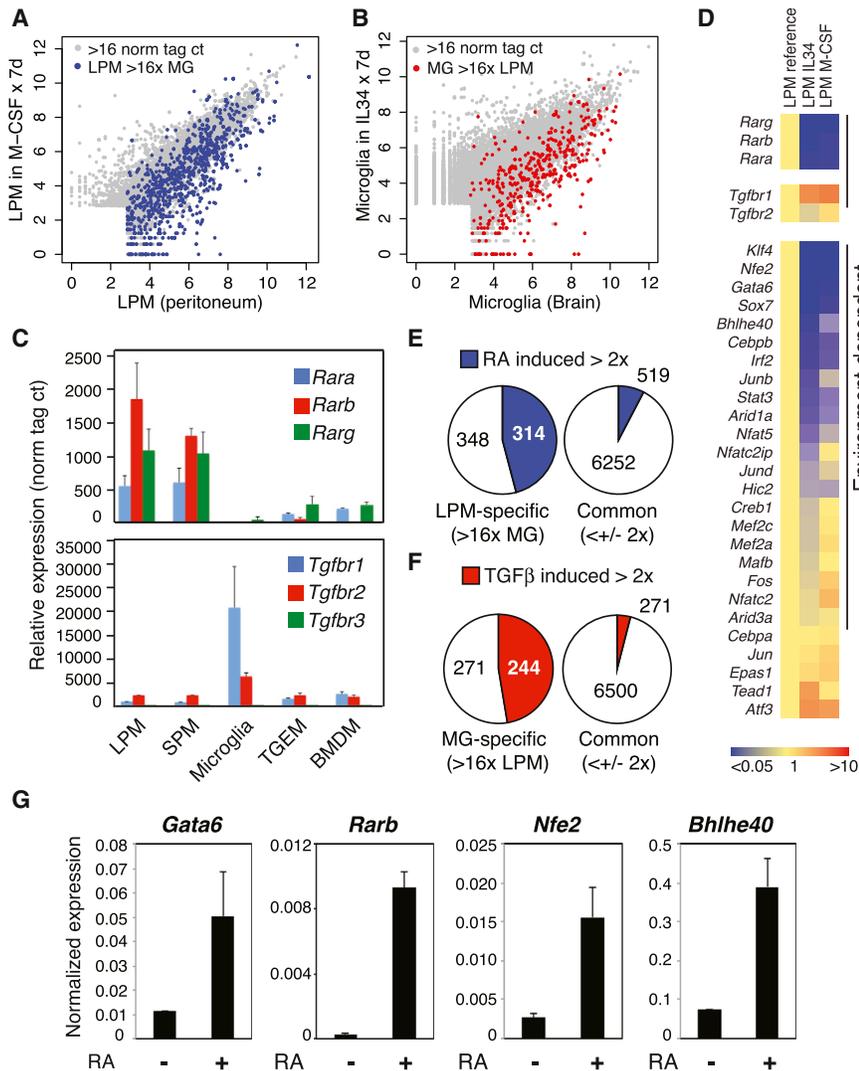
This analysis was repeated for each motif of interest in each macrophage subset (LPMs and MG) for the comparisons of C57 versus NOD and C57 versus SPRET. Vertical compression of the plot shown in Figure 5C allows stacking of plots for multiple motifs, indicated in Figure 5D. Overall, 37 motifs were found to reach statistical significance in at least one macrophage subset and strain (Figures 5E and S3). Many more motifs were found to be significant in comparisons of macrophages derived from C57 and SPRET mice than C57 and NOD mice, which is consistent with the much larger number of informative mutations. The most highly significant motifs corresponded to sequences recognized by ETS factors that are similar to motifs recognized

by PU.1 itself. Most of these motifs are closely situated to regions of PU.1 binding for which the PU.1 motif itself was considered to be intact. It is therefore unclear at present the extent to which these mutations directly affect PU.1 binding or represent binding sites for collaborative ETS factors.

Twelve non-ETS motifs were identified as being significantly associated with PU.1 binding in both LPMs and MG, including a C/EBP $\alpha$  motif, which is an established collaborative binding partner of PU.1 (Figure 5E). In contrast, 14 motifs exhibited preferential associations with strain-specific PU.1 binding in LPMs (Figure 5E). This list includes recognition motifs for KLF4, GATA factors, and AP-1 factors, independently identifying motifs discovered through de novo motif analysis. Finally, motifs for four factors were preferentially associated with PU.1 binding in MG, including a recognition motif for SMAD3.

**Tissue Environment Regulates Collaborative and Signal-Dependent TF Expression**

To investigate the importance of tissue environment in maintenance of specific macrophage phenotypes, we placed LPMs and MG into culture under the influence of IL-34 or M-CSF for 7 days. Whereas M-CSF is important to peritoneal macrophages (Witmer-Pack et al., 1993), IL-34 is critical for proper MG



**Figure 6. Environmental Influence on Gene Expression in LPMs and Microglia**

(A and B) Scatterplots illustrating relative gene expression of RNA transcripts in freshly isolated LPMs compared to LPMs maintained in culture for 7 days (A) and freshly isolated MG compared to MG in culture for 7 days (B). Genes specific to LPMs are colored blue in (A) and specific to MG are red in (B).

(C) Normalized gene expression values for members of the RAR and TGFβ receptor family members.

(D) Heatmap showing the fold-change of RNAs for the indicated transcription factors upon removal from the peritoneal cavity and culture with IL-34 or M-CSF.

(E and F) Effects of chronic stimulation with RA in M-CSF and/or IL34 (E) on LPM-specific or common mRNAs or TGFβ in M-CSF or IL34 (F) on MG-specific or common mRNAs.

(G) qPCR validation of maintained expression by RA of key transcription factors in cultured LPMs (error bars indicate SD).

See also Figure S4.

in each macrophage subset. The mRNAs encoding all three RA receptors (*Rara*, *Rarb*, and *Rarg*) are highly and selectively expressed in LPMs and SPMs, whereas mRNAs encoding the TGFβ receptors *Tgfb1* and *Tgfb2* are preferentially expressed in MG (Figure 6C). Interestingly, expression of all three retinoic acid receptors is markedly reduced when LPMs are placed into culture in the presence of M-CSF or IL-34, whereas the expression of *Tgfb1* is markedly increased under these conditions (Figure 6D). Thus, environment

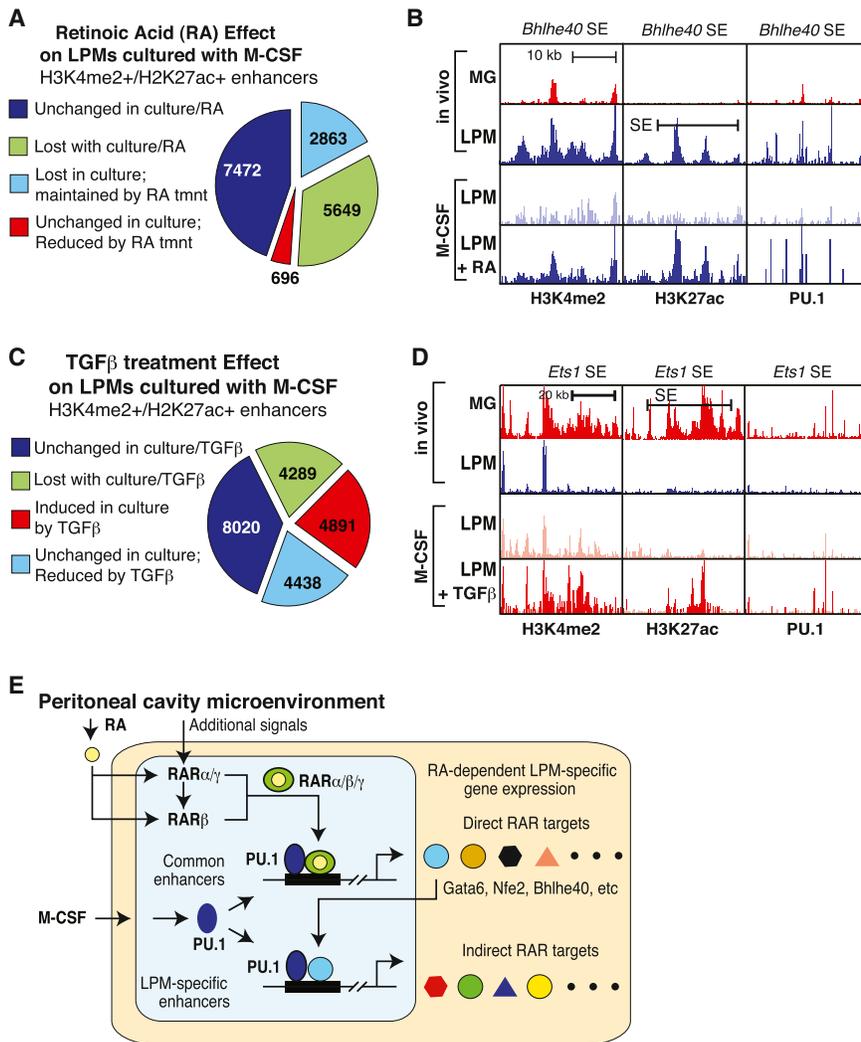
controls the expression of genes responsible for responses to environment-specific signals.

To investigate the extent to which RA and TGFβ influence subset-specific patterns of gene expression, we treated LPMs with RA or TGFβ for 7 days and performed RNA-seq analysis. RA treatment induced expression of nearly half of the LPM-specific genes by more than 2-fold, while inducing about 8% of genes expressed at similar levels in LPMs and MG (Figure 6E). Conversely, nearly 50% of the genes induced more than 2-fold by TGFβ in LPMs in culture are preferentially expressed by MG in vivo, whereas only 4% of the genes expressed at similar levels in LPMs and MG were induced by TGFβ in LPMs (Figure 6F). Thus, RA and TGFβ disproportionately regulate genes that specify LPM and MG-specific phenotypes, respectively.

We next evaluated the expression of TFs that recognize motifs identified as putative binding sites for collaborative partners of PU.1 in LPMs through analysis of strain-specific PU.1 binding. Remarkably, expression of the majority of TFs best matched to motifs identified by strains analysis was environment dependent (Figure 6D). A similar pattern was observed when considering all

ontology and/or survival in vivo (Greter et al., 2012; Wang et al., 2012). This environmental transition resulted in vast changes in gene expression (Table S7). Comparison of the gene expression program of LPMs freshly purified from the peritoneal cavity with LPMs maintained in M-CSF for 7 days is illustrated in Figure 6A. Data points colored in blue represent genes that are expressed more than 16-fold higher in LPMs than MG, indicating that the LPM-specific program of gene expression is preferentially lost in culture. Comparison of the gene expression program of MG freshly isolated from the brain or maintained in culture in the presence of IL-34 for 7 days is illustrated in Figure 6B. Data points colored in red represent genes that are expressed more than 16-fold higher in MG than LPMs, indicating that the MG-specific program of gene expression is preferentially lost in culture. In both LPMs and MG, many genes exhibiting low levels of expression in vivo are markedly upregulated in culture.

In view of recent findings indicating important roles of TGFβ signaling in MG and RA signaling in peritoneal macrophages, we examined the expression of the main receptors for these factors



**Figure 7. Environmental Influence on Enhancer Landscapes in LPMs and Microglia**

(A) Effects of culture environment and RA chronic stimulation on the enhancer landscape of LPMs. (B) UCSC browser images displaying effects of culture environment and RA chronic stimulation on H3K4me2, H3K27ac, and PU.1 binding at the *Bhlhe40* locus in LPMs. (C) Effects of culture environment and chronic stimulation with TGFβ on the enhancer landscape of LPMs. (D) UCSC browser images displaying effects of culture environment and chronic stimulation with TGF-β1 on H3K4me2, H3K27ac, and PU.1 binding at the *Ets1* locus in LPMs. (E) Hierarchical model for mechanisms by which the peritoneal environment induces the enhancer landscape and gene expression signature of LPMs. See Discussion for details. See also Table S7.

the enhancer-like regions (Figure 7A). One-third of these lost enhancer elements were maintained by RA treatment (Figure 7A). Of the 302 LPM-specific SEs identified in Figure 3, 223 (74%) no longer met SE criteria, indicating a disproportionate sensitivity to loss of environmental signals. This pattern is exemplified by the SE associated with *Bhlhe40*, which, in addition to substantial reduction in the histone signature of enhancers, also exhibits reduced PU.1 binding (Figure 7B). Notably, H3K4me2, H3K27ac, and PU.1 binding are largely maintained by RA treatment.

Maintenance of LPMs in M-CSF plus TGFβ resulted in marked changes in the

LPM enhancer landscape in comparison to culture in M-CSF alone, which is consistent with the preferential effects of TGFβ on a MG-specific program of gene expression (Figure 7C). Treatment with TGFβ increased the enhancer signature by more than 2-fold at ~25% of pre-existing enhancers. Conversely, TGFβ reduced enhancer signatures at ~25% enhancer-like elements that were stable upon transfer to culture in M-CSF. Induced enhancers are exemplified by a genomic region in the vicinity of the *Ets1* gene (Figure 7D) that is preferentially expressed in MG and is highly induced by TGFβ. Culture of LPMs in M-CSF results in appearance of PU.1 binding and H3K4me2 modification, with TGFβ treatment leading to substantial increases in H3K27ac and a marked increase in gene expression.

**DISCUSSION**

**Mechanisms Underlying Tissue-Specific Enhancer Selection and Activation**

The present studies provide evidence for a hierarchical model in which the distinct environments of the brain and peritoneal cavity differentially activate a common set of primed enhancers and

members of each TF family capable of recognizing these motifs (Figure S4A). RNA-seq analysis further suggested that several of these factors were inducible by RA. This response was confirmed under M-CSF treatment conditions for *Gata6* and *Rarb*, consistent with previous studies (Okabe and Medzhitov, 2014), as well as for *Bhlhe40* and *Nfe2* (Figures 6G and S4B). However, expression of *Rara*, *Rarg*, and most of the other factors illustrated in Figure 6C was not RA inducible. Thus, the environment modulates in LPMs the expression of collaborative and SDTFs through both RA-dependent and RA-independent mechanisms.

**Hierarchical Effects of Environment on Macrophage Enhancer Landscapes**

To gain insights into mechanisms underlying effects of environment on macrophage gene expression, we performed ChIP-seq analysis for H3K4me2, H3K27ac, and PU.1 in LPMs maintained in M-CSF and the presence or absence of RA or TGFβ for 7 days. Transition of LPMs from the peritoneal cavity to a tissue culture environment containing M-CSF led to a >2-fold reduction in H3K4me2 and/or H3K27ac at approximately half of

their target genes that, in turn, promote the selection and activation of subset-specific enhancer repertoires. The combinatorial activation of both common and subset-specific enhancers enables context-dependent regulation of genes required for specialized functions of MG and RPMs. Aspects of this model as they pertain to the RA-dependent program of gene expression specific to peritoneal macrophages are illustrated in [Figure 7E](#). Common to all macrophage subsets, stimulation of signaling pathways downstream of the M-CSF receptor by M-CSF and/or IL-34, which are present in the environment in a largely tissue-non-specific manner, ensures survival and promotes PU.1 expression ([Sarrazin et al., 2009](#)). PU.1 is a critical LDTFs required for all macrophage subsets that functions to select common and cell-specific enhancers through collaborative interactions with other TFs. These regions of PU.1 binding in turn serve as subset-specific sites of action for various types of SDTFs.

Within the peritoneal cavity, environment-specific signals control the expression and activities of TFs that act upon primed enhancers that are common to multiple macrophage subsets ([Figure 7E](#)). A particularly important signal is omentum-derived RA, which has been shown to be essential for development and function of LPMs through its activation of RAR $\beta$  and induction of *Gata6* ([Okabe and Medzhitov, 2014](#)). We find that all three high-affinity retinoic acid receptor genes (*Rara*, *Rarb*, and *Rarg*) are preferentially expressed in the peritoneal cavity and that this expression requires continual maintenance by the peritoneal cavity environment. However, only *Rarb* expression is preserved by RA treatment in culture, indicating that expression of *Rara* and *Rarg* is under the control of as-yet-unknown factors. We speculate that the expression of RAR $\alpha$  and RAR $\gamma$  is necessary for full induction of RAR $\beta$  expression in response to environmental RA and that this positive feedback loop is important for amplification of the RA signal and activation of direct RA target genes. These findings imply that at least two environmental signals are required for initiating the RA-dependent peritoneal macrophage phenotype, one being RA itself and the second being a signal or signals required for RAR $\alpha$  and RAR $\gamma$  expression.

Activated retinoic acid receptors primarily function as SDTFs that act at a common set of primed enhancers established by PU.1 and other LDTFs that are expressed across macrophage subsets. Importantly, direct RA target genes include *Gata6*, *Bhlhe40*, and *Nfe2*, which were identified as putative interacting partners of PU.1 through analysis of effects of natural genetic variation. We propose that RAR-dependent induction of these factors results in collaborative interactions with PU.1 that drive environment-specific selection of LPM-specific enhancers ([Figure 7E](#)).

Of note, *Gata6*, *Bhlhe40*, and *Nfe2*, as well as all three retinoic acid receptors, reside in or near peritoneal macrophage-specific SEs that are lost when LPMs are removed from the peritoneal cavity. Our findings suggest that an analogous hierarchy operates in MG, driven in part by TGF $\beta$  signaling and SMAD TFs. Although the present studies have focused on PU.1, we expect that additional macrophage LDTFs function in an analogous manner to set up macrophage-specific, PU.1-independent enhancers.

### Use of Natural Genetic Variation to Validate and Discover Collaborative TFs

Here, we demonstrate the use of the natural genetic variation provided by inbred strains of mice as a powerful means to validate and discover collaborative TFs. By measuring strain-specific binding of PU.1 in macrophages derived from genetically diverse strains of mice, we identified motifs for several different classes of TFs in which strain-specific mutations were highly correlated with the loss of binding of PU.1 to nonmutated PU.1 recognition motifs. Interestingly, the expression of a significant fraction of the TFs recognizing these motifs is dependent on environment.

Many of the motifs identified by analysis of strain-specific binding of PU.1 are recognized by TFs that have well-established roles in macrophage biology. Some, such as C/EBP $\alpha$  and C/EBP $\beta$ , are documented to function as factors that enable collaborative binding of PU.1 in macrophages ([Heinz et al., 2010](#)), supporting the validity of the approach. Although the biological role of *Gata6* in the development and function of LPMs is established ([Okabe and Medzhitov, 2014](#); [Rosas et al., 2014](#)), the present studies suggest that a key molecular function of *Gata6* is to collaborate with PU.1, and likely other macrophage LDTFs, to drive the selection of LPM-specific enhancers. *Bhlhe40* and *Nfe2* represent examples of putative collaborative partners of PU.1 that have not as yet been linked to macrophage-specific functions. *Bhlhe40*, also known as *Dec1*, *Stra13*, and *Sharp2*, has previously been shown to be inducible by RA and to act as both as a repressor and activator ([Boudjelal et al., 1997](#); [Ivanova et al., 2004](#)), raising the possibility that it could contribute to selection of LPM-specific enhancers, as well as suppress genes that become active when LPMs are removed from the peritoneal cavity. *Nfe2* is a bZip transcription factor that is broadly expressed in the hematopoietic system and has been established to play important roles in erythropoiesis and megakaryocyte development ([Andrews, 1998](#)). The present findings provide a rationale for further investigation of roles of *Bhlhe40*, *Nfe2*, and other TFs identified as putative collaborative binding partners of PU.1.

The use of natural genetic variation as a strategy for identification of TFs required for enhancer selection can in principle be applied to any cell type in which ChIP-seq can be performed for an index LDTF. In addition, although not a focus of the present studies, the variation in enhancer selection and activity observed in macrophages derived from different inbred strains of mice was associated with strain-specific differences in LPM and MG gene expression. Such changes in gene expression are presumably linked to both molecular phenotypes such as eQTLs and to the marked phenotypic differences exhibited by these mice that are influenced by tissue resident macrophage populations, such as relative susceptibility or resistance to metabolic, cardiovascular, infectious, and neurodegenerative diseases ([Civelek and Lusis, 2014](#); [Threadgill and Churchill, 2012](#)). The principle of collaborative binding, which serves as the basis for the motif discovery method described here, is directly applicable to investigating mechanisms by which noncoding variants may exert phenotypic effects in a cell-type-specific and/or context-dependent manner. In concert, these approaches enable insights into gene-by-environment interactions and the genetic architecture of molecular and complex disease traits.

### Tuning Enhancer Landscapes and Gene Expression to Context-Specific Functions

The present studies reveal that each macrophage subset uniquely possesses a distinct set of active enhancers, including subset-specific SEs, which are associated with strong preferential expression of nearby genes. In LPMs, for example, which populate a very potent immunogenic environment, *Gbp2b* and *Alox15* are associated with SE activity, and we note that the protein products of these genes are critical regulators of immunity, in particular inflammation and tolerance (Pilla et al., 2014; Uderhardt et al., 2012; Yamamoto et al., 2012). In contrast to LPMs, MG reside in the immune-privileged environment of the brain. As with LPMs, however, our observations suggest that MG adopt a unique phenotype that is again strongly contributed by distinct enhancers and SEs to accomplish tissue-specific functions required for brain homeostasis. For example, SEs in MG include genomic loci associated with the *Cx3cr1* and *Gpr56* genes, among others. Interestingly, both genes are highly relevant to brain functions, regulating synaptic pruning and efficient cortical patterning during brain development (Paolicelli et al., 2011; Piao et al., 2004). Together, our studies reveal an intricate relationship between the organization of the genome of tissue macrophage and their surrounding environment.

### Divergent Macrophage Gene Expression in a Common Environment

Distinct macrophage populations can coexist in a similar environment, as illustrated by the copresence of LPMs and SPMs in the peritoneum. Although these cells are highly concordant with respect to gene expression and organization of their enhancer landscapes, consistent with exposure to common tissue-derived signals, strong points of divergence can nonetheless discriminate the two. These observations raise the possibility that differences in origin and ontology play important roles in determining these later-stage differences (Perdiguerro et al., 2014; Schulz et al., 2012). Thus, the impact of developmental history on the regulation of enhancer repertoires and gene expression of different tissue macrophages remains a fundamental open question to be addressed in future studies.

## EXPERIMENTAL PROCEDURES

### Mice

Seven-week-old C57BL/6J, NOD/ShiLtJ, and SPRET/EiJ male mice were purchased from Jackson Labs and used at 8 to 9 weeks of age. All animal procedures were in accordance with University of California, San Diego research guidelines for the care and use of laboratory animals.

### Microglia Isolation

Mice were anaesthetized with CO<sub>2</sub> and quickly perfused intracardially with ice-cold DPBS. Whole brains were removed and gently mechanically homogenized on ice. Cells were fractionated by Percoll gradient centrifugation, and microglia-enriched fractions were further purified by cell sorting according to the scheme described in Figure S1A and Extended Experimental Procedures.

### Peritoneal Macrophage Isolation

Following euthanization, peritoneal cells were collected by lavage of the peritoneum with ice-cold staining buffer. LPM and SPM subsets were purified based on relative expression of MHCII and other markers described in Figure S1B and Extended Experimental Procedures.

### Thioglycollate-Elicited and Bone-Marrow-Derived Macrophages Cultures

TGEMs were harvested by peritoneal lavage with 20 ml ice-cold PBS 4 days after peritoneal injection of 3 ml Thioglycollate broth. Both TGEMs and BMDMs were cultured as described in Heinz et al. (2010). See also Extended Experimental Procedures.

### ChIP-Seq

Macrophages were fixed at room temperature with 1% paraformaldehyde/PBS containing 1 mM sodium butyrate for 10 min and quenched with glycine.  $2.0 \times 10^5$  to  $1.0 \times 10^6$  cells were used for ChIP, and samples were processed as previously described (Heinz et al., 2010), with minor modifications noted in the Extended Experimental Procedures. Sequencing libraries were prepared as previously described (Heinz et al., 2010).

### RNA Isolation

For RNA-seq, TRIzol (Life Technologies) isolated RNA was either PolyA-selected (MicroPoly(A) Purist kit, Ambion) or subjected to RiboZero rRNA removal (Epicenter).

### Quantitative PCR, RNA-Seq Library Preparation, and Sequencing

Libraries for RNA sequencing were generated as previously described (Heinz et al., 2013). See Extended Experimental Procedures for details and qRT-PCR primer sequences.

### Data Analysis

Fastq files from sequencing experiments were mapped to individual genomes for the mouse strain of origin using default parameters for STAR (Dobin et al., 2013) (RNA-seq) and Bowtie2 (Langmead and Salzberg, 2012) (ChIP-seq). NOD/ShiLtJ and SPRET/EiJ custom genomes were generated from invariant positions of the mm10 sequence with alleles replaced by those reported in VCF files from the Mouse Genomes Project (Keane et al. 2011). Mapped data were analyzed with HOMER (Heinz et al., 2010), custom R, and Perl scripts.

### ACCESSION NUMBERS

Raw and processed data are provided in the Gene Expression Omnibus (GEO) under accession number GSE62826.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.023>.

### AUTHOR CONTRIBUTIONS

D.G. and C.K.G. designed the study. D.G., G.J.F., D.Z.E., N.J.S., J.D.S., H.B.C., and H.G. performed experiments. D.G., V.L., C.E.R., F.G., and C.K.G. analyzed and interpreted the data. D.G., V.L., C.E.R., and C.K.G. wrote the manuscript.

### ACKNOWLEDGMENTS

The authors would like to thank Lynn Bautista, Andrea Crotti, and Renee F. Glass for valuable assistance with preparation of the manuscript and Sven Heinz for assistance with the ChIP-seq library preparation protocol. The authors would also like to thank the UCSD Human Embryonic Stem Cell Core Facility for assistance with cell sorting. These studies were primarily supported by NIH grants DK091183, CA17390, and DK063491 and the San Diego Center for Systems Biology (GM085764). D.G. was supported by a Canadian Institutes of Health Research Fellowship. C.E.R. was supported by the American Heart Association (12POST11760017) and the NIH Pathway to Independence Award (1K99HL12348). D.Z.E. was supported by an American Heart Association predoctoral fellowship (12PRE11610007). F.G. was supported by a Wellcome Trust Senior Investigator award (WT101853MA).

Received: November 2, 2014  
 Revised: November 16, 2014  
 Accepted: November 17, 2014  
 Published: December 4, 2014

## REFERENCES

- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461.
- Andrews, N.C. (1998). The NF-E2 transcription factor. *Int. J. Biochem. Cell Biol.* 30, 429–432.
- 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.
- Boudjelal, M., Taneja, R., Matsubara, S., Bouillet, P., Dolle, P., and Chambon, P. (1997). Overexpression of Stra13, a novel retinoic acid-inducible gene of the basic helix-loop-helix family, inhibits mesodermal and promotes neuronal differentiation of P19 cells. *Genes Dev.* 11, 2052–2065.
- Brown, J.D., Lin, C.Y., Duan, Q., Griffin, G., Federation, A.J., Paranal, R.M., Bair, S., Newton, G., Lichtman, A.H., Kung, A.L., et al. (2014). NF- $\kappa$ B Directs Dynamic Super Enhancer Formation in Inflammation and Atherogenesis. *Mol. Cell* 56, 219–231.
- Butovsky, O., Jedrychowski, M.P., Moore, C.S., Cialic, R., Lanser, A.J., Gabriely, G., Koeglsperger, T., Dake, B., Wu, P.M., Doykan, C.E., et al. (2014). Identification of a unique TGF- $\beta$ -dependent molecular and functional signature in microglia. *Nat. Neurosci.* 17, 131–143.
- Cardona, A.E., Pioro, E.P., Sasse, M.E., Kostenko, V., Cardona, S.M., Dijkstra, I.M., Huang, D., Kidd, G., Dombrowski, S., Dutta, R., et al. (2006). Control of microglial neurotoxicity by the fractalkine receptor. *Nat. Neurosci.* 9, 917–924.
- Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* 15, 34–48.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931–21936.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* 47, 810–822.
- Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubczik, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Gordonov, S., et al.; Immunological Genome Consortium (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* 13, 1118–1128.
- Gautier, E.L., Ivanov, S., Williams, J.W., Huang, S.C., Marcelin, G., Fairfax, K., Wang, P.L., Francis, J.S., Leone, P., Wilson, D.B., et al. (2014). Gata6 regulates aspartoacylase expression in resident peritoneal macrophages and controls their survival. *J. Exp. Med.* 211, 1525–1531.
- Geissmann, F., Manz, M.G., Jung, S., Sieweke, M.H., Merad, M., and Ley, K. (2010). Development of monocytes, macrophages, and dendritic cells. *Science* 327, 656–661.
- 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.
- Ghosn, E.E., Cassado, A.A., Govoni, G.R., Fukuhara, T., Yang, Y., Monack, D.M., Bortoluci, K.R., Almeida, S.R., Herzenberg, L.A., and Herzenberg, L.A. (2010). Two physically, functionally, and developmentally distinct peritoneal macrophage subsets. *Proc. Natl. Acad. Sci. USA* 107, 2568–2573.
- Gordon, S., Plueddemann, A., and Martinez Estrada, F. (2014). Macrophage heterogeneity in tissues: phenotypic diversity and functions. *Immunol. Rev.* 262, 36–55.
- Greter, M., Lelios, I., Pelczar, P., Hoeffel, G., Price, J., Leboeuf, M., Kündig, T.M., Frei, K., Ginhoux, F., Merad, M., and Becher, B. (2012). Stroma-derived interleukin-34 controls the development and maintenance of langerhans cells and the maintenance of microglia. *Immunity* 37, 1050–1060.
- He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., et al. (2010). Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* 42, 343–347.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
- Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orzoco, L.D., and Glass, C.K. (2013). Effect of natural genetic variation on enhancer selection and function. *Nature* 503, 487–492.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947.
- Ivanova, A.V., Ivanov, S.V., Zhang, X., Ivanov, V.N., Timofeeva, O.A., and Lerman, M.I. (2004). STRA13 interacts with STAT3 and modulates transcription of STAT3-dependent targets. *J. Mol. Biol.* 340, 641–653.
- Jung, S., Aliberti, J., Graemmel, P., Sunshine, M.J., Kreutzberg, G.W., Sher, A., and Littman, D.R. (2000). Analysis of fractalkine receptor CX(3)CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol. Cell Biol.* 20, 4106–4114.
- 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., and Glass, C.K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* 51, 310–325.
- Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289–294.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr. Bio.* 20, R754–R763.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334.
- Makwana, M., Jones, L.L., Cuthill, D., Heuer, H., Bohatschek, M., Hristova, M., Friedrichsen, S., Ormsby, I., Bueringer, D., Koppius, A., et al. (2007). Endogenous transforming growth factor beta 1 suppresses inflammation and promotes survival in adult CNS. *J. Neurosci.* 27, 11201–11213.
- Mullen, A.C., Orlando, D.A., Newman, J.J., Lovén, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF- $\beta$  signaling. *Cell* 147, 565–576.
- Okabe, Y., and Medzhitov, R. (2014). Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell* 157, 832–844.
- 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.
- Paolicelli, R.C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., Giustetto, M., Ferreira, T.A., Guiducci, E., Dumas, L., et al. (2011). Synaptic pruning by microglia is necessary for normal brain development. *Science* 333, 1456–1458.

- Perdiguerro, E.G., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M.F., Geissmann, F., et al. (2014). Tissue-resident macrophages originate from yolk sac-derived erythro-myeloid progenitors. *Nature*. Published online December 3, 2014. <http://dx.doi.org/10.1038/nature13989>.
- Piao, X., Hill, R.S., Bodell, A., Chang, B.S., Basel-Vanagaite, L., Straussberg, R., Dobyns, W.B., Qasrawi, B., Winter, R.M., Innes, A.M., et al. (2004). G protein-coupled receptor-dependent development of human frontal cortex. *Science* 303, 2033–2036.
- Pilla, D.M., Hagar, J.A., Haldar, A.K., Mason, A.K., Degrandi, D., Pfeffer, K., Ernst, R.K., Yamamoto, M., Miao, E.A., and Coers, J. (2014). Guanylate binding proteins promote caspase-11-dependent pyroptosis in response to cytoplasmic LPS. *Proc. Natl. Acad. Sci. USA* 111, 6046–6051.
- Rosas, M., Davies, L.C., Giles, P.J., Liao, C.T., Kharfan, B., Stone, T.C., O'Donnell, V.B., Fraser, D.J., Jones, S.A., and Taylor, P.R. (2014). The transcription factor Gata6 links tissue macrophage phenotype and proliferative renewal. *Science* 344, 645–648.
- Samstein, R.M., Arvey, A., Josefowicz, S.Z., Peng, X., Reynolds, A., Sandstrom, R., Neph, S., Sabo, P., Kim, J.M., Liao, W., et al. (2012). Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* 151, 153–166.
- Sarrazin, S., Mossadegh-Keller, N., Fukao, T., Aziz, A., Mourcin, F., Vanhille, L., Kelly Modis, L., Kastner, P., Chan, S., Duprez, E., et al. (2009). MafB restricts M-CSF-dependent myeloid commitment divisions of hematopoietic stem cells. *Cell* 138, 300–313.
- Schulz, C., Gomez Perdiguerro, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S.E., Pollard, J.W., et al. (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* 336, 86–90.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286.
- Steimle, V., Otten, L.A., Zufferey, M., and Mach, B. (1993). Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). *Cell* 75, 135–146.
- Threadgill, D.W., and Churchill, G.A. (2012). Ten years of the collaborative cross. *G3 (Bethesda)* 2, 153–156.
- Uderhardt, S., Herrmann, M., Oskolkova, O.V., Aschermann, S., Bicker, W., Ipseiz, N., Sarter, K., Frey, B., Rothe, T., Voll, R., et al. (2012). 12/15-lipoxygenase orchestrates the clearance of apoptotic cells and maintains immunologic tolerance. *Immunity* 36, 834–846.
- Wang, Y., Szretter, K.J., Vermi, W., Gilfillan, S., Rossini, C., Cella, M., Barrow, A.D., Diamond, M.S., and Colonna, M. (2012). IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nat. Immunol.* 13, 753–760.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.
- Witmer-Pack, M.D., Hughes, D.A., Schuler, G., Lawson, L., McWilliam, A., Inaba, K., Steinman, R.M., and Gordon, S. (1993). Identification of macrophages and dendritic cells in the osteopetrotic (op/op) mouse. *J. Cell Sci.* 104, 1021–1029.
- Wynn, T.A., Chawla, A., and Pollard, J.W. (2013). Macrophage biology in development, homeostasis and disease. *Nature* 496, 445–455.
- Yamamoto, M., Okuyama, M., Ma, J.S., Kimura, T., Kamiyama, N., Saiga, H., Ohshima, J., Sasai, M., Kayama, H., Okamoto, T., et al. (2012). A cluster of interferon- $\gamma$ -inducible p65 GTPases plays a critical role in host defense against *Toxoplasma gondii*. *Immunity* 37, 302–313.

## Chapter 2

# Tissue damage drives co-localization of NF- $\kappa$ B, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages

Dawn Z. Eichenfield\*, Ty D. Troutman\*, **Verena M. Link\***, Michael T. Lam, Han Cho, David Gosselin, Nathanael J. Spann, Hanna P. Lesch, Jenhan Tao, Jun Muto, Richard L. Gallo, Ronald M. Evans, Christopher K. Glass

*\* authors contribute equally*

*eLife* (2016) 5:e13024



# Tissue damage drives co-localization of NF- $\kappa$ B, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages

Dawn Z Eichenfield<sup>1,2†</sup>, Ty Dale Troutman<sup>1†</sup>, Verena M Link<sup>1,3†</sup>, Michael T Lam<sup>1,2,4</sup>, Han Cho<sup>5</sup>, David Gosselin<sup>1</sup>, Nathanael J Spann<sup>1</sup>, Hanna P Lesch<sup>1</sup>, Jenhan Tao<sup>1</sup>, Jun Muto<sup>6</sup>, Richard L Gallo<sup>6</sup>, Ronald M Evans<sup>5</sup>, Christopher K Glass<sup>1,4\*</sup>

<sup>1</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, San Diego, United States; <sup>2</sup>Biomedical Sciences Graduate Program, University of California, San Diego, San Diego, United States; <sup>3</sup>Department II, Faculty of Biology, Ludwig-Maximilian Universität München, Planegg-Martinsried, Germany; <sup>4</sup>Department of Medicine, University of California, San Diego, San Diego, United States; <sup>5</sup>Salk Institute for Biological Sciences, La Jolla, United States; <sup>6</sup>Department of Dermatology, University of California, San Diego, San Diego, United States

**Abstract** Although macrophages can be polarized to distinct phenotypes in vitro with individual ligands, in vivo they encounter multiple signals that control their varied functions in homeostasis, immunity, and disease. Here, we identify roles of Rev-erb nuclear receptors in regulating responses of mouse macrophages to complex tissue damage signals and wound repair. Rather than reinforcing a specific program of macrophage polarization, Rev-erbs repress subsets of genes that are activated by TLR ligands, IL4, TGF $\beta$ , and damage-associated molecular patterns (DAMPs). Unexpectedly, a complex damage signal promotes co-localization of NF- $\kappa$ B, Smad3, and Nrf2 at Rev-erb-sensitive enhancers and drives expression of genes characteristic of multiple polarization states in the same cells. Rev-erb-sensitive enhancers thereby integrate multiple damage-activated signaling pathways to promote a wound repair phenotype.

DOI: [10.7554/eLife.13024.001](https://doi.org/10.7554/eLife.13024.001)

\*For correspondence: [cglass@ucsd.edu](mailto:cglass@ucsd.edu)

†These authors contributed equally to this work

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 26

**Received:** 13 November 2015  
**Accepted:** 30 June 2016

**Reviewing editor:** Satyajit Rath, National Institute of Immunology, India

© Copyright Eichenfield et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Macrophages reside in all tissues of the body and play key roles in homeostasis, immunity, and disease. As immune cells, macrophages serve as sentinels of infection and injury and are active participants in both innate and adaptive immune responses. Detection of pathogens and tissue damage is mediated by a diverse array of pattern recognition receptors for pathogen associated molecular patterns (PAMPs) and damage associated molecular patterns (DAMPs), exemplified by the toll-like receptors (TLRs). Ligation of TLRs initiates profound changes in gene expression that include induction of chemokines, cytokines, anti-microbial peptides, and other factors that contribute to the innate immune response and influence adaptive immunity (Ostuni et al., 2013; Lawrence and Natoli, 2011). This response has been extensively characterized in vitro by treating cultured macrophages with specific TLR ligands such as bacterial lipopolysaccharide (LPS), a potent activator of TLR4 (Kaikkonen et al., 2013; Escoubet-Lozach et al., 2011; Raetz et al., 2006). TLR4 ligation regulates gene expression through signal transduction pathways culminating in the activation of latent signal-dependent transcription factors, which include members of the nuclear factor kappa-light-

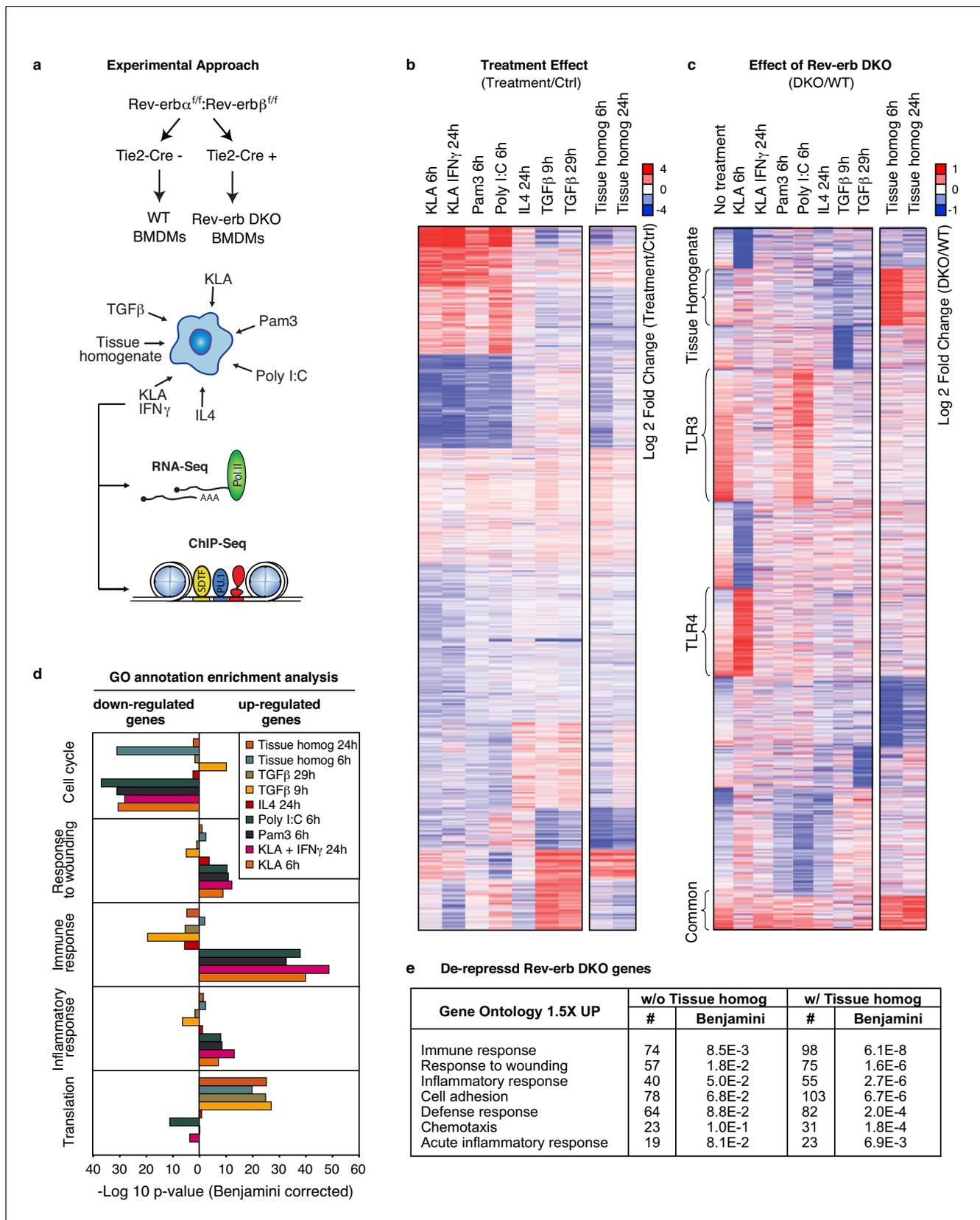
chain-enhancer of activated B cells (NF- $\kappa$ B), interferon regulatory factor (IRF), and activator protein 1 (AP-1) families (*Medzhitov and Horng, 2009*). In macrophages, these factors are primarily directed to macrophage-specific enhancers that are selected by macrophage lineage determining transcription factors, PU.1 and CCAAT-enhancer-binding proteins (C/EBPs) (*Heinz et al., 2010*). The macrophage activation phenotype resulting from selective treatment with LPS, or in some cases a combination of LPS and interferon  $\gamma$  (IFN $\gamma$ ), is referred to as M(LPS) or M(LPS+IFN $\gamma$ ) activation (*Murray et al., 2014*), and is considered vital for the host response to bacterial or viral infection.

Macrophages also play important roles in regulating the resolution phase of inflammation as well as the repair of tissue damage. These functions are controlled by complex microenvironmental pathways that include reductionist signals such as transforming growth factor  $\beta$  (TGF $\beta$ ) and interleukin 4 (IL4). TGF $\beta$  is generally considered to be an inducer of a 'de-activated' macrophage or M(TGF $\beta$ ) phenotype, although it also acts as a potent chemo-attractant for monocytes and can potentiate their transition into activated cells (*Li et al., 2006*). Macrophages respond to TGF $\beta$  in both an autocrine and paracrine manner. For example, phagocytosis of apoptotic cells results in increased macrophage-mediated secretion of TGF $\beta$  and subsequent inhibition of inflammatory cytokine production (*Li et al., 2006*). In addition to dampening inflammatory responses, secreted TGF $\beta$  plays key roles in accelerating wound healing and fibrosis (*Schuppan and Kim, 2013*). At the transcriptional level, TGF $\beta$  signal transduction pathways function primarily in a Mothers against decapentaplegic homolog (SMAD)-dependent manner through Smad2-, Smad3-, and Smad4-mediated activation, as well as Smad7-mediated inhibition (*Massagué, 2012*). Like other signal-dependent transcription factors, ligation of TGF $\beta$  receptors causes the localization of Smad3 to genomic loci containing lineage-determining transcription factors (*Mullen et al., 2011*).

Regulation of macrophage gene expression by IL4 plays roles in containment of parasitic infections and in homeostatic functions of adipose tissue. IL4 acts through the IL4 receptor to activate signal transducer and activator of transcription 6 (Stat6) (*Lefterova et al., 2010*), which positively regulates gene expression upon binding to recognition elements in promoters and enhancers of target genes (*Li et al., 2006*). IL4 signaling regulates genes that control tissue remodeling, phagocytosis, scavenging, and the arginase pathway. The macrophage activation phenotype resulting from selective treatment with IL4 is referred to as M(IL4) and is considered vital for the role of macrophages in wound repair (*Van Dyken and Locksley, 2013*).

While M(LPS) or M(LPS+IFN $\gamma$ ), M(TGF $\beta$ ), and M(IL4) macrophage phenotypes are clearly distinct *in vitro*, they result from selective activation of specific signaling pathways by strongly polarizing ligands. *In vivo*, macrophages encounter diverse combinations of signals that can change over time in response to physiological or pathological processes such as tissue injury. Recent studies show that these combinations of signals can influence the transcriptional landscape of macrophages in an input-specific fashion (*Lavin et al., 2014; Ginhoux et al., 2015; Gosselin et al., 2014*). However, how complex signals are integrated at the level of transcription and how reductionist stimuli (LPS, TGF $\beta$ , and IL4) can be used as a framework to predict how combinations of transcriptional regulators coordinate immune and tissue repair activities in complex tissue microenvironments remain largely unknown.

The Rev-erb nuclear receptor family consists of two members, Rev-erb $\alpha$  (also known as nuclear receptor subfamily 1, group D, member 1, NR1D1) and Rev-erb $\beta$  (also known as nuclear receptor subfamily 1, group D, member 2, NR1D2) (Rev-erbs), that regulate the expression of genes involved in the control of circadian rhythm (*Preitner et al., 2002; Liu et al., 2008; Cho et al., 2012*), metabolism (*Raspé et al., 2002; Le Martelot et al., 2009; Feng et al., 2011; Solt et al., 2012*), and inflammation (*Fontaine et al., 2008; Gibbs et al., 2012*). Rev-erbs mediate transcriptional repression through recruitment of the nuclear co-repressor (NCoR) and histone de-acetylase 3 (HDAC3) complex (*Yin and Lazar, 2005*). Rev-erbs lack the carboxy-terminal (AF2) transactivation domain, which is required for recruitment of co-activators (*Durand et al., 1994*). Genome-wide location analysis of Rev-erb $\alpha$  and Rev-erb $\beta$  in macrophages revealed thousands of binding sites, the vast majority of which resided at macrophage-specific enhancer-like regions of the genome established by PU.1 and other macrophage lineage determining factors (*Lam et al., 2013*). Gain and loss of function experiments indicated that Rev-erbs function to suppress the activities of these enhancers by repressing enhancer-directed transcription. While these studies provided insights into the functional significance of enhancer transcription, the biological consequences of the actions of Rev-erbs at these distal regulatory elements were not explored.



**Figure 1.** Overall impact of Rev-erb DKO on signal-dependent gene expression. (a) Schematic illustrating the experimental approach used in defining the global transcriptional program in WT and Rev-erb DKO bone marrow derived macrophages (BMDMs). (b) Heatmap showing genes captured by Figure 1 continued on next page

Figure 1 continued

RNA-Seq associated with Rev-erb control after treatment with the indicated ligands compared to the basal state. Genes shown are those more than 1.5-fold differentially expressed in Rev-erb DKO macrophages compared to WT. Data is represented as log<sub>2</sub> fold change between the basal state and treatment (untreated for 6 hr was used for comparison to KLA, Pam3, Poly I:C, TGFβ 9 hr, and tissue homogenate 6 hr; untreated for 24 hr was used for comparison to KLA + IFNγ, IL4, TGFβ 29 hr, and tissue homogenate 24 hr). Genes were clustered using k-means clustering (k = 10). For untreated samples, N = 4, for samples treated with Pam3, Poly I:C, KLA or KLA + IFNγ, tissue homogenate or TGFβ, N = 3, and for samples treated with IL4, N = 2. The data for this heat map is accessible in **Figure 1—source data 1**. (c) Heatmap showing genes captured by RNA-Seq as differentially expressed 1.5-fold in the Rev-erb DKO macrophage compared to WT as indicated. Data is represented as log<sub>2</sub> fold change between DKO and WT. Genes were clustered using k-means clustering (k = 10). For untreated samples, N = 4, for samples treated with Pam3, Poly I:C, KLA or KLA + IFNγ, tissue homogenate or TGFβ, N = 3, and for samples treated with IL4, N = 2. The data for this heat map is accessible in **Figure 1—source data 2**. (d) Gene ontology analysis using David (**Huang et al., 2009a, 2009b**) of genes shown in panel b. (e) Gene ontology analysis using David (**Huang et al., 2009a, 2009b**) of genes demonstrating de-repressed expression in Rev-erb DKO macrophages by more than 1.5-fold in all of the conditions combined (w/o Tissue homog considers de-repressed genes in columns 1–8 of panel c (N = 2315), while w/ Tissue homog considers de-repressed genes in all columns of panel c (N=2614)).

DOI: [10.7554/eLife.13024.002](https://doi.org/10.7554/eLife.13024.002)

The following source data and figure supplement are available for figure 1:

**Source data 1.** Source data for **Figure 1b** where each value represents the average normalized log<sub>2</sub> fold change between the basal state and treatment state per column.

DOI: [10.7554/eLife.13024.003](https://doi.org/10.7554/eLife.13024.003)

**Figure supplement 1.** .Rev-erb deletion efficiency.

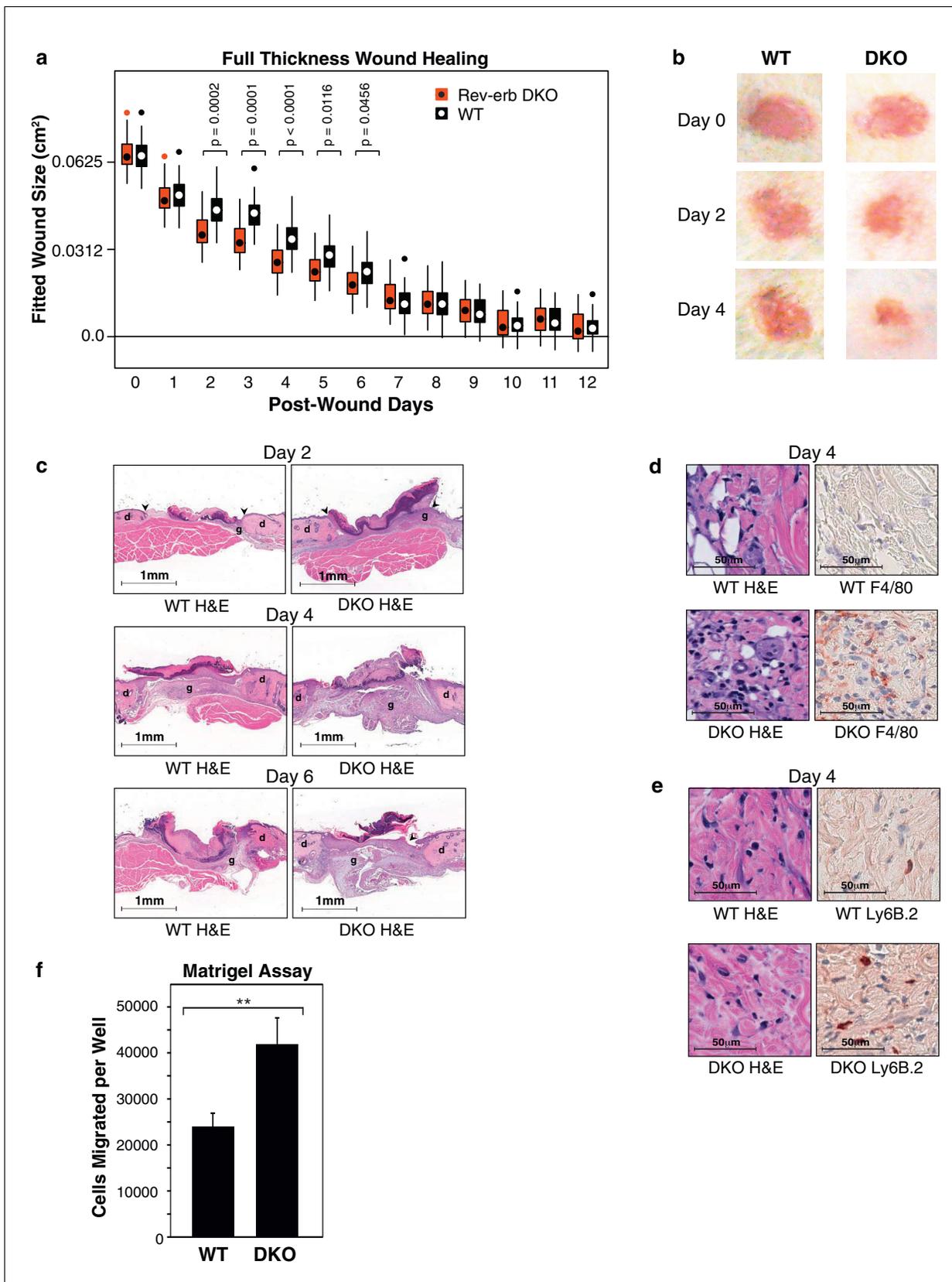
DOI: [10.7554/eLife.13024.004](https://doi.org/10.7554/eLife.13024.004)

Here, we provide evidence that Rev-erbs repress the transcription and function of signal-dependent enhancers that are targets of TLR, IL4, TGFβ, and DAMP signaling. Rather than exerting a pattern of repression that reinforces a particular polarization phenotype, Rev-erbs regulate subsets of signal responsive genes that span those associated with M(LPS) or M(LPS+IFNγ), M(TGFβ), and M(IL4) phenotypes, enriching for functions associated with wound repair. Consistent with these in vitro observations, deletion of Rev-erbs from the hematopoietic lineage in vivo results in accelerated wound repair. Unexpectedly, we found that a complex tissue injury signal directs genomic binding patterns for NF-κB p65 (p65), FBJ murine osteosarcoma viral oncogene homolog (Fos – a member of the activator protein 1, or AP-1, family), and Smad3 that differ substantially from those observed following selective treatments with a TLR4 agonist or TGFβ. In addition, by analyzing changes in enhancer signatures, we identified Nrf2 as an additional mediator of the transcriptional response to the tissue injury signal. While these transcription factors exhibit relatively little co-localization in response to single polarizing ligands, we observe substantial co-localization and enhancer activation in response to the complex tissue injury signal, resulting in transcriptional outcomes that are qualitatively different than the sum of single polarizing signals. These observations provide insights into how combinations of signals are integrated at a transcriptional level to result in context-specific patterns of gene expression.

## Results

### Rev-erb transcriptional activity varies according to polarizing signal

Our previous findings that Rev-erbs regulate transcription from signal-dependent enhancers (**Lam et al., 2013**) led us to investigate possible biological roles of Rev-erbs in influencing macrophage phenotypes (**Figure 1a**). To study the phenotypic contribution of Rev-erbs to signal-dependent gene expression in macrophages, we performed RNA-Sequencing (RNA-Seq) of poly(A) mRNA isolated from wild-type macrophages and those deficient for both Rev-erbα and Rev-erbβ (**Figure 1a**). Rev-erb double knockout (DKO) macrophages were generated from bone marrow differentiation of Tie2-Cre Rev-erbα<sup>fllox/fllox</sup> Rev-erbβ<sup>fllox/fllox</sup> (Rev-erb DKO) mice and compared to control macrophages derived from Cre-negative littermates (WT). Deletion of Rev-erbβ exons to generate a non-functional Rev-erbβ mRNA results in marked de-repression of Rev-erbα expression and increased expression of a DNA binding domain deleted form of Rev-erbα mRNA (**Sud et al., 2007**) (**Figure 1—figure supplement 1a–b**). Similar effects can be seen following deletion of Rev-erbα exons (corresponding to the DNA-binding domain) with respect to Rev-erbβ de-repression



**Figure 2.** Rev-erb DKO bone marrow transplanted animals display enhanced wound closure in a full thickness wound healing model. (a) Wound size ( $\text{cm}^2$ ) as fitted from a linear mixed effects model. Boxes denote the interquartile range and the median, whiskers denote the minimum and maximum. Figure 2 continued on next page

## Figure 2 continued

values excluding outliers, and dots outside of the whiskers denote outlier observations. Data are pooled from three independent experiments as described in more detail in the Materials and methods. The p-values shown reflect comparisons with a p-value less than 0.05, as determined by the linear mixed effects model. (b) Macroscopic digital photographs of wound closure in WT and Rev-erb DKO bone marrow transplanted animals. (c) Histological images of wound healing in WT and Rev-erb DKO bone marrow transplanted animals taken at 2.5x magnification after 2, 4, and 6 days. Arrowheads show differential re-epithelialization between WT and Rev-erb DKO bone marrow transplanted animals. Abbreviations: g=granulation tissue, d=dermis. Images representative of two independent animals. (d) Day 4 hematoxylin and eosin (H&E), as well as F4/80 stained histological images taken at 20x magnification. Images representative of two independent animals. (e) Day 4 hematoxylin and eosin (H&E), as well as Ly6B.2 stained histological images taken at 20x magnification. Images representative of two independent animals. (f) Migration of WT and Rev-erb DKO macrophages through matrigel extracellular matrix for 24 hr (\*\*p-value <0.01 two-tailed test, Data represent mean + SD from one of three experiments using 8 wells with cells pooled from 3 independent mice).

DOI: [10.7554/eLife.13024.005](https://doi.org/10.7554/eLife.13024.005)

The following figure supplement is available for figure 2:

**Figure supplement 1.** . Engraftment efficiency and quantification of circulating blood cells in WT and DKO chimeras.

DOI: [10.7554/eLife.13024.006](https://doi.org/10.7554/eLife.13024.006)

(**Figure 1—figure supplement 1a–b**). Reduction of targeted Rev-erb exonic mRNA averaged 90% for Rev-erb $\alpha$  and 80% for Rev-erb $\beta$  (**Figure 1—figure supplement 1c**).

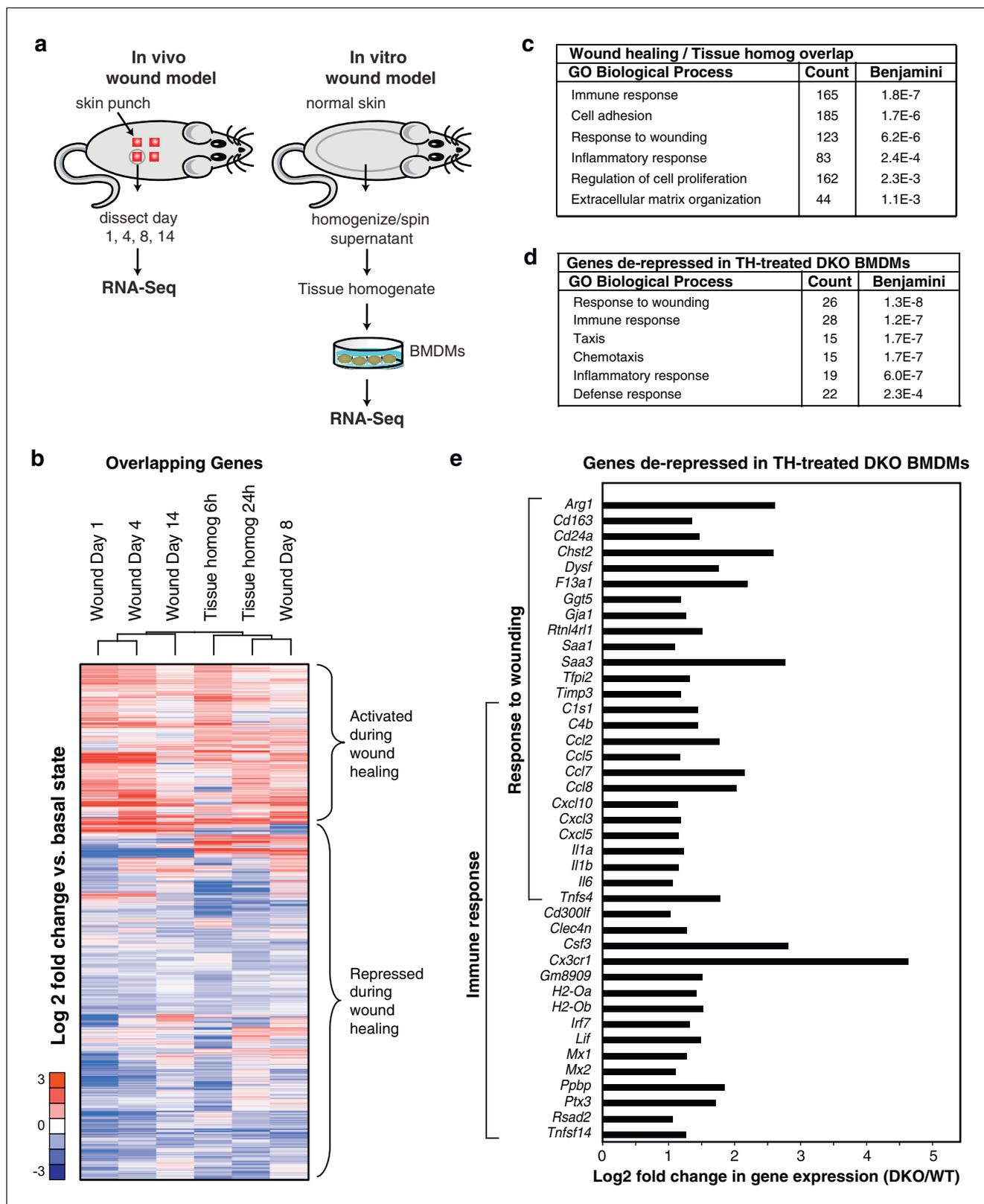
Activation of TLR3 with a synthetic double-stranded RNA analog, polyinosinic-polycytidylic acid (Poly I:C), TLR4 with Kdo2-lipid A (KLA), TLR1/2 with a synthetic triacylated lipopeptide, Pam3CSK4 (Pam3), and co-activation with KLA and IFN $\gamma$  induced characteristic pro-inflammatory gene signatures (**Figure 1b and d**) in WT macrophages. In contrast, IL4 or TGF $\beta$  stimulation of macrophages resulted in the expected alternatively activated and de-activated gene profiles, respectively (**Figure 1b and d**).

Comparing the gene expression signature from WT and Rev-erb DKO macrophages, for the majority of genes, the magnitude of differential expression between WT and Rev-erb DKO macrophages varied depending on the polarization state (**Figure 1c**), in some cases only being observed under basal conditions, and in other cases only observed in response to a particular stimulus. These results suggest that the magnitude of differential expression in WT compared to Rev-erb DKO macrophages is highly dependent on polarization state.

### Rev-erb deficient animals display enhanced wound healing

Gene ontology analysis of mRNAs exhibiting differential expression (>1.5-fold de-repressed in DKO macrophages) in at least one of the single polarizing conditions revealed significant enrichment for genes involved in the response to wounding (**Figure 1e**). Notably, genetic loss of *Cx3cr1* and *Arg1* has been shown to hinder efficient wound healing in mice (**Campbell et al., 2013; Ishida et al., 2008**), suggesting that mice lacking Rev-erbs in cells of hematopoietic origin might exhibit more rapid wound healing. To test this hypothesis, we utilized a full thickness wound healing model (**Figure 2a**) in mice after bone marrow reconstitution with either WT or Rev-erb DKO bone marrow (**Figure 2—figure supplement 1a**). Bone marrow reconstitution efficiency exceeded 94% (**Figure 2—figure supplement 1b**). We found from three independent experiments that Rev-erb deficiency in bone marrow derived hematopoietic cells resulted in accelerated wound closure (**Figure 2a–b**). This was especially apparent on days 2–6 post-injury (**Figure 2a**), consistent with Rev-erb deficiency resulting in a faster response during the immune phase of wound healing.

Wounds from the Rev-erb DKO chimeric mice displayed greater immune cell infiltration and faster wound healing progression, characterized by enhanced re-epithelialization and increased granulation tissue development (**Figure 2c**), characteristics correlated with an accelerated immune response during wound healing. In addition, Rev-erb DKO bone marrow transplanted mice displayed more macrophages at the wound site on day 4 post-injury (**Figure 2d**), while neutrophil persistence at the wound site remained similar between WT and Rev-erb DKO transplanted mice (**Figure 2e**). Moreover, matrigel migration assays show increased extravasation of Rev-erb DKO macrophages when compared to their WT counterparts (**Figure 2f**). Flow cytometry analysis of circulating blood leukocytes from WT and Rev-erb DKO bone marrow transplanted animals (**Figure 2—figure supplement 1c–d**) showed no differences in the populations of Ly6C<sup>low</sup>/Ly6C<sup>high</sup> circulating monocytes. These experiments suggest that the increased migration of macrophages into wounds may be cell autonomous changes in transcriptional output.



**Figure 3.** Rev-erb DKO macrophages display increased inflammatory responses to damaged tissue. (a) Schematic illustrating the experimental approach used comparing the transcriptional profile of in vivo wounds on days 1, 4, 8, or 14 post-wounding with macrophages treated in vitro with tissue homogenate after 6 or 24 hr. (b) Heatmap showing genes differentially expressed both in the in vivo mouse wound and in macrophages after in vitro treatment with tissue homogenate. (c) GO enrichment analysis of genes overlapping between wound healing and tissue homogenate overlap. (d) GO enrichment analysis of genes de-repressed in TH-treated DKO BMDMs. (e) Horizontal bar chart showing Log2 fold change in gene expression (DKO/WT) for genes de-repressed in TH-treated DKO BMDMs, categorized by Response to wounding and Immune response.

Figure 3 continued

vitro stimulation with tissue homogenate. Mouse wound genes from Days 1, 4, 8, or 14 post-injury and macrophage tissue homogenate genes at 6 or 24 hr post-stimulation were compared to uninjured in vivo skin or unstimulated controls, respectively. Differentially expressed genes were those induced or repressed more than 1.5-fold compared to baseline. Genes were clustered using k-means clustering ( $k = 10$ ). For unstimulated macrophages for 6 or 24 hr,  $N = 2$ , wound samples from Day 1, 8, or 14,  $N = 2$ , macrophages stimulated with tissue homogenate for 6 or 24 hr,  $N = 3$ , and wound samples from Day 0, or 4,  $N = 4$ . (c) Summary of gene ontology analysis using DAVID (Huang et al., 2009a, 2009b) of overlapping wound healing and homogenate genes shown in b ( $N = 5590$ ). **Figure 3—source data 1.** (d) Summary of gene ontology analysis using DAVID (Huang et al., 2009a, 2009b) of genes de-repressed more than two-fold in Rev-erb DKO macrophages treated with tissue homogenate in comparison to WT macrophages (maximum de-repression after tissue homogenate treatment for 6 or 24 hr,  $N = 282$ ). (e) Bar graphs depicting representative genes de-repressed more than two-fold (in log<sub>2</sub> scale) in Rev-erb DKO macrophages after tissue injury (maximum de-repression after tissue homogenate treatment for 6 or 24 hr). Genes correspond to those associated with response to wounding and immune response categories in panel d.  $N$  as described in 3b.

DOI: [10.7554/eLife.13024.007](https://doi.org/10.7554/eLife.13024.007)

The following source data is available for figure 3:

**Source data 1.** Source data for **Figure 3b** where each value represents the average normalized log<sub>2</sub> fold change between the basal state and treatment state per column.

DOI: [10.7554/eLife.13024.008](https://doi.org/10.7554/eLife.13024.008)

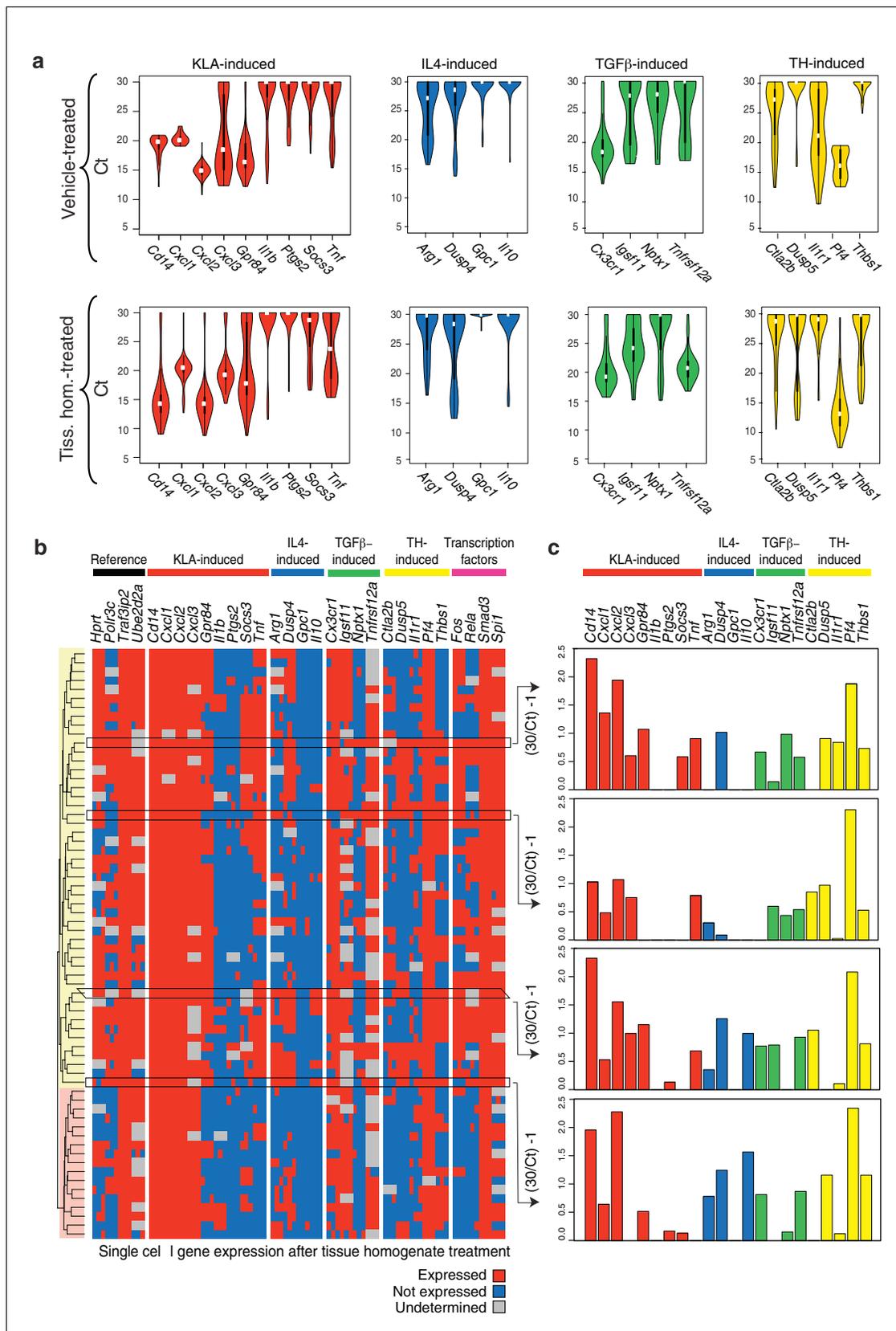
### Rev-erbs integrate macrophage responses to a complex wound signal

Classically, tissue injury of the skin, muscle, or organ systems induces an initial local inflammatory response, which is followed by subsequent regenerative processes involving macrophages and other immune cells, as well as mesenchymal stem cells (Novak and Koh, 2013). To devise an in vitro model of the acute in vivo response to wounding, we prepared a supernatant from homogenized skin (Figure 3a). This tissue homogenate provides a complex signal derived from components of disrupted cells (damage associated molecular patterns; DAMPs), the skin microbiome (microbial associated molecular patterns; MAMPs), and factors residing in the extracellular matrix (e.g., TGF $\beta$ ). Tissue homogenate was used to stimulate WT and Rev-erb DKO macrophages for 6 and 24 hr, followed by RNA-Seq analysis. The gene expression signature of tissue homogenate-stimulated macrophages showed both similarities and differences when compared to the responses observed after treatment with TLR agonists, IL4, or TGF $\beta$  (Figure 1b).

In parallel, we performed temporal transcriptomic analysis of biopsied wounds during wound healing and compared them to unwounded skin (Figure 3a). Although myeloid cells represent only a small fraction of the total cells analyzed in the wound biopsy, 5590 genes exhibited concordant changes in expression with those observed following stimulation of macrophages with tissue homogenate (Figure 3b). Gene ontology analysis of this set of genes indicated significant enrichment for biological process terms related to the response to wounding, immune response, and cell adhesion (Figure 3c). Response to wounding was the most highly enriched gene ontology term associated with genes de-repressed greater than two-fold in Rev-erb DKO tissue homogenate treated macrophages (282) followed by immune response and taxis (Figure 3d). De-repressed genes in Rev-erb DKO macrophages with gene ontology annotations linked to response to wounding and immune response are indicated in Figure 3e. These results indicate that tissue homogenate induces a Rev-erb-sensitive program of macrophage gene expression that substantially overlaps with the pattern of gene expression observed in response to wounding in vivo.

### Genes characteristic of alternate polarization states are co-expressed within individual cells

The approaches used thus far evaluated populations of cells. Genes associated with distinct polarization states resulting from activation with single ligands but exhibiting co-expression following treatment with tissue homogenate could reflect co-expression at the single cell level or mutually exclusive expression in subpopulations. To address this question, we performed RT-Q-PCR analysis of mRNA isolated from single cells maintained under control conditions or treated with tissue homogenate for 6 hr. We evaluated panels of mRNAs in triplicates corresponding to genes selectively activated by LPS or LPS+IFN $\gamma$ , IL4, TGF $\beta$ , or tissue homogenate signals, as well as informative transcription factors and reference genes. After filtering for dead/duplicate cells and eliminating probes with altered melting curves, data was obtained for 30 genes in 80 control cells and 70 homogenate-treated cells. The distributions of expression values of genes in individual cells under



**Figure 4.** Genes characteristic of alternate polarization states are co-expressed within individual cells. (a) Violin plots of expression values for genes in the indicated categories as determined by single cell RT-Q-PCR from bone marrow derived macrophages treated for 6 hr with vehicle or tissue  
 Figure 4 continued on next page

Figure 4 continued

homogenate. Y-axis shows RT-Q-PCR CTs. Higher values indicate lower expression (30: gene product could not be detected). Values are averaged over 3 PCR replicates per gene. (b) Hierarchical clustering with Euclidean distance of single bone marrow derived macrophages treated with tissue homogenate based on expression (red) or lack of expression (blue) of the genes indicated at the top. Genes with alternating melting curves were treated as undefined (grey). PCR replicates are shown sequentially (N = 3). (c) RT-Q-PCR expression values for genes indicated above for four representative cells. Y-axis normalized to (30/CT) – 1. Higher values indicate higher expression (0: gene product could not be detected). Values are averaged over 3 PCR replicates per gene.

DOI: 10.7554/eLife.13024.009

control or tissue homogenate treatment conditions are illustrated in **Figure 4a**. Cells treated with tissue homogenate were clustered in a binary fashion, according to whether the gene was expressed or not expressed. Notably, evaluating individual genes by column, a subset from each category of polarization states is expressed in the majority of cells (e.g., *Cxcl1*, *Dusp4*, *Cx3cr1*, *Pf4*) (**Figure 4b**). Conversely, evaluating the total set of genes across individual cells, genes from each polarization state can be expressed at similar levels in the same cell (**Figure 4c**). Of interest, clustering revealed two main groups that were distinguished by lack of detectable expression of *Fos* and *Rela*. Cells lacking *Fos* and *Rela* expression also exhibit reduced expression of subsets of genes in the M(LPS) or M(LPS+IFN $\gamma$ ), tissue homogenate, and transcription factor categories. Collectively, these findings indicate that while there is substantial heterogeneity in gene expression at the single cell level, genes characteristic of M(LPS) or M(LPS+IFN $\gamma$ ), M(TGF $\beta$ ), and M(IL4) polarization states can be co-expressed in individual cells.

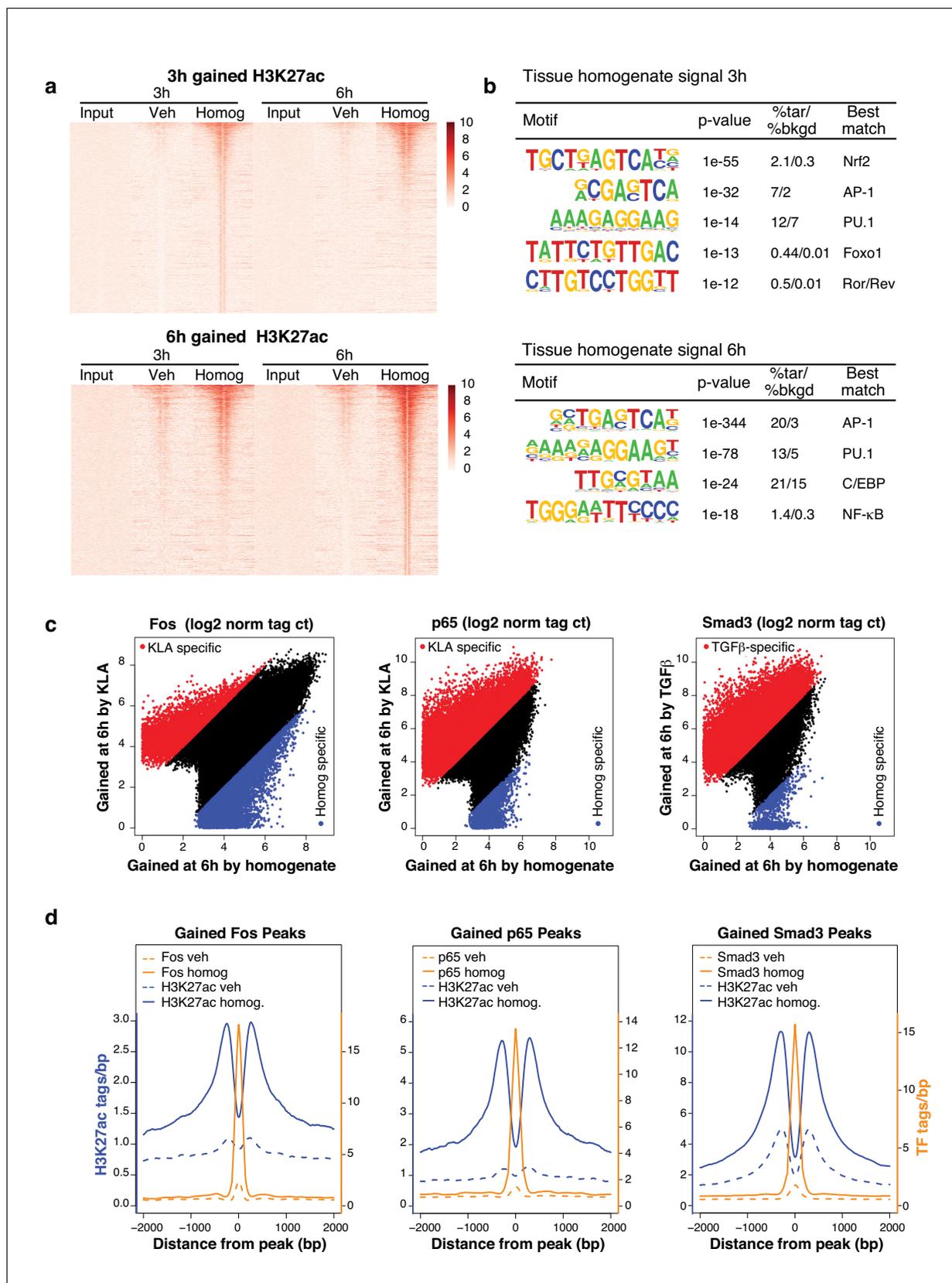
### Complex signals re-allocate transcription factors to novel genomic loci

To investigate mechanisms underlying effects of tissue homogenate on gene expression, we performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) for histone 3 lysine 27 acetylation (H3K27ac), a histone modification associated with active enhancers and promoters (Creyghton et al., 2010), after 3 or 6 hr of control or tissue homogenate stimulation. Treatment with tissue homogenate induced H3K27ac at ~2500 regions after 3 hr and ~5000 regions after 6 hr (**Figure 5a**). *De novo* motif analysis revealed binding sites for Nrf2, AP-1, and NF- $\kappa$ B motifs as among the most highly enriched sequences in these regions (**Figure 5b**).

Based on these motif findings, we initially performed ChIP-Seq analysis for p65 and Fos in macrophages treated with control or tissue homogenate. In addition, because SMAD motifs are difficult to retrieve using *de novo* motif analysis and tissue homogenate stimulation resembled treatment with TGF $\beta$  (**Figure 1b**), we performed corresponding ChIP-Seq analysis of Smad3. In each case, we observed that tissue homogenate induced a pattern of genomic binding sites that substantially differed from the pattern resulting from stimulation with the single ligands, KLA (p65 and Fos) or TGF $\beta$  (Smad3) (**Figure 5c**). These binding sites were also highly associated with tissue homogenate-induced gain of H3K27ac, consistent with their contribution to these changes in active chromatin (**Figure 5d**).

Examples of the binding patterns of Fos, p65, Smad3, and PU.1 in the vicinity of highly regulated genes are illustrated in **Figure 6a**, with responses of corresponding mRNAs to KLA, TGF $\beta$ , and tissue homogenate in WT and Rev-erb DKO macrophages shown in **Figure 6b**. Each genomic location contains numerous binding sites for each factor. PU.1 and Fos exhibit a high degree of constitutive binding, consistent with roles as pioneering factors that collaborate with each other and other macrophage lineage-determining factors, but also show quantitative changes in response to KLA and tissue homogenate. Smad3 and p65 both exhibit strong signal-dependent increases in ChIP-Seq signal at the majority of their binding sites. We note here that the starting conditions for KLA induction and tissue homogenate treatment differ, resulting in more constitutive binding of p65 in the vehicle control for tissue homogenate experiments.

Overall, there is a strong co-occurrence of p65 and Smad3 with pre-existing binding of Fos and PU.1, consistent with roles of these factors in establishing open regions of chromatin. Despite exhaustive efforts, we were not successful in determining high-confidence cistromes for endogenous Rev-erbs in BMDMs. We therefore considered the genomic locations of 7889 high-confidence binding sites occupied by both Rev-erb $\alpha$  and Rev-erb $\beta$  defined by ChIP-Seq of biotin-tagged proteins in RAW264.7 macrophages (Lam et al., 2013). For the de-repressed genes in the Rev-erb DKO, such as *Cx3cr1*, *Mmp9*, *Arg1*, and *Socs3* (**Figure 6a**), strong Rev-erb peaks coincide with at least one



**Figure 5.** Complex transcriptional signals re-allocate transcription factors to novel genomic loci. (a) H3K27ac regions differentially gained upon treatment for 3 or 6 hr with tissue homogenate compared to treatment for 3 or 6 hr with the vehicle control. The heatmap shows a 6 kb window of *Figure 5 continued on next page*

Figure 5 continued

normalized H3K27ac tag counts of the 2510 or 5005 homogenate gained regions at 3 or 6 hr, respectively, centered on the nucleosome free region (nfr). Input shows genomic background at these regions. N = 1. (b) Motifs enriched in the vicinity of gained H3K27ac sites after treatment with tissue homogenate for 3 or 6 hr using *de novo* motif enrichment analysis. (c) Comparison of Fos, p65, or Smad3 tag counts at genomic regions that contain Fos, p65, or Smad3 binding after stimulation with KLA, TGF $\beta$ , or tissue homogenate. Peaks found to be differentially gained (four-fold more tags) with KLA or TGF $\beta$  are colored red, while peaks found to be differentially gained (four-fold more tags) upon tissue homogenate treatment are colored blue. N = 1. (d) Quantification of H3K27ac, Fos, p65, and Smad3 ChIP-Seq tag counts in the 6 hr vehicle or tissue homogenate treated states centered on homogenate gained (using HOMER) Fos, p65, or Smad3 binding events. Dashed lines represent ChIP-Seq signal of the vehicle state and solid lines represent the signal after 6 hr of tissue homogenate stimulation. Blue represents H3K27ac signal, orange represents signal of the respective transcription factor. N = 1.

DOI: 10.7554/eLife.13024.010

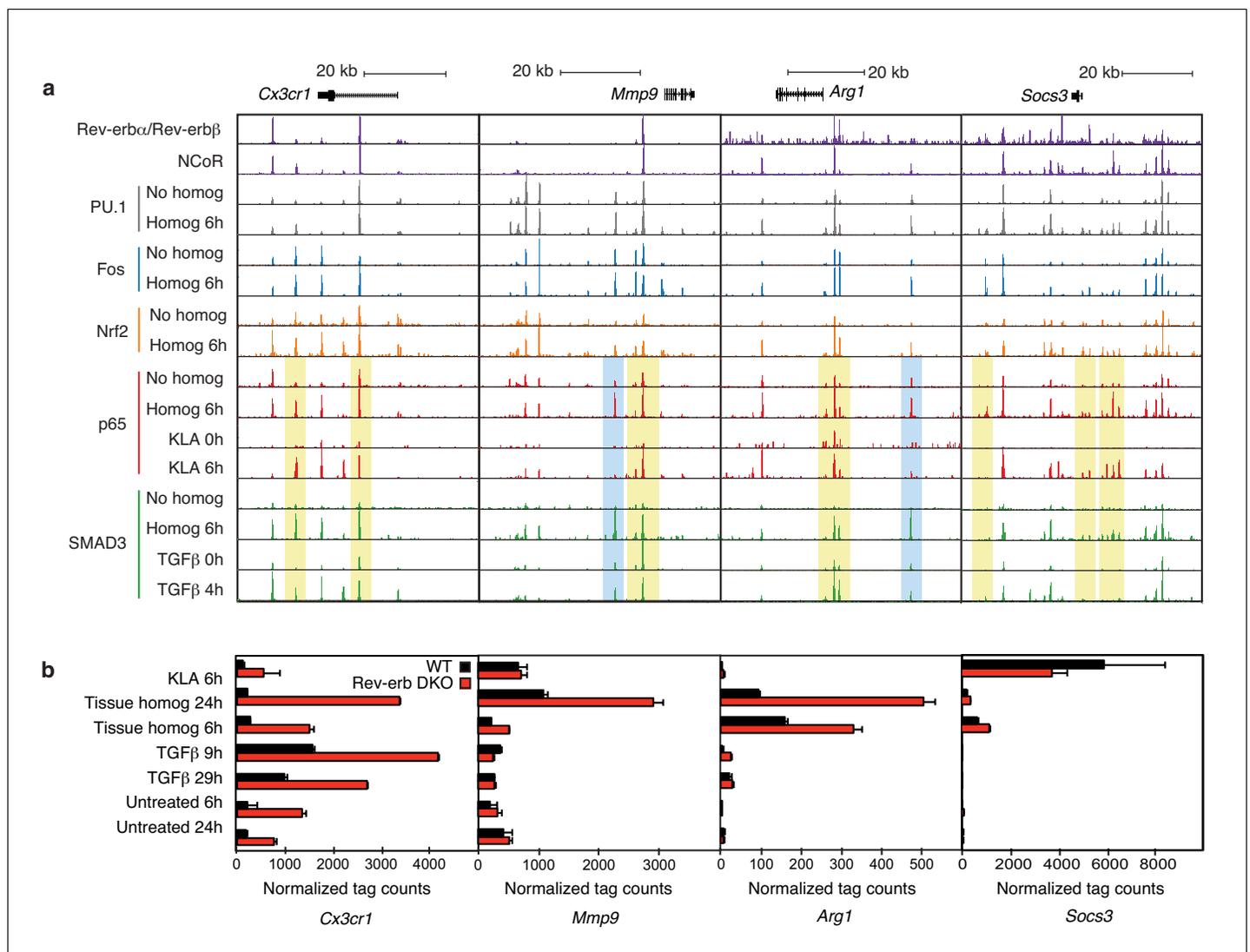
nearby enhancer-like region occupied by PU.1 and/or Fos, as well as p65 and/or Smad3 (e.g., **Figure 6a**). Furthermore, the majority of Rev-erb binding sites identified in RAW264.7 macrophages co-localize with binding sites for their obligate co-repressor NCoR in BMDMs, strongly suggesting that Rev-erbs occupy a similar cistrome in these cells (**Figure 6a**).

Two observations were unexpected and noteworthy. First, a subset of enhancer-like regions occupied by Smad3 in cells treated with TGF $\beta$  were occupied by p65 in cells treated with KLA. Under conditions of stimulation with either KLA or TGF $\beta$  alone, the expectation is that these regions would be occupied by one factor, but not the other. However, in the context of tissue homogenate treatment, both factors are simultaneously bound (**Figure 6a**, yellow boxes). Second, and consistent with the results presented in **Figure 5c**, tissue homogenate treatment leads to binding sites for p65 and Smad3 that are not observed following treatment with KLA or TGF $\beta$ , respectively (**Figure 6a**, blue boxes). Furthermore, many of the new binding sites for p65 co-localize with Smad3 and vice versa. Consistent with these findings at individual genomic locations, motif analysis of tissue homogenate-specific SMAD binding sites (from **Figure 5c**) using TGF $\beta$ -specific SMAD sites as the background returned an NF- $\kappa$ B recognition motif as the second most highly enriched motif (**Figure 7a**).

To investigate whether co-localization of p65 and Smad3 in tissue homogenate-treated cells was a specific consequence of the complex signal, we performed ChIP-Seq analysis of p65 in macrophages selectively treated with TGF $\beta$ . We observed ~7400 p65 peaks, 5465 of which overlapped with the 39,825 peaks for Smad3 observed in TGF $\beta$ -treated cells, representing an overlap with 6% of the Smad3 peaks (**Figure 7b**). In contrast, we observed 20,858 p65 peaks and 13,975 Smad3 peaks in homogenate-treated cells, with p65 co-localizing with 11,379 (82%) of the Smad3 binding sites. Therefore, the complex tissue homogenate signal drives substantial co-localization of p65 with Smad3 that is not observed following selective treatment with TGF $\beta$ . These relationships are further illustrated for two representative genes, *Arg1* and *Cxcl2*, in **Figure 7c**, in which yellow shading indicates regions where tissue homogenate induced p65 binding to regions occupied by Smad3 under either TGF $\beta$  or tissue homogenate treatment, whereas blue shading indicates regions in which both p65 and Smad3 binding are selectively observed following treatment with tissue homogenate.

## Nrf2 target genes and Nrf2 genomic binding are induced by tissue damage signals

Unexpectedly, the top enriched motif in tissue homogenate-specific SMAD sites is a binding site for NFE2L2, also known as Nrf2 (**Figure 7a**). This was also the top motif recovered from motif analysis of genomic regions exhibiting a gain in H3K27ac 3 hr following stimulation with tissue homogenate (**Figure 5b**). Nrf2 is a latent basic leucine zipper (bZIP) protein that is activated in response to cell injury and inflammation, and regulates the expression of antioxidant proteins that protect against oxidative damage (**Chen et al., 2015**). Evaluation of RNA-seq data from both the *in vivo* wound model and tissue homogenate-treated macrophages revealed up regulation of numerous Nrf2 target genes, including *Txn1*, *Sod2*, *Hmox1*, *Prdx6*, and *Nqo1*, suggesting that Nrf2 is activated in macrophages as part of the wound response (**Figure 7d**). We therefore performed ChIP-Seq analysis for Nrf2 in macrophages before and after tissue homogenate treatment. These experiments demonstrated that tissue homogenate increased the genome-wide binding of Nrf2 at thousands of genomic locations, a substantial fraction of which were observed to overlap with the tissue homogenate-induced binding sites for p65, Fos, and Smad3 (e.g., **Figure 6a**).



**Figure 6.** Locus-specific effects of Rev-erbs and signal-dependent transcription factors. (a) UCSC genome browser images depicting the genomic regions surrounding Rev-erb target genes, *Cx3cr1*, *Mmp9*, *Arg1*, or *Socs3*. The image shows the co-localization of PU.1, Rev-erbs, Fos, Nrf2, p65, and Smad3 upon homogenate treatment. Yellow denotes gain of signal-dependent transcription factor peaks (p65 or Smad3) and Nrf2 after tissue homogenate stimulation that is not seen after treatment with one single polarizing signal. Blue denotes gain of signal-dependent transcription factor peaks (p65 and Smad3) and Nrf2 after tissue homogenate stimulation that is not seen after treatment with either single polarizing signal. (b) Comparison of normalized RNA-Seq tag counts from WT or Rev-erb DKO macrophages stimulated as indicated. Error bars show standard deviation. For unstimulated 6 hr and 24 hr samples, N = 2, samples stimulated with KLA, tissue homogenate for 6 or 24 hr or TGFβ for 9 or 29 hr, N = 3.

DOI: [10.7554/eLife.13024.011](https://doi.org/10.7554/eLife.13024.011)

## Tissue damage signals drive co-localization of PU.1, p65, Fos, Smad3, and Nrf2

To further explore the signal-dependent binding patterns of p65, Fos, Smad3, and Nrf2, we performed unbiased hierarchical clustering using peaks gained after stimulation with tissue homogenate or single stimuli. This analysis revealed that genomic occupancy of PU.1, Smad3, Nrf2, Fos, and p65 was most similar upon treatment of macrophages with the tissue homogenate signal, whereas patterns of transcription factor binding were more varied upon treatment of macrophages with individual stimuli (Figure 7e). This co-binding of transcription factors is further emphasized when comparing the overlap of the investigated transcription factors upon treatment of cells with tissue homogenate or the vehicle control (Figure 7f and g). This approach demonstrated co-localization of only 1.4% (893) of peaks in the vehicle state, which increased to 12.05% (7758) overlap when cells

were treated with tissue homogenate. This eight-fold increase in co-localization suggests that the combination of signals present in tissue homogenate induce co-binding of multiple transcription factors to enhancers that mediate the tissue injury response.

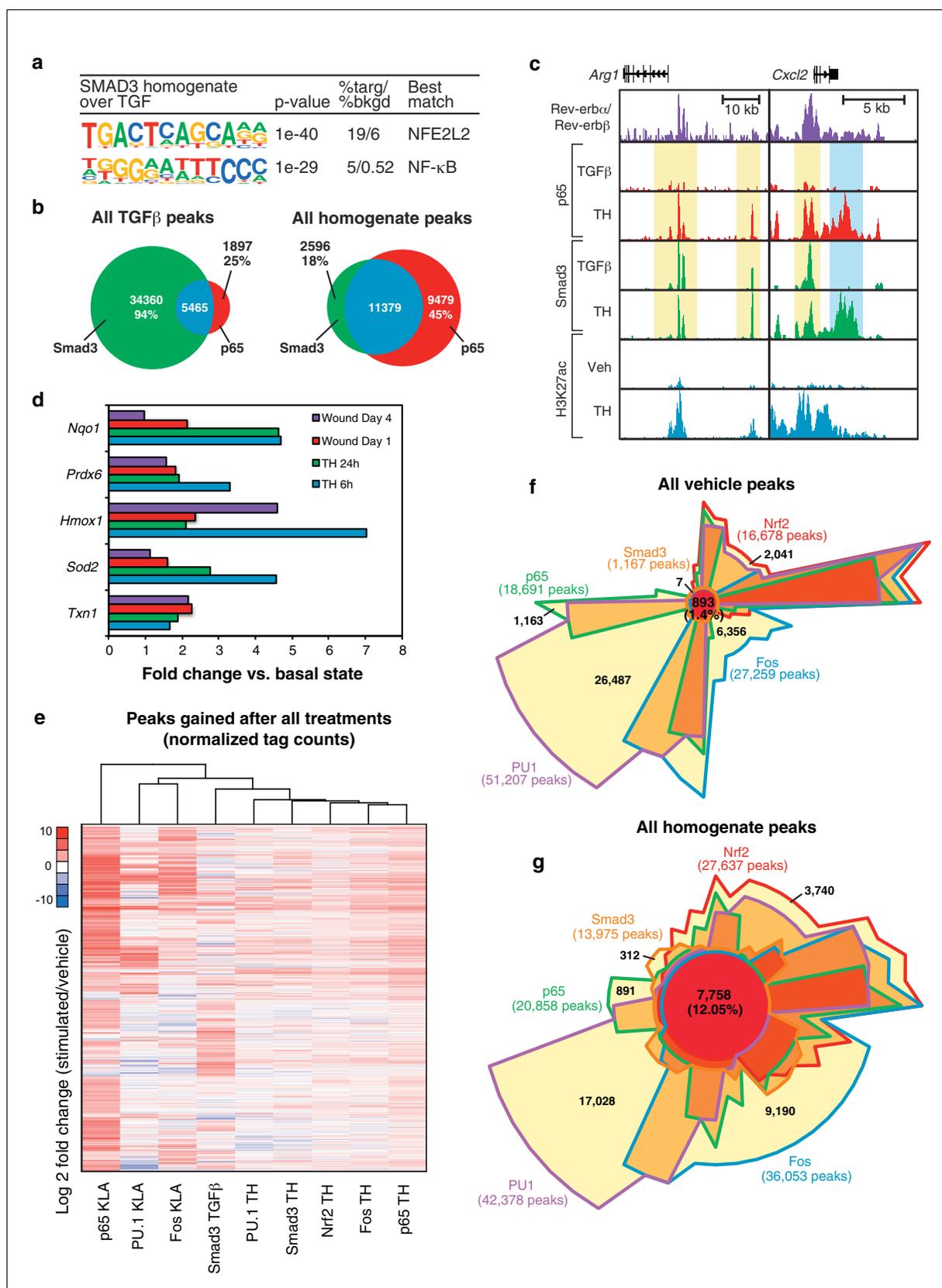
## NF- $\kappa$ B, Smad, Nrf2, and Rev-erb signaling pathways contribute to the integrated tissue damage response

Tissue homogenate contains a combination of DAMPs, MAMPs and other factors that have the potential to activate numerous signaling pathways. While ChIP-Seq experiments documented that tissue homogenate induces genomic binding of p65, Smad3, and Nrf2, these studies do not establish functional roles of these factors in the integrated transcriptional response. To address this question, we evaluated effects of chemical inhibitors of NF- $\kappa$ B, Smad3, and Nrf2 on gene expression in tissue homogenate-treated macrophages, using the IKK inhibitor VII to inhibit NF- $\kappa$ B activity, SB-43154 to inhibit TGF $\beta$  signaling, and glutathione to block the activation of Nrf2 (**Figure 8a–c**). These studies support the idea that each factor contributes to the integrated response to tissue homogenate. For example, activation of *Cx3cr1* by tissue homogenate was decreased upon targeting the NF- $\kappa$ B, TGF $\beta$  receptor, and Nrf2 pathways, supporting the involvement of all of these pathways in the regulation of this gene. Conversely, tissue homogenate activation of other genes was more dependent on specific signal-dependent pathways. For instance, *Dusp5* activation was sensitive to NF- $\kappa$ B inhibition (**Figure 8a**) while *Nptx1* activation was unaffected by NF- $\kappa$ B inhibition (**Figure 8a**). Surprisingly, *Socs3* activation was sensitive to both inhibition of NF- $\kappa$ B and TGF $\beta$  receptor signaling (**Figure 8a and b**) and *Ctla2b* was selectively sensitive to inhibition of TGF $\beta$  receptor signaling (**Figure 8b**). Established Nrf2 target genes *Txn1* and *Hmox1*, which were also induced by tissue homogenate, were repressed by glutathione co-treatment (**Figure 8c**). Finally, we investigated the ability of the Rev-erb agonist SR-9009 to influence the responses to tissue homogenate. This agonist repressed a subset of genes in tissue homogenate-treated macrophages, exemplified by *Cx3cr1*, *Gpr84*, and *Pgd* (**Figure 8d**). These results are consistent with these genes being de-repressed in Rev-Erb DKO macrophages.

## Discussion

Rev-erbs have been established to play general roles in the regulation of promoters of ubiquitously expressed genes such as *Bmal* that control the circadian rhythm (**Preitner et al., 2002; Liu et al., 2008; Cho et al., 2012**). However, the great majority of Rev-erb binding sites in macrophages are located at cell-specific enhancers, which are selected by macrophage lineage-determining factors such as PU.1 (**Lam et al., 2013**). These observations predicted that in addition to cell autonomous regulation of the circadian rhythm, Rev-erbs would also regulate a macrophage-specific program of gene expression. Here, using loss of function, transcriptomic, and epigenetic analyses, we demonstrate that Rev-erbs function to repress a network of genes associated with the response to wounding. Consistent with altered transcriptional responses observed in vitro, loss of Rev-erb expression in cells derived from the bone marrow compartment resulted in accelerated wound healing in the skin. As the Rev-erbs are deleted from all hematopoietic lineages in these experiments, further studies will be required to establish the relative contributions of macrophages and other bone marrow-derived cells to this phenotype. How this function of Rev-erbs might contribute to normal tissue homeostasis is as yet unclear. In vivo, Rev-erb expression is circadian (**Cho et al., 2012**), implying that the effects on macrophage gene expression observed in the present studies are likely to vary over the course of the day. Rev-erbs may thus act in a circadian manner to regulate aspects of tissue macrophage gene expression required for the normal turnover of extracellular matrix, tissue remodeling, and wound healing.

By evaluating the consequences of Rev-erb deficiency on macrophage gene expression in response to distinct polarizing signals in vitro, we found that the consequences of loss of function of Rev-erbs were dependent on the specific polarizing signal. Consistent with this, Rev-erbs co-localize with NF- $\kappa$ B p65 and AP-1 family member Fos at enhancers activated by TLR ligands, and with Smad3 at enhancers activated by TGF $\beta$ . Although of interest from a mechanistic standpoint, these findings are of uncertain relevance to functions of macrophages within tissue environments, which contain a multitude of signaling molecules that are sensed simultaneously. To model the complex environment associated with acute tissue damage, we treated macrophages with a supernatant of a



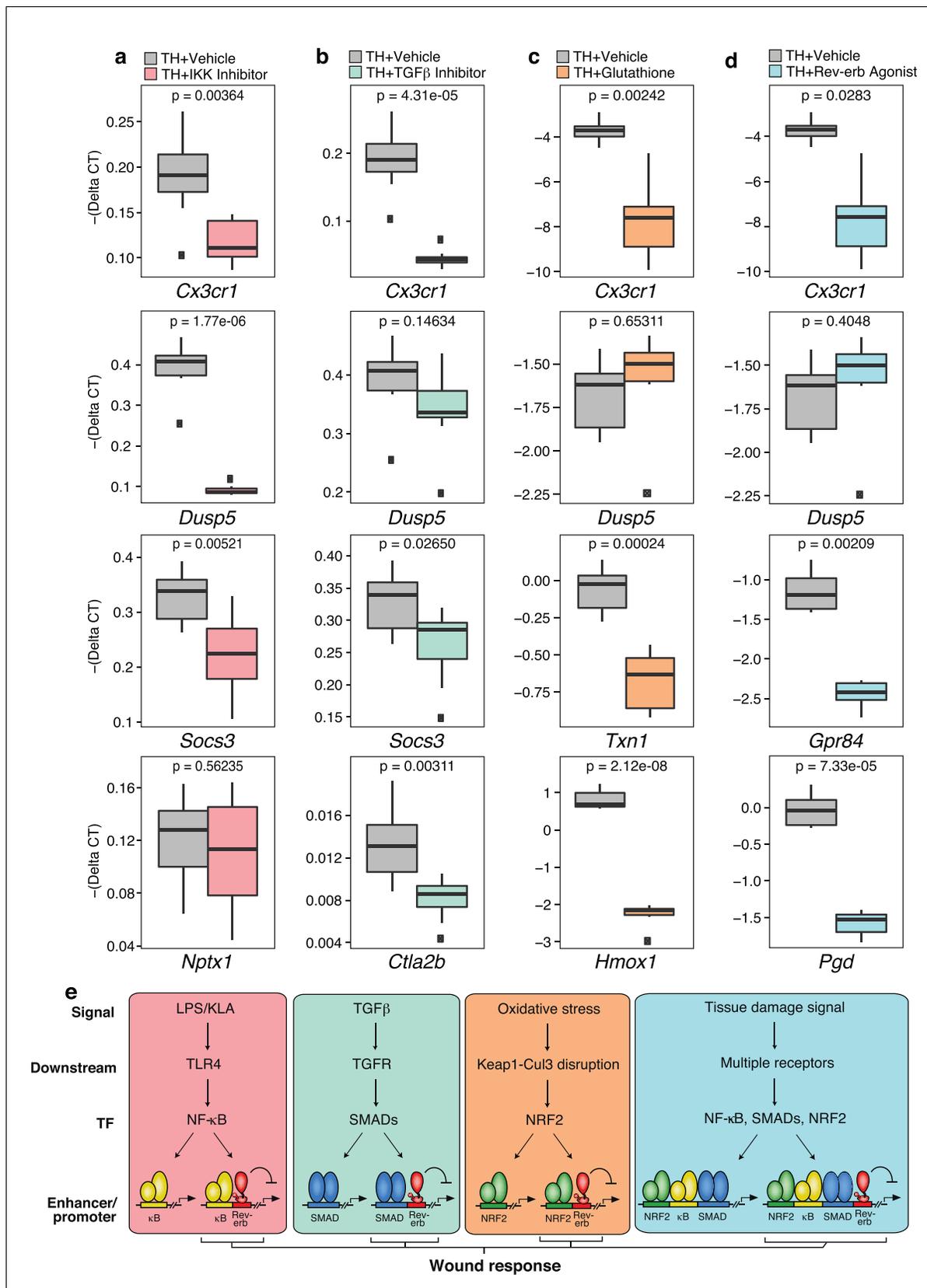
**Figure 7.** Signal-dependent transcription factors collaborate in response to complex stimuli. (a) Motifs enriched in the vicinity (200 bp) of Smad3 binding sites that are specific to tissue homogenate (induced four-fold), as compared to Smad3 binding sites that are specific to TGFβ (induced four-fold). *Figure 7 continued on next page*

Figure 7 continued

fold) using *de novo* motif enrichment analysis. (b) Venn diagrams depicting overlap of Smad3 and p65 after treatment with TGF $\beta$  (left) or tissue homogenate (right). Peaks have a minimal normalized tag count of 16. N = 1. (c) UCSC genome browser images depicting the genomic regions surrounding genes highly induced by tissue homogenate and not by TGF $\beta$ . Yellow denotes gain of signal-dependent transcription factor peaks (p65 or Smad3) after tissue homogenate stimulation that is not seen after treatment with one single polarizing signal. Blue denotes gain of signal-dependent transcription factor peaks (p65 and Smad3) after tissue homogenate stimulation that is not seen after treatment with either single polarizing signal. (d) Bar graphs depicting canonical Nrf2 genes induced during stimulation with tissue homogenate (green and blue) or during wound healing (purple and red). N as described in 3b. (e) Heatmap showing the log<sub>2</sub> fold change of transcription factor tag counts at all genomic regions (minimum of 64 normalized tag counts in at least one condition per row) that are differentially gained (four-fold more tags) after stimulation as indicated. N = 1. (f) Chow-Ruskey Venn diagrams depicting the overlap of all p65, PU.1, Fos, Smad3, and Nrf2 peaks after treatment with vehicle. Peaks have a minimal normalized tag count of 16. N = 1. (g) Same as (f) but depicting the overlap of all peaks gained after treatment with tissue homogenate. N = 1.  
DOI: [10.7554/eLife.13024.012](https://doi.org/10.7554/eLife.13024.012)

skin homogenate. While the specific identities and concentrations of the DAMPs, MAMPs, and other bioactive molecules in this homogenate are unknown, we provide evidence that the transcriptomic response of the macrophage to this mixture overlaps significantly with the transcriptional response observed in a skin wound, thereby validating its use. Through ChIP-Seq experiments, we demonstrate that this complex signal coordinately induces binding of NF- $\kappa$ B, AP-1, and Smad transcription factors. Furthermore, *de novo* motif analysis of activated enhancers led to the unexpected discovery that the tissue damage signal also acutely activates Nrf2. This finding illustrates the utility of enhancer analysis to identify transcriptional mediators of unknown environmental factors, providing a basis for subsequent directed analysis of corresponding upstream signaling pathways. Accordingly, the use of glutathione to neutralize reactive oxygen species, thus blocking the downstream disruption of the Kelch-like ECH-associated protein 1 (Keap1)-Cuilin 3 (Cul3) complex required for activation of Nrf2 (Gorrini *et al.*, 2013; Shibata *et al.*, 2013) provides evidence for its functional importance in the transcriptional response to the tissue damage signal. Similarly, the use of inhibitors of NF- $\kappa$ B and TGF $\beta$  provided corresponding support for functionally important roles of these transcription factors. Of course, there are likely to be many other signaling pathways and downstream transcription factors involved in the tissue damage response. Furthermore, Rev-erb deficiency likely modifies both basal and signal dependent transcriptional programs. To distinguish between 'prior' versus 'post-activation' roles of Rev-erbs in macrophages during wound healing may require the use of inducible Cre-expression strategies, as well as measurements of target gene expression in situ in macrophage infiltrated wounds.

Three additional observations are of particular interest. The first is that the complex signal provided by tissue homogenate induced co-expression of genes characteristic of distinct macrophage polarization states within individual cells. Second, we found that the tissue homogenate signal induced different genomic locations of p65, Fos, and Smad3 than were observed following KLA or TGF $\beta$ , respectively, resulting in co-binding at a large number of enhancer-like regions in the vicinity of tissue homogenate-induced genes. An important implication of these findings is that transcription factors binding maps are context-dependent and must be interpreted accordingly. We speculate that the observed co-localization of factors in response to the complex signal enables the appropriate integration of multiple relevant signaling components necessary for the initial acquisition of a wound repair phenotype (Figure 8e). Third, the present findings may have practical applications based on the development of small molecules that enhance or inhibit Rev-erb repressive activity (Solt *et al.*, 2012). Delayed wound healing is observed in a number of pathological contexts, including in diabetics (Falanga, 2005; Sen *et al.*, 2009) and in immunocompromised individuals (Chen *et al.*, 2013; Lin *et al.*, 2011). In these settings, it is possible that Rev-erb antagonists could be evaluated as a means of enhancing wound repair. Alternatively, a large number of devastating and largely untreatable diseases are characterized by exaggerated tissue fibrosis, such as idiopathic pulmonary fibrosis, interstitial renal fibrosis, and liver fibrosis (Schuppan and Kim, 2013). We demonstrate that a Rev-erb agonist can suppress a subset of genes that are de-repressed in the Rev-erb DKO and are regulated by the complex wound signal. Overall, our findings suggest that Rev-erbs act to repress a specific combination of genes downstream of multiple signaling pathways that collectively function in an integrated manner to promote the response to wounding (Figure 8e). It will therefore be of interest to evaluate whether defects in Rev-erb signaling are associated with these diseases and whether pharmacological modulation of Rev-erb might be of therapeutic benefit.



**Figure 8.** Chemical inhibition of multiple signal-dependent pathways results in decreased response to complex tissue homogenate. (a) Box and whisker plots of expression values for genes as determined by RT-Q-PCR from bone marrow derived macrophages treated with tissue homogenate for 6 hr, Figure 8 continued on next page

Figure 8 continued

and pre-treated for 1 hr with vehicle (gray) or 1  $\mu$ M IKK inhibitor VII (pink). Y-axis shows RT-Q-PCR  $-(\Delta\text{CT})$ , which is normalized to the housekeeping gene 36B4. Lower values indicate lower expression. N = 8 mice. p-values are shown comparing vehicle versus inhibitor treatment as determined by unpaired t-test. (b) Same as (a) but with pre-treatment for 1 hr with vehicle (gray) or 12.5  $\mu$ M TGF $\beta$  inhibitor SB-43154 (green). N = 8 mice. (c) Same as (a) but samples were co-treated with tissue homogenate and vehicle (gray) or tissue homogenate and 15 mM glutathione (orange) for 6 hr. N = 6 mice. (d) Same as (a) but with pre-treatment for 1 hr with vehicle (gray) or 15  $\mu$ M Rev-erb agonist SR-9009 (blue). N = 6 mice. (e) Working model showing that signal-specific stimuli (KLA/TGF $\beta$ /oxidative stress) activate their respective signal-dependent transcription factors, NF- $\kappa$ B, Smads, and Nrf2, which bind to distinct sets of enhancers and promoters. Tissue damage signal activates all three factors simultaneously, which can co-occupy enhancers and promoters to generate a transcriptional response that is different than the sum of TGF $\beta$ , KLA, and oxidative stress mediated signaling. A subset of these sites that are co-bound and repressed by Rev-erbs are involved in regulating the macrophage response to wounding.

DOI: 10.7554/eLife.13024.013

## Materials and methods

### Cell culture

Mouse bone marrow derived macrophages were obtained and cultured as previously described (Heinz *et al.*, 2010). For cytokine stimulation studies, macrophages were cultured in RPMI-1640 (Invitrogen, Waltham, MA) supplemented with 16.7 ng/ml CSF1 (Shenandoah Biotechnology, Warwick, PA) and 0.5% heat-inactivated fetal bovine serum (FBS) (Hyclone, Logan, UT) overnight and then stimulated with Pam3CSK4 (300 ng/ml, InvivoGen, San Diego, CA), polyinosinic-polycytidylic acid (50 ng/ml, GE Healthcare Bioscience, Pittsburgh, PA), KDO2-Lipid A (100 ng/ml, Avanti Polar Lipids, Alabaster, AL), recombinant interferon  $\gamma$  (10 U/ml, R&D Systems, Minneapolis, MN), interleukin 4 (20 ng/ml, R&D Systems), or tumor growth factor  $\beta$  (1 ng/ml, Cell Signaling, Danvers, MA) for the indicated time points. For ChIP-Seq experiments, cells treated with TGF $\beta$  or KLA were compared to untreated cells as a control.

For whole-skin tissue homogenate studies, skin from shaved wild type congenic mice was harvested and homogenized in RPMI-1640 supplemented with 0.5% heat-inactivated FBS using a Precellys 24 tissue homogenizer (6500 RPM, 4  $\times$  20 s) and metal beads (2.8 mm beads, 2 mL tubes, Cayman Chemical, Ann Arbor, MI) according to the manufacturer's instructions. Skin homogenate was then centrifuged at 4000 RPM for 30 min at 4°C, and the supernatant filtered through a 0.2  $\mu$ m filter (Nalgene, ThermoScientific, Rochester, NY). Approximately 50 ml of tissue homogenate was collected per mouse. To stimulate macrophages, macrophages were cultured in RPMI-1640 (Invitrogen) supplemented with 16.7 ng/ml CSF1 and 0.5% heat-inactivated FBS overnight. The following morning, the media was switched to either tissue homogenate or vehicle. 16.7 ng/ml CSF1 was added exogenously to both the homogenate and vehicle treatment conditions. For RNA-Seq replicates using BMDMs (where applicable), bone marrow of several mice were pooled and then cultured in different wells and processed independently.

For inhibitor experiments, macrophages were generated and cultured as described above, then pre-treated with 1  $\mu$ M IKK inhibitor VII (Calbiochem, Billerica, MA), 12.5  $\mu$ M TGF $\beta$  RI kinase inhibitor SB-43154 (Calbiochem), or 5  $\mu$ M Rev-erb agonist SR-9009 (Burriss laboratory) (Lewis *et al.*, 2013) for 1 hr before treatment with tissue homogenate. For anti-oxidant experiments, macrophages were co-treated with tissue homogenate and 15 mM glutathione (Sigma, St. Louis, MO).

### Rev-erb DKO mice and genotyping

Rev-erb $\alpha$  and Rev-erb $\beta$  double floxed mice were generated as previously described (Cho *et al.*, 2012) and crossed with Tie2-Cre (Lam *et al.*, 2013). Breeding and genotyping were performed as previously described (Lam *et al.*, 2013). Only males were used for wound healing experiments while both males and females were used for flow cytometry experiments to enumerate monocyte populations from peripheral blood. Littermates without the Tie2-Cre transgene were used as WT controls. All animal procedures were performed in accordance with the University of California, San Diego research guidelines for the care and use of laboratory animals (Permit Number: S01015).

### RNA isolation and RT-Q-PCR

Total RNA was harvested from tissue and cells using the RNeasy Mini Kit (Qiagen, Hilden, Germany) with in column DNase digestion performed according to the manufacturer's instructions. DNase-

treated RNA was used for cDNA synthesis using Superscript III (Invitrogen) according to the manufacturer's instructions.

For the IKK inhibitor VII and TGF $\beta$  inhibitor experiments, cDNA from biological replicates (N = 8) were assessed by quantitative polymerase chain reaction using SYBR GreenER Master Mix (Invitrogen) or SYBR Fast qPCR Master Mix (Kapa Biosystems, Wilmington, MA) on an Applied Biosystems 7300 Real-time PCR system or Step One Plus. For the glutathione and Rev-erb agonist experiments, cDNA from biological replicates (N = 6) were synthesized and assessed in technical triplicates by quantitative PCR using a Fluidigm Biomark HD (Fluidigm, San Francisco, CA). For statistical analysis, the delta CT was calculated for each biological replicate using 36b4 as the reference gene. Data were compared statistically using the t-test command in R.

### Single Cell RT-Q-PCR and analysis

BMDMs on petri plates were treated for 6 hr with vehicle or tissue homogenate in the presence of recombinant CSF-1. Following treatment, cells were removed by scraping and captured on a Fluidigm 17–25 micrometer C1 Single-Cell Auto Prep Array IFC or a 10–17 micrometer C1 Single-Cell Auto Prep Array IFC for homogenate or vehicle treated cells, respectively, according to the manufacturer's instructions. IFC positions having a single viable cell were noted and gene specific priming and pre-amplification was performed using the Fluidigm C1 instrument and the instrument protocol number 100–4904 H1. After cDNA synthesis, samples were harvested and stored at  $-20^{\circ}\text{C}$  prior to detection of cDNA using Fluidigm 96.96 Dynamic arrays using the instrument protocol number 100–9792 A1. cDNA from individual cells was assessed in triplicate using the primers listed below.

For analysis of the data, melting curves of the triplicates were compared and samples with different melting curves or melting curves with more than one product were defined as NA. Gene products that could not be detected by Q-PCR were set to a CT of 30. A majority analysis was applied to the triplicates to calculate the average CT per primer pair per single cell (for violin plots [Font-Vizcarra et al., 2012]). Data was converted into binary data for gene expression heatmap, using 1 for expressed, 0 for not expressed (CT equals 30) and NA for undetermined. For cases where one sample of the triplicates had value 1, one had value 0 and one had value NA, NA was used as consensus. The heatmap for single cell analysis was created using hierarchical clustering with Euclidian distance in R.

### RT-Q-PCR Primers

| Gene target         | Primer sequence              |
|---------------------|------------------------------|
| <b>Arg1-Forward</b> | <b>TTTATAGGGTTACGGCCGGTG</b> |
| Arg1-Reverse        | CCTCGAGGCTGTCTTTTGA          |
| Cd14-Forward        | CAGAGAACACCACCGCTGTA         |
| Cd14-Reverse        | CACGCTCCATGGTCGGTAGA         |
| Cd86-Forward        | CAGCACGGACTTGAACAACC         |
| Cd86-Reverse        | CTCCACGGAAACAGCATCTGA        |
| Ctla2b-Forward      | CTCATGCACTAGCCTCC            |
| Ctla2b-Reverse      | AGCAGGAAGACAGCACTGAA         |
| Cx3cr1-Forward      | CCATCTGCTCAGGACCTCAC         |
| Cx3cr1-Reverse      | CACCAGACCGAACGTGAAGA         |
| Cxcl1-Forward       | ACCCAAACCGAAGTCATAGCC        |
| Cxcl1-Reverse       | TTGTCAGAAGCCAGCGTTCA         |
| Cxcl2-Forward       | TGAACAAAGGCAAGGCTAACTG       |
| Cxcl2-Reverse       | CAGGTACGATCCAGGCTTCC         |
| Cxcl3-Forward       | ACCCAGACAGAAGTCATAGCCA       |
| Cxcl3-Reverse       | CTTCATCATGGTGAGGGGCT         |
| Dusp4-Forward       | CATCGAGTACATCGACGCGAG        |
| Dusp4-Reverse       | ATGAAGCTGAAGTTGGGCGA         |

|                           |                         |
|---------------------------|-------------------------|
| <i>Dusp5</i> -Forward     | GCACCACCCACCTACACTAC    |
| <i>Dusp5</i> -Reverse     | CCTTCTTCCTGACACAGTCAAT  |
| <i>Fos</i> -Forward       | TTTCAACGCCGACTACGAGG    |
| <i>Fos</i> -Reverse       | TCTGCGCAAAAGTCCTGTGT    |
| <i>Gpc1</i> -Forward      | GCCATGGAACTCCGGACC      |
| <i>Gpc1</i> -Reverse      | GCAGGTGCTCACCCGAGAT     |
| <i>Gpr84</i> -Forward     | AAACTGGGAACCTCAGTCTCCA  |
| <i>Gpr84</i> -Reverse     | GCCCAACACAGACTCATGGTA   |
| <i>Hmox1</i> -Forward     | GAGCAGAACCAGCCTGAACT    |
| <i>Hmox1</i> -Reverse     | AAATCCTGGGGCATGCTGTC    |
| <i>Hprt</i> -Forward      | GTTGGGCTTACCTACTGCT     |
| <i>Hprt</i> -Reverse      | TCATCGCTAATCACGACGCT    |
| <i>Igsf11</i> -Forward    | GTGTCGCTGCTCGGTGT       |
| <i>Igsf11</i> -Reverse    | AGAATGACCTGTTCCGGGCTG   |
| <i>Il10</i> -Forward      | GGTTGCCAAGCCTTATCGGA    |
| <i>Il10</i> -Reverse      | GGGGAGAAATCGATGACAGC    |
| <i>Il1b</i> -Forward      | TGCCACCTTTTGACAGTGATG   |
| <i>Il1b</i> -Reverse      | TGATGTGCTGCTGCGAGATT    |
| <i>Il1r1</i> -Forward     | GCTGACTTGAGGAGGCAGTT    |
| <i>Il1r1</i> -Reverse     | CATACGTCAATCTCCAGCGAC   |
| <i>Nptx1</i> -Forward     | TGGAGAACCTCGAGCAGTACA   |
| <i>Nptx1</i> -Reverse     | GTCAAGGCGCTCTCGATCTT    |
| <i>Pf4</i> -Forward       | CCCGAAGAAAGCGATGGAGAT   |
| <i>Pf4</i> -Reverse       | TTCAGGGTGGCTATGAGCTGG   |
| <i>Pgd</i> -Forward       | CTCCTCGACTCTGCTTCGTC    |
| <i>Pgd</i> -Reverse       | GCACAGACCACAAATCCATGA   |
| <i>Polr3c</i> -Forward    | TCTAAGAAGGGGCGATGGGA    |
| <i>Polr3c</i> -Reverse    | AGCCTCAGAACTCAGGGTCG    |
| <i>Ptgs2</i> -Forward     | AGCCAGGCAGCAAATCCTT     |
| <i>Ptgs2</i> -Reverse     | GGGTGGGCTTACAGCAGTAAT   |
| <i>Rela</i> -Forward      | CGGATTCGGGCAGTGAC       |
| <i>Rela</i> -Reverse      | GAGGGGAAACAGATCGTCCA    |
| <i>Smad3</i> -Forward     | AAGAAGCTCAAGAAGACGGGG   |
| <i>Smad3</i> -Reverse     | CAGTGACCTGGGGATGGTAAT   |
| <i>Socs3</i> -Forward     | TAGACTTCACGGCTGCCAAC    |
| <i>Socs3</i> -Reverse     | CGGGGAGCTAGTCCCGAA      |
| <i>Spi1</i> -Forward      | AAGCAGGGGATCTGACCAAC    |
| <i>Spi1</i> -Reverse      | AGTCATCCGATGGAGGGGC     |
| <i>Thbs1</i> -Forward     | GACAATTTTCAGGGGGTGCT    |
| <i>Thbs1</i> -Reverse     | AGAAGGACGTTGGTAGCTGAG   |
| <i>Tnf</i> -Forward       | GATCGGTCCCAAAGGGATG     |
| <i>Tnf</i> -Reverse       | GTGGTTTGTGAGTGTGAGGGT   |
| <i>Tnfrsf12a</i> -Forward | CAATCATGGCTTCGGCTTGG    |
| <i>Tnfrsf12a</i> -Reverse | CTGCGGCGCCTGGTG         |
| <i>Traf3ip2</i> -Forward  | CCTGCTCCACCACTTACCTG    |
| <i>Traf3ip2</i> -Reverse  | TCTAGTTTCTAAGATCGCCACCG |
| <i>Txn1</i> -Forward      | AGCCCTTCTTCCATTCCCTC    |

|                 |                        |
|-----------------|------------------------|
| Txn1-Reverse    | GGAAGGTCGGCATGCATTTG   |
| Ube2d2a-Forward | AGCTGAGTGGGGCCTCG      |
| Ube2d2a-Reverse | TCAATTCCTTGTGGATTCTTCA |

## RNA-Seq

Detailed protocols for RNA-Seq experiments have been previously described ([Kaikkonen et al., 2013](#); [Heinz et al., 2013](#)). Briefly, total RNA was isolated using TRIzol LS (ThermoFisher Scientific) and resuspended with UltraPure water (ThermoFisher Scientific) supplemented with 1  $\mu\text{M}$  SUPERase-In (Ambion) then treated with TURBO DNA-free kit (Ambion). Poly(A) selection was performed using the MicroPoly(A)Purist kit (Invitrogen) according to the manufacturer's instructions. Poly(A) RNA was fragmented using RNA Fragmentation Reagents (Ambion) for 10 min at 70°C and purified by running through a Micro Bio-Spin P-30 column (Bio-Rad, Irvine, CA) according to the manufacturer's instructions. 30 ng RNA was utilized for subsequent library preparation.

For the following RNA samples: two replicates of the four day 0 in vivo wound samples, day 1 wound samples, day 4 wound samples, and day 14 wound samples, RNA library preparation was performed as previously described ([Kaikkonen et al., 2013](#)). Fragmented RNA was de-phosphorylated using 1  $\mu\text{L}$  T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) and 5  $\mu\text{L}$  5x PNK buffer (0.5 M MES, 50 mM  $\text{MgCl}_2$ , 50 mM mercaptoethanol, 1.5 M NaCl, pH 5.5–5.8) supplemented with 1  $\mu\text{M}$  SUPERase-In for 45 min at 37°C, an additional 1  $\mu\text{L}$  T4 polynucleotide kinase was added to the reaction, followed by incubated for 45 min, and subsequent heat-inactivation for 5 min at 70°C and ethanol precipitation overnight with glycogen. The pellet was resuspended in 5.5  $\mu\text{L}$  nuclease free water supplemented with 1  $\mu\text{M}$  SUPERase-In and denatured for 5 min at 65°C. Poly(A)-tailing reaction was performed using 3.75  $\mu\text{L}$  E. coli poly(A)-polymerase (New England Biolabs) in 10x poly(A)-polymerase buffer supplemented with ATP (50:1 molar ratio to RNA) and 1  $\mu\text{M}$  SUPERase-In for 30 min at 37°C. Reverse transcription was performed using Superscript III (Invitrogen). 8  $\mu\text{L}$  RNA from the previous reaction, 1  $\mu\text{L}$  10 mM dNTP and 1  $\mu\text{L}$  of the following oligo with custom barcodes (underlined and bolded): 5'-Phos CA/TG/AC/GT-GATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGGCATAACGATTTTTTTTTTTTTTTTTTTTTTTVN-3' were incubated for 3 min at 75°C and then chilled on ice. 1.7  $\mu\text{L}$  10x RT buffer, 3  $\mu\text{L}$  25 mM  $\text{MgCl}_2$ , 1.7  $\mu\text{L}$  0.1 M DTT, 0.5  $\mu\text{L}$  SUPERase-In, and 0.9  $\mu\text{L}$  Superscript III reverse transcriptase was added to the reverse transcription reaction and then incubated for 30 min at 48°C. After cDNA synthesis, 2  $\mu\text{L}$  exonuclease I (New England Biolabs) was added to the reaction and incubated for 30 min at 37°C. The enzyme was inactivated and RNA hydrolyzed by adding 1  $\mu\text{L}$  of 2 M NaOH and incubating for 20 min at 98°C. The reaction was then neutralized with 1  $\mu\text{L}$  2 M HCl. The cDNA was run on a 10% TBE-Urea gel (Invitrogen) and the gel was stained using SYBR gold (ThermoFisher Scientific). cDNA sized ~120–350 nucleotides were cut, gel purified, and precipitated overnight with ethanol and glycogen. Afterwards, cDNA was circularized by resuspending precipitated DNA in 10  $\mu\text{L}$  circularization mix (7.5  $\mu\text{L}$  of water, 1  $\mu\text{L}$  10x Reaction Buffer (Epicentre, Madison, WI), 0.5  $\mu\text{L}$  of 1 mM ATP (final 0.05 mM), 0.5  $\mu\text{L}$  of 50 mM  $\text{MnCl}_2$  (final 2.5 mM), 0.5  $\mu\text{L}$  CirLigase (100  $\mu\text{M}$ ), (Epicentre)). Circularization was performed for 1 hr at 60°C, and the reaction was heat-inactivated for 15 min at 85°C. Circular single-stranded DNA was re-linearized by adding 3.3  $\mu\text{L}$  of re-linearization mix (4x mix containing 100 mM KCl and 2 mM DTT) followed by 1  $\mu\text{L}$  of APE 1 (15  $\mu\text{M}$ ; New England Biolabs). The reaction was incubated for 45 min at 37°C; an additional 1  $\mu\text{L}$  APE 1 was added and the reaction was incubated for another 45 min. The enzyme was inactivated by incubating for 20 min at 65°C. The cDNA was amplified for 10–14 cycles using 0.1  $\mu\text{L}$  Phusion polymerase (New England Biolabs), 2  $\mu\text{L}$  5x HF buffer, 0.2  $\mu\text{L}$  10 mM dNTP, 1  $\mu\text{L}$  5 M betaine, 4.7  $\mu\text{L}$  water, and 0.5  $\mu\text{L}$  of the following 10  $\mu\text{M}$  primers: 5'-CAA GCA GAA GAC GGC ATA-3' and 5'-AAT GAT ACG GCG ACC ACC GAC AGG TTC AGA GTT CTA CAG TCC GACG-3'. The subsequent product was then gel purified from a 10% TBE gel (Invitrogen) using the ChIP DNA Clean & Concentrator Kit (Zymo Research Corporation, Irvine, CA).

For the following RNA-Seq samples: one replicate of no treatment 6 hr, one replicate of no treatment 24 hr, three replicates of polyinosinic-polycytidylic acid treatment 6 hr, two replicates of Pam3CSK4 treatment 6 hr, one replicate of Kdo2-lipid A treatment 6 hr, one replicate of IL4 treatment 24 hr, and one replicate of Kdo2-lipid A and interferon- $\gamma$  treatment 24 hr, strand-specific RNA

sequencing libraries were prepared from poly(A) mRNA using a method similar to that previously described ([Wang, 2011](#)) with modifications described herein. Briefly, poly(A) enriched mRNA was fragmented, in 2x Superscript III first-strand buffer with 10 mM DTT (Invitrogen), by incubation at 94°C for 9 min, then immediately chilled on ice before the next step. The 10 µL of fragmented mRNA, 0.5 µL of random primer (Invitrogen), 0.5 µL of Oligo dT primer (Invitrogen), 0.5 µL of SUPERase-In (Ambion), 1 µL of dNTPs (10 mM), and 1 µL of DTT (10 mM) were heated at 50°C for three minutes. At the end of incubation, 5.8 µL of water, 1 µL of DTT (100 mM), 0.1 µL Actinomycin D (2 µg/µL), 0.2 µL of 1% Tween-20 (Sigma), and 0.2 µL of Superscript III (Invitrogen) were added and incubated in a PCR machine using the following conditions: 25°C for 10 min, 50°C for 50 min, and a 4°C hold. The product was then purified with RNAClean XP beads according to manufacturer's instructions and eluted with 10 µL nuclease-free water. The RNA/cDNA double-stranded hybrid was then added to 1.5 µL of Blue Buffer (Enzymatics, Beverly, MA), 1.1 µL of dUTP mix (10 mM dATP, 10 mM dCTP, 10 mM dGTP, and 20 mM dUTP), 0.2 µL of RNase H (5 µg/µL), 1.05 µL of water, 1 µL of DNA polymerase I (Enzymatics), and 0.15 µL of 1% Tween-20. The mixture was incubated at 16°C for 1 hr. The resulting dUTP-marked dsDNA was purified using 28 µL of Sera-Mag Speedbeads (Thermo Fisher Scientific), diluted with 20% PEG8000, 2.5 M NaCl to final of 13% PEG, eluted with 40 µL EB buffer (10 mM Tris-HCl, pH 8.5), and frozen at -80°C. The purified dsDNA (40 µL) subsequently underwent end repair by blunting, poly(A)-tailing, and adapter ligation as described below.

All other RNA-Seq samples were prepared as described ([Heinz et al., 2013](#)). After RNA fragmentation and re-buffering with the Micro Bio-Spin P-30 column (Bio-Rad) according to the manufacturer's instructions, samples were resuspended with 16.5 µL of water. For de-capping using tobacco acid pyrophosphatase (TAP) (Epicentre), the following was added to the reaction: 2 µL 10x TAP buffer, 1 µL (20 µ) SUPERase-In (Ambion), 0.5 µL TAP; the reaction was then incubated for 2 hr at 37°C. Samples were then 3' de-phosphorylated using T4 polynucleotide kinase (New England Biolabs); 0.5 µL 10x TAP buffer, 1.5 µL water, 0.5 µL 0.25 M MgCl<sub>2</sub>, 0.5 µL 10 mM ATP, and 1 µL PNK was added to the reaction and incubated for 50 min at 37°C. After de-phosphorylation, samples were subsequently 5' phosphorylated using T4 polynucleotide kinase in order to facilitate subsequent adapter ligation processes; 10 µL 10x T4 DNA ligase buffer, 63 µL water, and 2 µL PNK was added to the reaction and incubated for 60 min at 37°C. TRIzol LS was used to quench the reaction and extract phosphorylated RNA. RNA was resuspended in 4.5 µL water. For indexed library preparation, the 3' adapter (0.5 µL 9 µM of a 5'-adenylated sRNA 3' MPX adapter /5Phos/AG ATC GGA AGA GCA CAC GTC TGA /3AmMO/ (Integrated DNA Technologies, San Jose, CA)) was heat-denatured together with the RNA for 2 min at 70°C, placed on ice, and ligated with 100 µ truncated T4 RNA ligase 2 (K227Q, New England Biolabs) in 10 µL 1x T4 RNA ligase buffer without ATP, containing 20 µg/µL SUPERase-In and 15% PEG8000 for 2 hr at 16°C. Afterwards, 0.5 µL 10 µM MPX\_RT primer 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3' (Integrated DNA Technologies, desalted) was added and annealed to the ligation product by incubating at 75°C for 2 min, then 37°C for 30 min, then 25°C for 15 min. To ligate the 5' adapter, 0.5 µL 5 µM hybrid DNA/RNA sRNA 5'h adapter 5'-GTT CAG AGT TCT ACA rGrUrC rCrGrA rCrGrA rUrC-3' (Integrated DNA Technologies) was ligated to the 5' end by adding 2 µL T4 RNA ligase buffer, 6 µL 50% PEG8000, 1 µL 10 mM ATP, 9.5 µL water, and 0.5 µL T4 RNA ligase 1 (New England Biolabs) for 90 min at 20°C. The reaction was then split in half (15 µL each) and 0.5 µL 10 µM MPX\_RT primer was added to one 15 µL reaction. The reactions were incubated at 70°C for 1 min, then placed on ice. Reverse transcription was performed by adding 3 µL 10x RT buffer, 4.5 µL water, 1.5 µL 10 mM dNTP, 3 µL 0.1 M DTT, 1.5 µL RNaseOUT, and 1 µL Superscript III reverse transcriptase, then incubating for 30 min at 50°C. The cDNA was amplified for 10–14 cycles using 0.5 µL Phusion polymerase, 10 µL 5x HF buffer, 1 µL 10 mM dNTP, 5 µL 5 M betaine, and 0.25 µL of the following 100 µM primers: 5'-AAT GAT ACG GCG ACC ACC GAC AGG TTC AGA GTT CTA CAG TCC GAC G-3' and TruSeq-compatible indexed primers (e.g. 5'-CAA GCA GAA GAC GGC ATA CGA GAT iii iii GTG ACT GGA GTT CAG ACG TGT GCT CTT-3' (desalted, Integrated DNA Technologies, i signifies index nucleotides)). The subsequent product was then size selected for 175–225 base pair product and gel purified from a 10% TBE gel (Invitrogen) using the ChIP DNA Clean & Concentrator Kit. Libraries were PCR-amplified for 9–14 cycles, size selected by gel extraction, and quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

## ChIP-Seq

Previously published Rev-erb $\alpha$  and Rev-erb $\beta$  ChIP-Seq, and NCoR ChIP-Seq experiments, deposited as GSE45914 ([Lam et al., 2013](#)) and GSE27060 ([Barish et al., 2012](#)), respectively, were utilized for analyses. Detailed protocols for ChIP-Seq experiments have been previously described ([Kaikkonen et al., 2013](#); [Heinz et al., 2010](#), [Heinz et al., 2013](#); [Li et al., 2013](#)). Antibodies against Fos (sc-7202), Nrf2 (sc-13032x), p65 (sc-372), and PU.1 (sc-352x) were purchased from Santa Cruz Biotechnology (Dallas, TX), against Smad3 (ab28379) from Abcam (Cambridge, UK), and against H3K27ac (39135) from Active Motif (Carlsbad, CA). Briefly, for Fos, Nrf2, p65, and Smad3 ChIPs, macrophages were first cross-linked in 2 mM disuccinimidyl glutarate (Pierce 20593, Thermo Fischer) in PBS for 30 min, followed by subsequent 1% formaldehyde (Sigma) cross-linking in PBS for 10 min at room temperature. For H3K27ac and PU.1 ChIPs, cells were cross-linked using 1% formaldehyde in PBS for 10 min at room temperature. After cross-linking, glycine (Sigma) was added to a final concentration of 0.2625 M to quench the reaction. Subsequently, cross-linked macrophages were centrifuged (5 min, 1,200 RPM, 4°C), washed twice with PBS, and pellets were snap frozen and stored at –80°C. For ChIP of H3K27Ac, p65, PU.1, Nrf2 or Smad3, frozen cell pellets were resuspended in cell lysis buffer (10 mM HEPES/KOH pH 7.9, 85 mM KCl, 1 mM EDTA, 1.0% IGEPAL CA-630 (Sigma), 1x protease inhibitor cocktail (Roche, Basel, Switzerland), 1 mM PMSF). After 5 min lysis on ice, cells were centrifuged (5 min, 4000 RPM, 4°C), and the supernatant was removed. The pellet was then resuspended in nuclear lysis buffer (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine, 1x protease inhibitor cocktail, and 1 mM PMSF) and the chromatin was sheared by sonication on wet ice with a Bioruptor Standard Sonicator (Diagenode, Denville, NJ) for three 15 min cycles each alternating 30 s on and 30 s off on the high setting. Additional Triton X-100 was added to the sonicated chromatin to 10% of the final volume and the lysate was cleared by centrifugation (5 min, 14,000 RPM, 4°C). Input was then saved for subsequent analysis.

For Fos ChIP, pellets were suspended in 50 mM Tris pH 8.0, 60 mM KCl, 0.5% IGEPAL, 1x protease inhibitor cocktail, and 1 mM PMSF, followed by 10 min of incubation on ice and centrifugation at 2000  $\times$ g for 3 min at 4°C. The pellet was then suspended in 0.5% SDS, 10 mM EDTA, 0.5 mM EGTA, 50 mM Tris pH 8.0, 1x protease inhibitor cocktail, and 1 mM PMSF. The chromatin suspension was sheared by sonication on wet ice with a Bioruptor Standard Sonicator for three 15 min cycles each alternating 30 s on and 30 s off on the high setting, followed by centrifugation for 10 min at 15,000 RPM at 4°C. The chromatin was diluted 5x with 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris pH 8.0, 1x protease inhibitor cocktail, and 1 mM PMSF. An input sample was saved for subsequent analysis.

Protein A or G Dynabeads (Invitrogen) pre-bound with antibody was added to the diluted cell lysate overnight at 4°C. Immunoprecipitated complexes were washed three times with 20 mM Tris/HCl pH 7.4/150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, three times with 10 mM Tris/HCl pH 7.4/250 mM LiCl, 1% Triton X-100, 1% sodium deoxycholate, 1 mM EDTA, and two times with Tris-EDTA plus 0.1% Tween-20 before eluting two times with 50  $\mu$ L elution buffer (TE, 1% SDS, 30 and 10 min, room temperature). Elution buffer was also added to the input. After pooling the eluted samples, the sodium concentration was adjusted to 300 mM and cross-links were reversed overnight at 65°C. Samples were treated with 0.5 mg/ml proteinase K for 1 hr at 55°C and 0.25 mg/ml RNase A for 1 hr at 37°C before DNA was isolated using the ChIP DNA Clean and Concentrator Kit according to the manufacturer's instructions. For library preparation, NEXTflex DNA barcode adaptors (BioO Scientific, Austin, TX) were ligated to the genomic DNA. Polymerase chain reaction mediated library amplification was performed and final libraries were size selected on 10% TBE gels (Invitrogen).

## High-throughput sequencing and data processing

RNA-Seq and ChIP-Seq libraries were sequenced for 50 cycles on an Illumina Hi-Seq 2000 (Illumina, San Diego, CA), sequenced for 51 cycles on an Illumina Hi-Seq 4000, or sequenced for 51 cycles on an Illumina NextSeq 2500 according to the manufacturer's instructions. ChIP-Seq reads were mapped to the mouse NCBI37/mm9 ([Ferreyra Garrott et al., 2013](#)) assembly using Bowtie ([Langmead et al., 2009](#)), allowing up to two mismatches. RNA-Seq reads were mapped to the mouse NCBI37/mm9 ([Ferreyra Garrott et al., 2013](#)) assembly using Tophat ([Trapnell et al., 2009](#)). Mapped reads were visualized using the UCSC genome browser ([Kent et al., 2002](#)) and

downstream data processing was performed using HOMER (Heinz *et al.*, 2010), and R (García-Oltra *et al.*, 2013).

### Genome-wide gene expression analysis with RNA-Seq

RNA-Seq analysis of genome-wide gene expression was performed using HOMER (Heinz *et al.*, 2010). Differential expression was defined by a fold-change of at least 1.5-fold averaging over replicated datasets. For heatmap analysis, genes were clustered using k-means clustering ( $k = 10$ ) in R. Gene ontology analysis was performed using DAVID Bioinformatics Resources 6.7 (Huang *et al.*, 2009a, 2009b).

### ChIP-Seq analysis

Genomic histone acetylation regions and transcription factor peaks were determined with HOMER using the findPeaks command default parameters of four-fold enrichment over the input, four-fold enrichment over local background, and normalization to 10 million mapped reads. For transcription factors, peaks were called using the '-style factor' parameter while histone acetylation regions were called using the '-style region' parameter. Histone regions were centered on nucleosome free regions using the '-nfr' parameter. For comparisons, called peaks from different data sets were merged using the mergePeaks command. Merging of transcription factor peaks or histone regions was done using the parameter '-size given'. To obtain differentially bound peaks/regions, tags were quantified from two data sets using the getDifferentialPeaks command. Peaks/regions were called as differentially gained if they had a four-fold enrichment of tag counts over the untreated/vehicle condition and a cumulative Poisson p-value less than 0.001. For heatmap analysis, peaks were clustered using hierarchical clustering in R.

### Bone marrow transplantation

Bone marrow harvested from WT and Rev-erb DKO mice was injected via the retro-orbital route into lethally irradiated (10 Gy) B6.SJL-Ptprc<sup>a</sup> Pepc<sup>b</sup>/BoyJ (CD45.1) (Jackson Lab, Sacramento, CA) or C57BL/6J (Harlem (now Envigo), Indianapolis, IN) 8 week old wild type congenic mice. Approximately 6–7 million bone marrow cells were injected per mouse. Transplanted mice were housed in autoclaved cages (changed every two days) and supplemented with antibiotics the day before irradiation until two weeks post-transplantation.

### Evaluating bone marrow transplant efficiency

To evaluate bone marrow transplant efficiency, whole blood from WT and Rev-erb DKO bone marrow transplanted mice was collected through cardiac puncture into EDTA tubes (Becton Dickinson, Franklin Lakes, NJ). 100  $\mu$ L whole blood was washed once with PBS and resuspended in 2% FBS in PBS. Samples were blocked with 1  $\mu$ L anti-mouse CD16/32 (eBioscience, San Diego, CA, 14-0161-82) for 15 min at room temperature. The following antibodies were utilized for staining: CD45 (Biolegend, San Diego, CA, 103122) and CD45.2 (Biolegend, 109813). Samples were incubated with directly labeled antibodies for 40 min (4°C in the dark). Stained cells were washed with 0.1% BSA in PBS, pelleted (1200 RPM, 5 min, 4°C), and lysed with hemolysin (Beckman Coulter, Brea, CA) for 20 s. Samples were quenched with 10x PBS, diluted, and gently washed before analysis using a LSR II flow cytometer (BD Bioscience, San Jose, CA). Unstained and single stains were used for setting up compensations and gating. Events were first gated on forward and side scatter to determine single events, before evaluation of other fluorescent markers.

### Monocyte enumeration from peripheral blood

Blood was collected from 16 chimeric mice per genotype into 0.5 ml K3 EDTA coated tubes. The volume of the blood was determined by pipetting and transferred to 5 ml round bottom tubes with 50  $\mu$ l of Life Technologies (Carlsbad, CA) CountBright Absolute Counting Beads. Erythrocytes were lysed by addition of 4 ml eBioscience RBC lysis buffer with incubation at 4°C for 5 mins. Cells were collected by centrifugation and the supernatant was carefully removed. Cells were washed once more and resuspended in buffer containing anti-CD16/CD32 (clone 93, BioLegend) and Zombie Aqua fixable viability dye (BioLegend). After 10 min, cells were stained with the following 2X antibody cocktail: anti-mouse CD11b BD Horizon BUV395 (clone M1/70, BD Biosciences), anti-mouse

CD19 BD Horizon BUV737 (clone1D3, BD Biosciences), anti-mouse CD115 Brilliant Violet 421 (clone AFS98, BioLegend), anti-mouse CD90.2 Brilliant Violet 785 (clone 30-H12, BioLegend), anti-mouse Ly6G FITC (clone 1A8, BioLegend), anti-mouse CD45.2 PE (clone 104, BioLegend), anti-mouse CD45.1 Alexa Fluor 647 (clone A20, BioLegend), and anti-mouse Ly6C APC/Cy7 (clone HK1.4, BioLegend). After 20 min, cells were washed and counted on a Beckman Colter MoFlo Astrios EQ equipped with 355 nm, 405 nm, 488 nm, 561 nm and 640 nm lasers. Cells per  $\mu\text{l}$  were determined by following the manufacturer protocol for CountBright Absolute Counting Beads. Cells of interest were identified by excluding Zombie Aqua that fell within consecutive singlet gates using SSC and FSC. Donor derived monocytes were identified as CD45.2+, CD19-, CD90.2-, CD115+, Ly6G-, and CD11b+. Monocytes were further segregated based on expression of Ly6C. To test the dependence on hematopoietic derived expression of Rev-erb  $\alpha/\beta$  on peripheral blood cell populations we used a Welch two sample t-test using R.

## Wound healing studies

Wound healing studies were conducted 6–10 weeks post-transplantation. Briefly, 15.5 mL tert-amyl alcohol was added to 25 grams of 2,2,2 tribromoethanol (Sigma Aldrich Chemical) and dissolved overnight in a dark bottle to generate a stock solution. The subsequent solution was diluted with PBS, dissolved overnight, and filtered through a 0.2  $\mu\text{m}$  filter to generate a working solution (20 mg/ml). To achieve anesthesia, 0.4–0.75 mg/g was administered intra-peritoneally. A 3 mm punch biopsy (Miltex, York, PA) was used to generate four wounds on the dorsal skin of each animal. Wounds were systematically photographed from a fixed distance daily. For macroscopic analysis, genotypes were blinded and the size of the wound was analyzed by Adobe Photoshop (San Jose, CA), and normalized to its size on Day 0. Mice were housed singularly throughout the duration of the study.

To assess the contribution of Rev-erb to wound healing, data from three independent experiments were combined and analyzed using a linear mixed effects model (*García-Gil et al., 2012*) using the R package 'nlme' (R script: wound model <- lme('wound size' ~ 'genotype' \* 'time point', random=~1 | 'independent experiment' /'independent mouse' /'nested observation', data=data.file, na.action='na.exclude'). Genotype, time point, and their interaction, were modeled as fixed effects, whereas the observations at wound sites were treated as a random effect nested within the independent mouse, which in turn was treated as a random effect nested within the independent experiment, to account for the hierarchical nature of the study design. The numbers of biologically independent mice per time point are summarized in the below table.

| Day | WT chimera | Rev-erb DKO chimera |
|-----|------------|---------------------|
| 0   | 29         | 28                  |
| 1   | 28         | 27                  |
| 2   | 26         | 25                  |
| 3   | 17         | 17                  |
| 4   | 21         | 20                  |
| 5   | 22         | 20                  |
| 6   | 22         | 20                  |
| 7   | 13         | 12                  |
| 8   | 20         | 16                  |
| 9   | 12         | 9                   |
| 10  | 7          | 7                   |
| 11  | 12         | 9                   |
| 12  | 7          | 7                   |

## Histological analyses

At the indicated time points, mice were euthanized and wounds were harvested using a 6 mm punch biopsy (Miltex). Harvested wounds were cut along the mid-sagittal plane and paraffin-embedded.

Genotypes were blinded for subsequent histological analyses. The first section along the mid-sagittal plane was utilized for hematoxylin and eosin staining. Subsequent sections were utilized for immunohistochemical analysis using the following primary antibodies: biotinylated anti-F4/80 (AbD Serotec (now Bio-Rad), MCA4978, 1:50 dilution), IgG (Dako, Glostrup Municipality, Denmark), rat anti-Ly6B.2 (AbD Serotec (now Bio-Rad), MCA771GA, 1:200), and the following secondary antibodies: biotinylated anti-rat (1:500, BD), as well as HRP-conjugated streptavidin (1:500, Jackson Laboratory). Briefly, slides were de-paraffinized and washed three times in 0.1% Tween-20 PBS. Blocking was performed sequentially using 3% hydrogen peroxide (10 min), 1% BSA in 0.1% Tween-20 PBS (10 min), 0.1% avidin (10 min), and 0.01% biotin (10 min). Three washes were performed between each blocking step using 0.1% Tween-20 PBS. Antigen retrieval was performed using proteinase K (Dako, S3020), followed by three washes and subsequent overnight incubation with the indicated primary antibodies. After three washes, slides were incubated with the indicated secondary antibodies for 30 min and developed using AEC Peroxidase Substrate Kit (Vector Labs, Burlingame, CA, SK-4200) according to the manufacturer's instructions. Counterstaining was performed using Mayer's Hematoxylin (Sigma, MHS16), after which samples were mounted in an aqueous gel mount (Vectamount, Vector Labs, H-5501).

### In vitro matrigel migration assays

In vitro matrigel migration assays were performed as previously described (Ogawa *et al.*, 2004). Briefly, macrophages were cultured in RPMI-1640 (Invitrogen) supplemented with 0.5% heat-inactivated FBS (Hyclone) for 24 hr and resuspended at a density of 1 million cells per milliliter. 100  $\mu$ L of macrophages was added to the top chamber of a transwell (Corning, Corning, NY) while 650  $\mu$ L of media was added to the bottom chamber. Macrophages were allowed to migrate through basement membrane extract (Corning, 3458) for 24 hr. Afterwards, the wells were briefly washed with PBS, and migrated macrophages were dissociated from the membrane and incubated with Calcein AM. Relative fluorescence was measured using a SpectraMax M3 plate reader (Molecular Devices, Sunnyvale, CA) and the SoftMax Pro software (485 nm excitation, 520 nm emission). A standard curve was used to convert relative fluorescence to cell numbers.

### Sequencing data

All sequencing data used in this manuscript has been submitted to GEO under the accession GSE72964. This data can be accessed by reviewers through the following link: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=ejixaiswxlqnjiv&acc=GSE72964>.

### Acknowledgements

We would like to thank Martina P Pasillas for assistance with flow cytometry and Liwen Deng at the UCSD Mouse Phenotyping Core for preparation of histological slides and immunohistochemical staining. We would also like to thank Lynn Bautista and Leslie Van Ael for their assistance with figure and manuscript preparation, as well as Dr. Thomas Burris for his kind gift of the SR-9009 Rev-erb ligand. These studies were supported by NIH Grants DK091183, DK063491, CA173903, and GM085764. DZE was supported by NIH training grant (T32-GM007198), and an American Heart Association Predoctoral Fellowship (12PRE11610007). TDT was supported by the National Cancer Institute of the National Institutes of Health under Award Number T32CA009523. HPL was supported by the Finnish Cultural Foundation, the North Savo Regional Fund, the Finnish Foundation for Cardiovascular Research, and the Maud Kuistila Memorial Foundation.

---

### Additional information

#### Funding

| Funder                        | Grant reference number | Author             |
|-------------------------------|------------------------|--------------------|
| National Institutes of Health | T32-GM007198           | Dawn Z Eichenfield |
| American Heart Association    | 12PRE11610007          | Dawn Z Eichenfield |
| National Cancer Institute     | T32CA009523            | Ty Dale Troutman   |

|                               |  |                     |
|-------------------------------|--|---------------------|
| Suomen Kulttuurirahasto       | North Savo Regional Fund                           | Hanna P Lesch       |
| Pohjois-Savon Rahasto         |  | Hanna P Lesch       |
| Maud Kuistilan Muistosäätiö   |  | Hanna P Lesch       |
| Suomen Kulttuurirahasto       | The Finnish Foundation for Cardiovascular Research | Hanna P Lesch       |
| National Institutes of Health | DK091183   | Christopher K Glass |
| National Institutes of Health | DK063491   | Christopher K Glass |
| National Institutes of Health | CA173903   | Christopher K Glass |
| National Institutes of Health | GM085764   | Christopher K Glass |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

DZE, CKG, Designed the study, Performed experiments, Analyzed and interpreted the data; TDT, Performed experiments, Analyzed and interpreted the data; VML, Performed experiments, Provided essential mouse models; MTL, Performed experiments; HC, Performed experiments, Developed RNA-Seq methods, Analysis and interpretation of data; DG, Designed the study, Performed experiments; NJS, HPL, Analyzed and interpreted the data; JT, Designed the study, Analyzed and interpreted the data; JM, Analyzed and interpreted the data, Provided essential mouse models; RLG, Designed the study, Analyzed and interpreted the data, Wrote the manuscript; RME, Designed the study, Performed experiments, Analyzed and interpreted the data, Wrote the manuscript

### Author ORCIDs

Dawn Z Eichenfield, <http://orcid.org/0000-0002-1511-3804>

Ty Dale Troutman, <http://orcid.org/0000-0001-8925-8080>

Verena M Link, <http://orcid.org/0000-0002-3207-312X>

Christopher K Glass, <http://orcid.org/0000-0003-4344-3592>

### Ethics

Animal experimentation: All animal procedures were in accordance with the University of California, San Diego research guidelines for the care and use of laboratory animals (Permit Number: S01015).

## Additional files

### Major datasets

The following dataset was generated:

| Author(s)  | Year | Dataset title  | Dataset URL   | Database, license, and accessibility information                                |
|--|------|--|---|---|
| Eichenfield DZ, Troutman TD, Link VM, Lam MT, Cho H, Gosselin D, Spann NJ, Lesch HP, Tao J, Muto J, Gallo RL, Evans RM, Glass CK | 2016 | Tissue damage signals drive co-localization of NF- $\kappa$ B, Smad3, and Nrf2 to direct a Rev-erb sensitive wound repair program in macrophages | <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72964">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72964</a> | Publicly available at the NCBI Gene Expression Omnibus (Accession no: GSE72964) |

The following previously published datasets were used:

| Author(s)   | Year | Dataset title   | Dataset URL   | Database, license, and accessibility information                                |
|---|------|---|---|---|
| Lam MT, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M, Lee CY, Watt A, Grossman TR, Rosenfeld MG, Evans RM, Glass CK     | 2013 | Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription | <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45914">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45914</a> | Publicly available at the NCBI Gene Expression Omnibus (Accession no: GSE45914) |
| Barish GD, Yu RT, Karunasiri MS, Becerra D, Kim J, Tseng TW, Tai LJ, Leblanc M, Diehl C, Cerchiatti L, Miller YI, Witztum JL, Melnick AM, Dent AL, Tangirala RK, Evans RM | 2012 | A Bcl6-Smrt/Ncor repression program controls atherosclerosis                              | <a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27060">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27060</a> | Publicly available at the NCBI Gene Expression Omnibus (Accession no: GSE27060) |

## References

- Barish GD**, Yu RT, Karunasiri MS, Becerra D, Kim J, Tseng TW, Tai LJ, Leblanc M, Diehl C, Cerchiatti L, Miller YI, Witztum JL, Melnick AM, Dent AL, Tangirala RK, Evans RM. 2012. The Bcl6-SMRT/NCOR cistrome represses inflammation to attenuate atherosclerosis. *Cell Metabolism* **15**:554–562. doi: [10.1016/j.cmet.2012.02.012](https://doi.org/10.1016/j.cmet.2012.02.012)
- Campbell L**, Saville CR, Murray PJ, Cruickshank SM, Hardman MJ. 2013. Local arginase 1 activity is required for cutaneous wound healing. *Journal of Investigative Dermatology* **133**:2461–2470. doi: [10.1038/jid.2013.164](https://doi.org/10.1038/jid.2013.164)
- Chen L**, Guo S, Ranzer MJ, DiPietro LA. 2013. Toll-like receptor 4 has an essential role in early skin wound healing. *Journal of Investigative Dermatology* **133**:258–267. doi: [10.1038/jid.2012.267](https://doi.org/10.1038/jid.2012.267)
- Chen B**, Lu Y, Chen Y, Cheng J. 2015. The role of Nrf2 in oxidative stress-induced endothelial injuries. *Journal of Endocrinology* **225**:R83–99. doi: [10.1530/JOE-14-0662](https://doi.org/10.1530/JOE-14-0662)
- Cho H**, Zhao X, Hatori M, Yu RT, Barish GD, Lam MT, Chong LW, DiTacchio L, Atkins AR, Glass CK, Liddle C, Auwerx J, Downes M, Panda S, Evans RM. 2012. Regulation of circadian behaviour and metabolism by REV-ERB- $\alpha$  and REV-ERB- $\beta$ . *Nature* **485**:123–127. doi: [10.1038/nature11048](https://doi.org/10.1038/nature11048)
- Creighton MP**, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**:21931–21936. doi: [10.1073/pnas.1016071107](https://doi.org/10.1073/pnas.1016071107)
- Durand B**, Saunders M, Gaudon C, Roy B, Losson R, Chambon P. 1994. Activation function 2 (AF-2) of retinoic acid receptor and 9-cis retinoic acid receptor: presence of a conserved autonomous constitutive activating domain and influence of the nature of the response element on AF-2 activity. *The EMBO Journal* **13**:5370–5382.
- Escoubet-Lozach L**, Benner C, Kaikkonen MU, Lozach J, Heinz S, Spann NJ, Crotti A, Stender J, Ghisletti S, Reichart D, Cheng CS, Luna R, Ludka C, Sasik R, Garcia-Bassets I, Hoffmann A, Subramaniam S, Hardiman G, Rosenfeld MG, Glass CK. 2011. Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. *PLOS Genetics* **7**:e1002401. doi: [10.1371/journal.pgen.1002401](https://doi.org/10.1371/journal.pgen.1002401)
- Falanga V**. 2005. Wound healing and its impairment in the diabetic foot. *The Lancet* **366**:1736–1743. doi: [10.1016/S0140-6736\(05\)67700-8](https://doi.org/10.1016/S0140-6736(05)67700-8)
- Feng D**, Liu T, Sun Z, Bugge A, Mullican SE, Alenghat T, Liu XS, Lazar MA. 2011. A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* **331**:1315–1319. doi: [10.1126/science.1198125](https://doi.org/10.1126/science.1198125)
- Ferreira Garrott LG**, Soriano ER, Rosa JE, Navarta DA, Saucedo C, Scolnik M, Bedran Z, Sabelli M, García MV, Anselmi C, Galimberti R, Catoggio LJ, Husni ME, Qureshi AA. 2013. Validation in Spanish of a screening questionnaire for the detection of psoriatic arthritis in patients with psoriasis. *Rheumatology* **52**:510–514. doi: [10.1093/rheumatology/kes306](https://doi.org/10.1093/rheumatology/kes306)
- Font-Vizcarra L**, García S, Bori G, Martínez-Pastor JC, Zumbado A, Morata L, Mensa J, Soriano A. 2012. Long-term results of acute prosthetic joint infection treated with debridement and prosthesis retention: a case-control study. *The International Journal of Artificial Organs* **35**:908–912. doi: [10.5301/ijao.5000147](https://doi.org/10.5301/ijao.5000147)
- Fontaine C**, Rigamonti E, Pourcet B, Duez H, Duhem C, Fruchart JC, Chinetti-Gbaguidi G, Staels B. 2008. The nuclear receptor Rev-erb $\alpha$  is a liver X receptor (LXR) target gene driving a negative feedback loop on select LXR-induced pathways in human macrophages. *Molecular Endocrinology* **22**:1797–1811. doi: [10.1210/me.2007-0439](https://doi.org/10.1210/me.2007-0439)

- García-Gil FA**, Albendea CD, López-Pingarrón L, Royo-Dachary P, Martínez-Guillén J, Piedrafita E, Martínez-Díez M, Soria J, García JJ. 2012. Altered cellular membrane fluidity levels and lipid peroxidation during experimental pancreas transplantation. *Journal of Bioenergetics and Biomembranes* **44**:571–577. doi: [10.1007/s10863-012-9459-7](https://doi.org/10.1007/s10863-012-9459-7)
- García-Oltra E**, Bori G, Tomas X, Gallart X, Garcia S, Soriano A. 2013. Radiological evaluation of acetabular erosion after antibiotic-impregnated polymethylmethacrylate spacer (Spacer-G). *The Journal of Arthroplasty* **28**:1021–1024. doi: [10.1016/j.arth.2012.07.013](https://doi.org/10.1016/j.arth.2012.07.013)
- Gibbs JE**, Blaikley J, Beesley S, Matthews L, Simpson KD, Boyce SH, Farrow SN, Else KJ, Singh D, Ray DW, Loudon ASI. 2012. The nuclear receptor REV-ERB mediates circadian regulation of innate immunity through selective regulation of inflammatory cytokines. *Proceedings of the National Academy of Sciences of the United States of America* **109**:582–587. doi: [10.1073/pnas.1106750109](https://doi.org/10.1073/pnas.1106750109)
- Ginhoux F**, Schultze JL, Murray PJ, Ochando J, Biswas SK. 2015. New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nature Immunology* **17**:34–40. doi: [10.1038/ni.3324](https://doi.org/10.1038/ni.3324)
- Gorriani C**, Harris IS, Mak TW. 2013. Modulation of oxidative stress as an anticancer strategy. *Nature Reviews Drug Discovery* **12**:931–947. doi: [10.1038/nrd4002](https://doi.org/10.1038/nrd4002)
- Gosselin D**, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, Stender JD, Chun HB, Garner H, Geissmann F, Glass CK. 2014. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**:1327–1340. doi: [10.1016/j.cell.2014.11.023](https://doi.org/10.1016/j.cell.2014.11.023)
- Heinz S**, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**:576–589. doi: [10.1016/j.molcel.2010.05.004](https://doi.org/10.1016/j.molcel.2010.05.004)
- Heinz S**, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD, Glass CK. 2013. Effect of natural genetic variation on enhancer selection and function. *Nature* **503**:487–492. doi: [10.1038/nature12615](https://doi.org/10.1038/nature12615)
- Huang da W**, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**:1–13. doi: [10.1093/nar/gkn923](https://doi.org/10.1093/nar/gkn923)
- Huang da W**, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**:44–57. doi: [10.1038/nprot.2008.211](https://doi.org/10.1038/nprot.2008.211)
- Ishida Y**, Gao JL, Murphy PM. 2008. Chemokine receptor CX3CR1 mediates skin wound healing by promoting macrophage and fibroblast accumulation and function. *Journal of Immunology* **180**:569–579. doi: [10.4049/jimmunol.180.1.569](https://doi.org/10.4049/jimmunol.180.1.569)
- Kaikkonen MU**, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, Glass CK. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular Cell* **51**:310–325. doi: [10.1016/j.molcel.2013.07.010](https://doi.org/10.1016/j.molcel.2013.07.010)
- Kent WJ**, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Research* **12**:996–1006. doi: [10.1101/gr.229102](https://doi.org/10.1101/gr.229102)
- Lam MT**, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M, Lee CY, Watt A, Grossman TR, Rosenfeld MG, Evans RM, Glass CK. 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**:511–515. doi: [10.1038/nature12209](https://doi.org/10.1038/nature12209)
- Langmead B**, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25)
- Lavin Y**, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, Jung S, Amit I. 2014. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**:1312–1326. doi: [10.1016/j.cell.2014.11.018](https://doi.org/10.1016/j.cell.2014.11.018)
- Lawrence T**, Natoli G. 2011. Transcriptional regulation of macrophage polarization: enabling diversity with identity. *Nature Reviews Immunology* **11**:750–761. doi: [10.1038/nri3088](https://doi.org/10.1038/nri3088)
- Le Martelot G**, Claudel T, Gatfield D, Schaad O, Kornmann B, Lo Sasso G, Moschetta A, Schibler U. 2009. REV-ERBalpha participates in circadian SREBP signaling and bile acid homeostasis. *PLOS Biology* **7**:e1000181. doi: [10.1371/journal.pbio.1000181](https://doi.org/10.1371/journal.pbio.1000181)
- Lefterova MI**, Steger DJ, Zhuo D, Qatanani M, Mullican SE, Tuteja G, Manduchi E, Grant GR, Lazar MA. 2010. Cell-specific determinants of peroxisome proliferator-activated receptor gamma function in adipocytes and macrophages. *Molecular and Cellular Biology* **30**:2078–2089. doi: [10.1128/MCB.01651-09](https://doi.org/10.1128/MCB.01651-09)
- Lewis SL**, Sonke B, Sunderland T, Begne SK, Lopez-Gonzalez G, van der Heijden GMF, Phillips OL, Affum-Baffoe K, Baker TR, Banin L, Bastin J-F, Beeckman H, Boeckx P, Bogaert J, De Canniere C, Chezeaux E, Clark CJ, Collins M, Djagbletey G, Djuikouo MNK, et al. 2013. Above-ground biomass and structure of 260 African tropical forests. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**:20120295. doi: [10.1098/rstb.2012.0295](https://doi.org/10.1098/rstb.2012.0295)
- Li MO**, Wan YY, Sanjabi S, Robertson AK, Flavell RA. 2006. Transforming growth factor-beta regulation of immune responses. *Annual Review of Immunology* **24**:99–146. doi: [10.1146/annurev.immunol.24.021605.090737](https://doi.org/10.1146/annurev.immunol.24.021605.090737)
- Li P**, Spann NJ, Kaikkonen MU, Lu M, Oh da Y, Fox JN, Bandyopadhyay G, Talukdar S, Xu J, Lagakos WS, Patsouris D, Armando A, Quehenberger O, Dennis EA, Watkins SM, Auwerx J, Glass CK, Olefsky JM. 2013. NCoR repression of LXRs restricts macrophage biosynthesis of insulin-sensitizing omega 3 fatty acids. *Cell* **155**:200–214. doi: [10.1016/j.cell.2013.08.054](https://doi.org/10.1016/j.cell.2013.08.054)
- Lin Q**, Fang D, Fang J, Ren X, Yang X, Wen F, Su SB. 2011. Impaired wound healing with defective expression of chemokines and recruitment of myeloid cells in TLR3-deficient mice. *Journal of Immunology* **186**:3710–3717. doi: [10.4049/jimmunol.1003007](https://doi.org/10.4049/jimmunol.1003007)

- Liu AC, Tran HG, Zhang EE, Priest AA, Welsh DK, Kay SA. 2008. Redundant function of REV-ERB $\alpha$  and  $\beta$  and non-essential role for Bmal1 cycling in transcriptional regulation of intracellular circadian rhythms. *PLoS Genetics* **4**:e1000023. doi: [10.1371/journal.pgen.1000023](https://doi.org/10.1371/journal.pgen.1000023)
- Massagué J. 2012. TGF $\beta$  signalling in context. *Nature Reviews Molecular Cell Biology* **13**:616–630. doi: [10.1038/nrm3434](https://doi.org/10.1038/nrm3434)
- Medzhitov R, Horng T. 2009. Transcriptional control of the inflammatory response. *Nature Reviews Immunology* **9**:692–703. doi: [10.1038/nri2634](https://doi.org/10.1038/nri2634)
- Mullen AC, Orlando DA, Newman JJ, Lovén J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, DeKoter RP, Young RA. 2011. Master transcription factors determine cell-type-specific responses to TGF- $\beta$  signaling. *Cell* **147**:565–576. doi: [10.1016/j.cell.2011.08.050](https://doi.org/10.1016/j.cell.2011.08.050)
- Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdt S, Gordon S, Hamilton JA, Ivashkiv LB, Lawrence T, Locati M, Mantovani A, Martinez FO, Mege JL, Mosser DM, Natoli G, Saeij JP, Schultze JL, Shirey KA, Sica A, et al. 2014. Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity* **41**:14–20. doi: [10.1016/j.immuni.2014.06.008](https://doi.org/10.1016/j.immuni.2014.06.008)
- Novak ML, Koh TJ. 2013. Macrophage phenotypes during tissue repair. *Journal of Leukocyte Biology* **93**:875–881. doi: [10.1189/jlb.1012512](https://doi.org/10.1189/jlb.1012512)
- Ogawa S, Lozach J, Jepsen K, Sawka-Verhelle D, Perissi V, Sasik R, Rose DW, Johnson RS, Rosenfeld MG, Glass CK. 2004. A nuclear receptor corepressor transcriptional checkpoint controlling activator protein 1-dependent gene networks required for macrophage activation. *Proceedings of the National Academy of Sciences of the United States of America* **101**:14461–14466. doi: [10.1073/pnas.0405786101](https://doi.org/10.1073/pnas.0405786101)
- Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**:157–171. doi: [10.1016/j.cell.2012.12.018](https://doi.org/10.1016/j.cell.2012.12.018)
- Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U. 2002. The orphan nuclear receptor REV-ERB $\alpha$  controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**:251–260. doi: [10.1016/S0092-8674\(02\)00825-5](https://doi.org/10.1016/S0092-8674(02)00825-5)
- Raetz CR, Garrett TA, Reynolds CM, Shaw WA, Moore JD, Smith DC, Ribeiro AA, Murphy RC, Ulevitch RJ, Fearn C, Reichart D, Glass CK, Benner C, Subramaniam S, Harkewicz R, Bowers-Gentry RC, Buczynski MW, Cooper JA, Deems RA, Dennis EA. 2006. Kdo2-Lipid A of *Escherichia coli*, a defined endotoxin that activates macrophages via TLR-4. *Journal of Lipid Research* **47**:1097–1111. doi: [10.1194/jlr.M600027-JLR200](https://doi.org/10.1194/jlr.M600027-JLR200)
- Raspé E, Duez H, Mansén A, Fontaine C, Fiévet C, Fruchart JC, Vennström B, Staels B. 2002. Identification of Rev-erbalpha as a physiological repressor of apoC-III gene transcription. *Journal of Lipid Research* **43**:2172–2179. doi: [10.1194/jlr.M200386-JLR200](https://doi.org/10.1194/jlr.M200386-JLR200)
- Schuppan D, Kim YO. 2013. Evolving therapies for liver fibrosis. *Journal of Clinical Investigation* **123**:1887–1901. doi: [10.1172/JCI66028](https://doi.org/10.1172/JCI66028)
- Sen CK, Gordillo GM, Roy S, Kirsner R, Lambert L, Hunt TK, Gottrup F, Gurtner GC, Longaker MT. 2009. Human skin wounds: a major and snowballing threat to public health and the economy. *Wound Repair and Regeneration* **17**:763–771. doi: [10.1111/j.1524-475X.2009.00543.x](https://doi.org/10.1111/j.1524-475X.2009.00543.x)
- Shibata N, Carlin AF, Spann NJ, Saijo K, Morello CS, McDonald JG, Romanoski CE, Maurya MR, Kaikkonen MU, Lam MT, Crotti A, Reichart D, Fox JN, Quehenberger O, Raetz CR, Sullards MC, Murphy RC, Merrill AH, Brown HA, Dennis EA, et al. 2013. 25-Hydroxycholesterol activates the integrated stress response to reprogram transcription and translation in macrophages. *Journal of Biological Chemistry* **288**:35812–35823. doi: [10.1074/jbc.M113.519637](https://doi.org/10.1074/jbc.M113.519637)
- Solt LA, Wang Y, Banerjee S, Hughes T, Kojetin DJ, Lundasen T, Shin Y, Liu J, Cameron MD, Noel R, Yoo SH, Takahashi JS, Butler AA, Kamenecka TM, Burris TP. 2012. Regulation of circadian behaviour and metabolism by synthetic REV-ERB agonists. *Nature* **485**:62–68. doi: [10.1038/nature11030](https://doi.org/10.1038/nature11030)
- Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK, Merrill AH, Murphy RC, Raetz CR, Russell DW, Subramaniam S. 2007. LMSD: LIPID MAPS structure database. *Nucleic Acids Research* **35**:D527–532. doi: [10.1093/nar/gkl838](https://doi.org/10.1093/nar/gkl838)
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**:1105–1111. doi: [10.1093/bioinformatics/btp120](https://doi.org/10.1093/bioinformatics/btp120)
- Van Dyken SJ, Locksley RM. 2013. Interleukin-4- and interleukin-13-mediated alternatively activated macrophages: roles in homeostasis and disease. *Annual Review of Immunology* **31**:317–343. doi: [10.1146/annurev-immunol-032712-095906](https://doi.org/10.1146/annurev-immunol-032712-095906)
- Wang L. 2011. A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PLoS One* **6**:e26426. doi: [10.1371/journal.pone.0026426](https://doi.org/10.1371/journal.pone.0026426)
- Yin L, Lazar MA. 2005. The orphan nuclear receptor Rev-erbalpha recruits the N-CoR/histone deacetylase 3 corepressor to regulate the circadian Bmal1 gene. *Molecular Endocrinology* **19**:1452–1459. doi: [10.1210/me.2005-0057](https://doi.org/10.1210/me.2005-0057)

# Chapter 3

## MARGE: Mutation Analysis of Regulatory Genomic Elements

Verena M. Link, Casey E. Romanoski, Dirk Metzler, and Christopher K. Glass

manuscript *bioRxiv*, 268839



## MARGE: Mutation Analysis for Regulatory Genomic Elements

Verena M. Link<sup>1,3</sup>, Casey E. Romanoski<sup>4</sup>, Dirk Metzler<sup>3</sup>, and Christopher K. Glass<sup>1,2,5</sup>

<sup>1</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, San Diego, United States

<sup>2</sup>Department of Medicine, University of California, San Diego, San Diego, United States

<sup>3</sup>Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilian Universität München, Planegg-Martinsried, Germany

<sup>4</sup>Department of Cellular and Molecular Medicine, University of Arizona, Tucson, United States

<sup>5</sup>Correspondence:

Office: 858-534-6011

Email: [ckg@ucsd.edu](mailto:ckg@ucsd.edu)

### Abstract

Cell-specific patterns of gene expression are determined by combinatorial actions of sequence-specific transcription factors at *cis*-regulatory elements. Studies indicate that relatively simple combinations of lineage-determining transcription factors (LDTFs) play dominant roles in the selection of enhancers that establish cell identities and functions. LDTFs require collaborative interactions with additional transcription factors to mediate enhancer function, but the identities of these factors are often unknown. We have shown that natural genetic variation between individuals has great utility for discovering collaborative transcription factors. Here, we introduce MARGE (**M**utation **A**nalysis of **R**egulatory **G**enomic **E**lements), the first publicly available suite of software tools that integrates genome-wide genetic variation with epigenetic data to identify collaborative transcription factor pairs. MARGE is optimized to work with chromatin accessibility assays (such as ATAC-seq or DNase I hypersensitivity), as well as transcription factor binding data collected by ChIP-seq. Herein, we provide investigators with rationale for each step in the MARGE pipeline and key differences for analysis of datasets with different experimental designs. We demonstrate the utility of MARGE using mouse peritoneal macrophages, liver cells, and human lymphoblastoid cells. MARGE provides a powerful tool to identify combinations of cell type-specific transcription factors while simultaneously interpreting functional effects of non-coding genetic variation.

### Introduction

Molecular mechanisms enabling cell-specific transcriptional responses to intra- and extra-cellular signals remain poorly understood. Genome-wide studies of most lineage-determining (LDTF) and signal-dependent transcription factors (SDTF) indicate that the vast majority of

their binding sites are in distal intra- and intergenic locations that frequently exhibit epigenomic features associated with enhancers (1-6) and are evolutionary well conserved (7-9). The complement of active *cis*-regulatory elements bound by LDTFs changes across cell types, whereas promoters stay the same. Therefore, these findings introduced the notion that enhancers are largely responsible for cell type-specific gene expression (10-12). The ENCODE consortium annotated epigenetic features associated with enhancers in several different cell lines, primary cells and tissues providing evidence for hundreds of thousands of such elements in the human genome (13), greatly exceeding the number of promoters.

Previous studies of macrophages and B cells provided the basis for a collaborative and hierarchical model (14-16). In this model, collaborative binding of two or more LDTFs opens up chromatin to establish enhancers (1), enabling cell-specific actions of broadly expressed SDTFs (17) (reviewed in (18)). The collaborative nature of LDTFs was further demonstrated by analysis of effects of genetic variation in macrophages provided by two inbred strains of mice (19).

Genome-wide association studies, or GWAS (20) have shown that most complex trait-associated genetic variation is located in non-gene/protein regions of the genome. Such non-coding variants have the potential to change conserved sequences recognized by LDTFs and thereby alter enhancer landscapes between different alleles. These differences could manifest between individuals (i.e., between individuals that are each homozygous for opposite alleles), or within an individual that is heterozygous for a functional enhancer variant. A straightforward mechanism by which enhancer function would be altered by genetic variation is where alleles alter the affinity of transcription factors to bind their motifs. Consistent with the enhancer model whereby transcription factors collaborate with each other to bind DNA motifs, reports have found that allelic variation that mutates DNA binding motifs reduces binding of the respective factor while at the same time reducing binding of collaborating factors within 100 base pairs (19,21,22). Since the DNA binding motif of the partner factor is not mutated, these examples demonstrate a coordinated action of transcription factors in accessing DNA. The implication for cell-specific gene regulation is that genetic variants altering collaborative factor binding at enhancers will only be functional in the appropriate cell type where the correct combinations of transcription factors are expressed. The practical implication of these observations is that the particular combinations of factors may be discovered with the general strategy in any cell type. In addition to the discovery of transcription factors, this method identifies the precise genomic loci where genetic variation has a functional role in factor binding that may influence higher order biological processes.

To facilitate discovery of novel collaborating transcription factors using the genetic variation approach, we developed MARGE (**M**utation **A**nalysis for **R**egulatory **G**enomic **E**lements). MARGE is a suite of software tools to analyze ChIP-seq, ATAC-seq, DNase I Hypersensitivity or other next generation sequencing (NGS) assays where genotyping or DNA sequence data is available.

MARGE requires two data types: 1) genetic variation, and 2) high-throughput sequencing data (ChIP-seq, ATAC-seq, DNase-seq). MARGE then integrates these data and provides visualization tools to interpret the results. Importantly, MARGE was built to test for functional effects of alternate alleles at single nucleotide polymorphisms (SNPs) as well as short insertion-deletions (InDels). MARGE performs traditional de-novo motif analysis on genomic sequence for each polymorphic allele to identify DNA binding motifs that potentially affect transcription factor binding based on sequence analysis alone. The next step is to test whether the set of potential variants that mutate a single DNA binding motif are enriched in a set of loci where differential binding/accessibility is observed. For this step, MARGE associates quantitative measures of binding or accessibility from the ChIP/ATAC/DNase-seq data with the list of potential mutations in motifs. It analyzes differences in two genotypes by comparing the transcription factor binding distribution in relation to motif mutations between both genotypes, and also takes advantages of a Linear Mixed Model (LMM) (23,24) to compare many different individuals at the same time

In this report, we apply MARGE and demonstrate its ability to reliably identify known key regulators of macrophage lineage. We further apply MARGE to three different ChIP-seq datasets from mouse liver cells and also show that MARGE can identify important B cell factors in a human PU.1 ChIP-seq dataset from lymphoblastoid cell lines (25). In conclusion, MARGE is the first publicly available tool that is created to identify combinations of collaborating transcription factors. This approach is agnostic to cell type and can be applied in any dataset where genotypes and epigenetic signatures are measured.

MARGE is based on the ChIP-seq analysis tool HOMER (1) (<http://homer.ucsd.edu/homer/>) and it is an extension to the software used in (19,21). The source code and installation package are freely available on GitHub:

[https://github.com/vlink/marge/blob/master/MARGE\\_v1.0.tar.gz](https://github.com/vlink/marge/blob/master/MARGE_v1.0.tar.gz)

## **MATERIAL AND METHODS**

### **Overview**

A schematic outlining the major steps of MARGE is shown in Fig 1. First, MARGE offers a complete pipeline to process VCF (Variant Call Format) files (26) and generate individualized diploid genomes by extrapolating genetic variants from VCF files and swapping in alternate alleles into a reference genome (Fig. 1a-b). Importantly, MARGE is able to analyze sequencing data from homozygous (e.g., inbred mouse strains) and heterozygous (e.g., human) genomes and includes analysis for Single Nucleotide Polymorphisms (SNPs) as well as short Insertion-Deletions (InDels). Because MARGE generates genomes for each individual in the VCF file, the investigator can map their sequencing data to the genome with all genetic variations by using user-defined mapping software (e.g. bowtie2 (27) or STAR (28)) (Fig. 1c). MARGE shifts positions of individual sequence to their corresponding reference coordinates for motif analysis and visualization (Fig. 1 d-g). MARGE offers de-novo

motif analysis for individualized genomes (Fig. 1g), as well as a new algorithm to identify transcription factor binding motifs associated with allele-specific transcription factor binding or open chromatin (Fig. 1h). Each step of the MARGE pipeline is discussed below.

### **Merge, filter, and pre-process VCF files**

The initial step of the MARGE pipeline is to generate a set of high-confidence sequence differences between the alleles of interest (Fig. 1a). MARGE allows some basic filtering of VCF files by quality scores, however VCFtools (26) provides more sophisticated tools for this purpose. For some sequencing projects like the mouse genome project (29), SNPs and InDels are annotated in separate files, whereas other projects like the 1000 Genome project (30) provides one large file with SNPs and InDels. When SNPs and InDels are provided separately, MARGE merges them as a first step. If a combined file is provided then the first processing step is skipped. In cases where SNPs overlap deletions or insertions within one genomic background the SNP is filtered out and the longer mutation is kept. MARGE also simplifies the annotation of the variants per genotype (Fig. 2a). In cases where more than one possible mutation occurs in a particular genomic location (e.g. two different genotypes have two different mutations in comparison to the reference genome), the mutation is not always annotated as the shortest mutation per genotype. As shown in Fig. 2a the genetic variant for genotype2 is annotated as GTT -> GTTGTT. MARGE processes each genotype separately and therefore calculates the shortest genetic variation for each genotype (in this case T -> TGTT).

### **Generating individualized genomes**

MARGE produces individualized genomes by inserting the alleles from the VCF file into the reference genome and generating fasta files, which then can be used to make indices for mapping software. For homozygous data, only one genomic sequence is generated. Generation of individualized genomes and interpretation of allele-specific mapping for heterozygous data requires an additional step. Specifically, alleles at heterozygous sites need to be assigned on the same chromosome as neighboring heterozygous alleles. In genetics, this is called knowing the *phase* of the genotypes. Phase is especially important for MARGE when variants are in close proximity, because most sequencing reads are between 50-200 base pairs in length. When multiple SNPs reside in the same read, the correct combination of alleles in the genomic index is essential for accurate mapping and downstream interpretation. MARGE inherently assumes that all heterozygous data is phased. There are good resources for phasing genotypes in human populations. For example, phasing can be achieved using BEAGLE (31) or SHAPEIT (32) in conjunction with known haplotype structure of large reference populations such as the 1000 Genomes Project. In cases where phasing is not easily possible (e.g. F2 generation of inbred mice) loci where mutations overlap within the

read length should be excluded from the analysis.

### **Mapping data to individualized genomes**

Mapping of sequencing experiments to the individualized genome provides better results and decreases the possibility of incorrect mapping due to technical bias (Fig. 1c). This is especially true in datasets with a large number of differences to the reference. In these cases, mapping to the reference can introduce bias and in the case of datasets containing heterozygous genotypes can lead to overestimation of allele-specific expression or binding (33-35). To assess the effect of individualized genomes on mapping, we used a ChIP-seq dataset from inbred strains of mice. This provided a simplified situation since their genomes are entirely homozygous and all sequence tags originated from a genome of known sequence. Specifically, we used a PU.1 ChIP-seq dataset from three strains of mice (C57BL/6J, NOD/ShiLtJ, and SPRET/EiJ) (21). C57BL/6J (C57) is the commonly used reference genome and differs to NOD/ShiLtJ (NOD) in about 5 million genetic variants (89% SNPs, 11% InDels), whereas SPRET/EiJ (SPRET) provides about 43 million variants (89% SNPs, 11% InDels). Mapping of the ChIP-seq data to their respective genomes affected the overall mappability of the reads (Fig 2b) and the percentage of uniquely mapped reads (Fig 2c). The difference in mapping is directly correlated to the number of differences between the genomes. After removing all reads that map to multiple locations, peaks were called on all datasets separately and compared. Peaks from the C57 ChIP-Seq mapped to C57 and NOD genomes show only small differences (Fig. 2d) (about 1% of peaks are unique to either genotype), but increasing the number of variation between the genotypes lead to many peaks uniquely called in one of the mapped datasets (up to 12%). Also when comparing a PU.1 ChIP-seq dataset in human lymphoblastoid cell lines (25) mapped to the reference versus the individual genomes only about 90% of reads were mapped to the same loci (Supp. Fig. 1a). The number of differences between the hg19 reference genome and the individualized genomes is smaller than for the mouse data, but still up to 4% of peaks were uniquely called on either the dataset mapped to the reference or the individual genome (Supp. Fig. 1b). Therefore, mapping the data to the correct individualized genome increases the mapping accuracy substantially, leading to a more precise downstream analysis.

### **Additional processing for heterozygous data**

Many studies in mice use hybrid mouse strains (F1) generating heterozygous mice from two homozygous parents (Fig 1d). Furthermore, all human genomes are heterozygous in many loci and due to the advantages in sequencing technology, have become more realistic to study genome-wide. To improve mapping for heterozygous data, statistical methods have been developed (e.g. WASP (36)). Unfortunately, these methods can only handle SNPs. In order to also analyze heterozygous data with InDels, we map our data to two reference

genomes corresponding to alternative parental alleles. To effectively analyze heterozygous data, allele-specific expression or binding needs to be calculated. For this step, MARGE filters all reads with perfect alignment followed by filtering of all reads spanning a variant between the two parental strains (Fig 2e). If the heterozygous data is not phased, all regions that contain more than one mutation within the length of one read should be excluded from the analysis. This procedure makes sure that it is possible to confidently identify the allele of origin. To assign allele-specific reads correctly all loci without any variation are annotated with half of the perfectly aligned reads, because half of the reads that are sequenced originate from allele 1 and the other half from allele 2. For loci with allele-specific sequences, the ratio of reads per allele is calculated based on the reads spanning variations. Then the loci are annotated with the corresponding ratio of all perfectly aligned reads mapped to this locus.

### **Shifting to reference coordinates**

A major challenge of mapping data to individual genomes is that the experiments cannot be easily compared because of insertions and deletions (Fig. 1e). For example, the chromosomal locations between individuals (and across homologous chromosomes within heterozygous individuals) do not correspond to each other anymore. Therefore, to be able to use external analysis software and to visualize the data in the UCSC genome browser (37), we designed MARGE to shift mapped data back to reference coordinates (Fig. 2f). To accomplish this, MARGE generates shifting vectors for each genome (or haploid genome in the case of human/heterozygous data). Motifs can overlap insertions (M2) and deletions (M1) in the reference genome (Fig. 2f). The M2 motif consists of 6 bases, but after shifting the length is shrank to 3 bases due to the deletion. Therefore, positional shifting has the potent to introduce problems. For example, InDels can cause potential TF binding motifs to disappear or appear, which is of interest because these cases likely have functional consequence. Another complication of shifting coordinates occurs in the identification of ChIP-seq peaks from variable chromosomal sequences (i.e. shifting can cause a loss of peaks). This is because ChIP-seq peak calling tools often require a minimum length in order to identify peaks and this might not be reached after shifting. To check how frequently a peak was lost, each PU.1 ChIP-Seq dataset performed in human lymphoblastoid cell lines (25) was mapped to its individual genome and peaks were called with HOMER both before and after shifting. There are up to 2 million genetic differences between the reference genome (hg19) and the allele-specific genomes per individuals, but only up to 11 peaks are lost after shifting (which corresponds to less than 0.1% of all peaks) (Fig. 2g, Sup. Table 1). Also when repeating this procedure for diverse mouse strains (with more than 40 million genetic differences) only about 0.2% of all peaks were lost (Sup. Table 2). Therefore, despite an opportunity for difference to emerge in peak calling, we conclude that this phenomenon is very rare and does not offset the advantages from more accurate mapping.

## Data Visualization

Tools like the Integrative Genomics Viewer (IGV) (38,39) allow visualization of individual genomes, but require the user to install the software locally, which is not preferable for data sharing. One of the most common software platforms to visualize next-generation sequencing data online is the UCSC genome browser (37). Although a powerful tool, it does not allow the usage of other genomes than the references. To account for this, after shifting the genomic coordinates from the individualized genomes to the reference genome, MARGE can generate UCSC genome browser files (e.g. bedGraphs and bigWig files) that take into account individual genomic features (Fig. 1f). In addition, MARGE can generate BED (Browser Extensible Data) (40) files with all alternate alleles relative to the reference coordinates for upload to the genome browser (Fig. 2h). We also provide basic tools to interact with the different individual genomes. For example, we make it possible to directly compare the number of polymorphisms between different datasets in a table format for either all variants (Table 1) or for all private variants (those which can only be found in a particular individual compared to all others) (Table 2). More importantly however, MARGE can align nucleotide sequences from different individuals or chromosome sequences such as nucleotides or protein sequences. This application integrates RefSeq (41) or common gene name information to provide alignment for genes of interest, but is also able to extract the sequence for every genomic location of interest. This provides a fast and easy way to check for differences in genes or non-coding regions for different genetic backgrounds. This also simplifies the design of primers or other constructs, because differences can be checked by simple alignments of VCF files. To enable some more user-specific analysis, MARGE annotates files containing genomic coordinates with all genetic variants and generates files with genotype-specific sequences.

## De novo motif analysis

One of the first steps in analyzing ChIP-seq data is motif analysis. The de-novo motif analysis software from HOMER (1) was adapted to allow the integration of the individual genomes (Fig. 1g). We extended the de-novo motif finding algorithm (1) with a function to extract the sequences of the different genotypes as inputs to make sure that the motif finding algorithm is applied to the correct sequences and finds the motifs enriched in the sequence of the genotype not of the reference. It is possible to use different genotypes for the foreground sequences and the background sequences when unique peaks in two different genotypes are compared as foreground and background. These extensions make MARGE a powerful tool in comparing enriched motifs in two different genotypes.

## Motif mutation analysis

MARGE was primarily developed to determine importance of various nearby transcription factor motifs on the binding of a given transcription factor (Fig. 1h). It can analyze transcription factor binding profiles for two genomes in a pairwise fashion, but is also able to

analyze the binding profiles of many different genomes together (Fig. 3). The first case is preferable when two datasets have many genetic differences (e.g. two diverse mouse strains), as it may be more cost effective experimentally (pairwise comparison). For the analysis of human samples, however, it may be preferable to have more individuals, as the number of differences between two human genomes is fewer. In this scenario, a larger sample size may be required to achieve statistical power (all-versus-all comparison). MARGE uses a list of hand-curated motifs from the JASPAR motif database (42) as default, but also allows user-defined input.

#### *Pairwise comparison*

For the pairwise comparisons, peak files of both genotype alignments are merged and annotated with read counts (Fig. 3a). To account for differences between the alleles, the individual genome sequence is extracted and scanned with the motif-scanning algorithm provided by HOMER. Each motif is analyzed separately. Peaks without the motif that is currently scanned for are excluded from the analysis of this particular motif, but are considered for other motifs. Therefore, the analysis of every transcription factor motif is done on a different number of peaks. The fold change of the normalized read counts between the two alleles is calculated. Finally, the distribution of the fold change is calculated for all peaks, all peaks with a mutation in the motif of interest in allele1 and all peaks with a mutation in the motif of interest in allele2. To ensure that a motif is not just considered allele-specific because its log-odd score was slightly below the arbitrarily defined threshold in one of the alleles, MARGE extracts the sequence of the potential motif from each allele and calculates the log odd score based on the provided position weight matrix (PWM). By default a motif is considered missing when the log odd score is smaller or equal to zero, but the user can change this value to whatever seems suitable. MARGE also provides the possibility to define a motif as missing when its log-odd score in one allele is less than n% of the log-odd score in the other allele. To determine the significance of every motif a Student's t-test is performed between the general fold change distribution and the fold change distribution of allele1 and allele2, respectively.

Furthermore, the p-value between the distributions of the two alleles is calculated. This procedure is repeated for all transcription factors of interest. All p-values are multiplied by the number of comparisons to correct for multiple testing.

Allele-specific binding can be observed due to the loss of the binding site for the collaborative factors or the measured transcription factor itself. Additionally to analyzing every peak with the motif of interest, MARGE can analyze only peaks where all loci with differences in the motif of the measured TF between genotypes are filtered out. A Student's t-test is performed on the remaining distributions and the p-values are multiplied by the number of comparisons. MARGE outputs a motif mutation plot showing the distribution of mutations in relation to the fold change for each transcription factor (bottom Fig. 3a, Sup. Fig. 2a). It further outputs a density distribution plot for the fold change distribution of all peaks with changes in the motif in allele1, allele2, and the background (Sup Fig. 2b).

### All-versus-all comparison

In order to perform an all-versus-all comparison on more than two genotypes, peaks are called for all genotypes individually (Fig. 3b) and annotated with read counts. In case of heterozygous genotypes, peaks should be called on alleles separately and also be annotated with allele-specific reads (Fig. 2e). Both alleles are then analyzed as if they were independent genotypes. Therefore, when comparing for example 3 heterozygous genotypes, MARGE actually analyzes 6 independent samples. All sequences of all genotypes are scanned for the motifs of interest. To model the impact of the motif on the binding of the measured factor a Linear Mixed Model (LMM) is used. The binding of the measured factor is modeled as the fixed effect motif existence or motif score (defined by the user) with random effects locus and genotype (Formula 1) with the lme4 package (43) in R (44).

$$binding_i = \alpha + \beta * motifexistence_i + \gamma_{locus_i} + \delta_{genotype_i} + \varepsilon_i$$

or

$$binding_i = \alpha + \beta * motifscore_i + \gamma_{locus_i} + \delta_{genotype_i} + \varepsilon_i$$

with

$$\begin{aligned}\gamma_{locus_i} &\sim N(0, \sigma_{locus}^2) \\ \delta_{genotype_i} &\sim N(0, \sigma_{genotype}^2) \\ \varepsilon_i &\sim N(0, \sigma^2)\end{aligned}$$

To calculate significance for each motif, the drop1 command is used. It compares a model including motif score (motif existence, respectively) with a model without motif score (motif existence, respectively) and reports the Akaike information criterion (AIC) (45) for the difference. To keep the run time reasonable, MARGE implements threading for this procedure.

### Data mapping

All data was mapped using bowtie2 (27) with default parameters. The data for the different inbred strain of mice and the human data were mapped to the individualized genomes. The individualized genomes were generated using bowtie2-build with default parameters. The data for C57BL/6J was mapped to the mm10 reference genome from the UCSC genome browser (37). The human reference genome was hg19. Uniquely mapped reads are all reads that were mapped to only one unique region of the genome.

To analyze the impact of the genome on the accuracy of the mapping, all mouse ChIP-seq data sets in LPMs (21) were mapped to the three strain genomes C57, NOD, and SPRET. For the human data (25) all data was mapped against the individualized genome for allele 1, allele 2 and the hg19 reference genome. To assess the impact of the mapping on peak calling all reads that were mapped to more than one region of the genome were removed.

### ChIP-seq analysis

All ChIP-seq data sets were analyzed with HOMER after being shifted to reference coordinates. Peaks were called using findPeaks with default parameters and `-style factor`. For the LPM data set inputs were used for the peak calling. In case of the liver data and the human data no input was available and peaks were called without inputs. After running MARGE on the data, the list of significant motifs was reduced and summarized using HOMER's compareMotifs.pl.

### **Simulation of a data set**

MARGE is based on the model of collaborative binding for TFs and important collaborative TF binding motifs therefore should be identified as significant. According to this model a TF can only bind if the collaborative factor can bind, too. Applying this idea to two different genotypes means that if the motif is missing in genotype1 the binding of the measured factor should be lost in genotype1 and be not affected in genotype2 (genotype-specific binding). It further means if the motif is found in both genotypes binding should be similar between them (genotype-similar binding).

For the synthetic dataset, ten motifs were randomly chosen and defined as important collaborative TF for PU.1 (Tead3, Ventx, and Zic1), somewhat collaborative (Rora, Znf354c, and Plag1) and not collaborative (Pax6, Nr4a2, Lin54, and Bhlha15) (Fig. 4a). The genomes from three mouse strains (C57BL/6J (C57), BALB/cJ (BALB), and SPRET/EiJ (SPRET)) were scanned for the occurrence of all motifs (including PU.1). Next a peak file was generated for all genomic locations where the motif of interest was within 200bp of the PU.1 motif. These files were merged between two strains (C57 and BALB, C57 and SPRET, BALB and SPRET). To model genotype-specific binding, the fold change was randomly chosen to be between 2 and 10fold. For genotype-similar binding the fold change between the strains was within 1.5 fold. In all cases the read counts were randomly chosen between 0 and 500. To include biological noise in this dataset 85% of peaks with genotype-specific TF binding motifs follow the genotype-specific binding for highly collaborative motifs. For somewhat collaborative motifs 50% follow this pattern, whereas in the case of not collaborative motifs only 10% of peaks with genotype-specific TF motifs also show genotype-specific binding. To model genotype-similar binding for all highly collaborative motifs 85% of all peaks with the same motif show genotype-similar binding, for somewhat collaborative motifs 50% of the peaks have genotype-similar binding, whereas for not collaborative motifs only 10% show genotype-similar binding. The rest of the peaks show genotype-specific binding randomly assigned to one of the two strains.

## **RESULTS**

### **MARGE recognizes collaborative motifs in synthetic dataset**

To test the accuracy of the method, a synthetic dataset was generated simulating a ChIP-seq experiment using an antibody against PU.1 (for more details see Material and Methods, Fig. 4a). Ten motifs were randomly chosen and defined as important collaborative TF for PU.1

(Tead3, Ventx, and Zic1), somewhat collaborative (Rora, Znf354c, and Plag1) and not collaborative (Pax6, Nr4a2, Lin54, and Bhlha15) (Fig. 4a). Data was simulated for three different homozygous mouse strains (C57, BALB, and SPRET). Comparing one representative of the different motif categories shows that the algorithm is able to detect very high significance for Tead3 (defined as highly collaborative), medium significant for Plag1 (defined as somewhat collaborative) and no significance for Nr4a2 (defined as not collaborative) (Fig 4b, Sup. Fig. 2a). In all three comparisons about 50% of all peaks had the motif of interest, so the significance is not dependent on the percentage of peaks having the motif. The algorithm is able to detect significance for all motifs that were collaborative and showed lower or no significance for all non-collaborative motifs (Fig. 4c). PU.1 was almost always recognized as a significant motif, which is expected as the peaks were modeled according to a PU.1 ChIP-seq experiment.

### **MARGE analysis output**

In order to learn more about important position in the motif of the candidate transcription factor, MARGE offers a motif mutation position analysis (Fig. 4d, Sup. Fig. 3a). Fig. 4d shows an example for mutations within the PU.1 motif for the comparison C57 versus BALB on the simulated data set for Ventx. Mutations with significant effects on binding are marked by dots, whereas stars mark mutations with non-significant effects. Each base is colored differently, so it is not only possible to see which positions are mutated (significantly and non-significantly), but also to which other base. In the simulated dataset, even highly conserved residues in the motif can have mutations without an effect on binding (e.g. Fig. 4d, the highly conserved guanine at position 8 has 21 mutations from G->A that are significant but also 5 mutations from G->A with no effect). In the simulated data this was inherently part of it due to the modeling of biological noise (15% of genotype-specific peaks did show genotype-similar binding). It also should be noted that most differences that could be found were InDels (63 significant versus 10 not significant) or multiple SNPs within one motif (27 significant versus 3 not significant). MARGE also provides a plot that shows the distribution of the Ventx motif around the anchor transcription factor motif PU.1 (Fig. 4e, Sup. Fig. 2b) to see if the motif overlaps the anchor TF motif or if it is only randomly distributed within the peak. This plot allows the user to explore how the motifs of interest are distributed around the center of the peak to get a better understanding of the effect of this motif on the binding of the anchor TF.

### **Pairwise analysis of mouse data**

To show that the method also works on real data we analyzed data previously published in (21) and (46). We assessed PU.1 (a macrophage LDTF) binding in large peritoneal macrophages (LPM) in three different inbred mouse strains C57BL/6J (C57), NOD/ShiLtJ (NOD), and SPRET/EiJ (SPRET). These strains differ substantially in mutations to each other (Table 1). To show the correctness of the method we generated a list of motifs that were previously discovered (21) to be involved in the establishment of PU.1 binding in macrophage

(PU.1, PU.1-IRF, ETS1, SpiB, CEBP, AP-1, Arid3a). Additionally, we chose some transcription factors not expressed in LPMs or with known binding patterns different from PU.1 in macrophages. We chose the motifs of Bcl6 (not expressed in LPM, with a known function in B cells (47)), NeuroD1 (not expressed in LPM, associated with neurons (48) and diabetes (49)), RORgt (not expressed in LPM and mainly associated with thymocytes (50,51)), and Gfi1b (not expressed in LPM and associated mainly with neutrophil differentiation (52)).

MARGE could reliably detect motifs that are significantly associated with PU.1 binding, independent of the number of peaks containing the motif, or the number of mutations in these peaks. For example mutations in CEBP, an important LDTF in macrophages, were detected as significantly associated with PU.1 binding (Fig 5a). The plot showing the positions of mutations within the motif shows enrichment for mutations in the conserved bases T (bases 2 and 3) and A (bases 8 and 9) in comparison to the rest of the bases in the motif (Fig. 5b). Most causal mutations are due to multiple SNPs or InDels, not merely one single SNP. The CEBP motif is distributed closely around the PU.1 motif (where PU.1 is bound) without any motifs overlapping the PU.1 binding site (Fig. 5c). Although the peaks are 200 base pairs with regard to the reference genome, the sequences analyzed can be longer due to long insertions in the different strains resulting in peaks with a size of 300 in this case. Figure 5d shows two examples of how SNPs can influence observed PU.1 binding. In the left panel PU.1 is only bound in SPRET. A SNP in SPRET in comparison to C57 and NOD adds a PU.1 binding motif adjacent to an existent CEBP motif resulting in the observed genotype-specific binding. The right panel shows how losing a CEBP binding motif in C57 and SPRET close to a PU.1 binding motif existing in all three strains can cause PU.1 binding to be lost. MARGE could not find any significant association between motif existence and binding for the motifs chosen to provide negative controls (Fig. 5e). Although the number of mutations between two genotypes correlates with the significance of the analysis result (due to a bigger sample size), even with a low number of genetic variations MARGE was able to detect almost all significant motifs. To further test MARGE, we applied it to ChIP-seq experiments in four different strains (C57BL/6J (C57), A/J (AJ), CAST/EiJ (CAST), and SPRET/EiJ (SPRET)) for three different factors (CEBPa, FOXA1, and HNF4A) in whole liver from (46). CEBPa is an important TF in hepatocytes (54,55) (which make up about 70% of all cells in the liver (56)) and macrophages. FOXA1 plays important roles for the development and maintenance of the liver, mainly in hepatocytes (15,57) and HNF4A is an important liver TF mainly associated with hepatocytes (reviewed in (58)). Figure 5f shows an example where the TF binding motifs for all three factors were found, but binding could only be observed in AJ, C57, and CAST. Binding in SPRET was lost due to the loss of an adjacent RORA motif. After applying MARGE to the data, all significant motifs were compared to each other and summarized (compare Materials & Methods). In almost all pairwise comparisons for the three different factors the measured factor and the two collaborative factors were found as highly significant (Fig. 5g).

Nuclear receptors, which play important roles in the liver (reviewed in (59)), were found as significant in all three comparisons.

#### **All-versus-all analysis of homozygous mouse data**

To show the correctness of the all-versus-all analysis, we reanalyzed the mouse ChIP-seq datasets for CEBPa, FOXA1, and HNF4A from whole liver (Fig. 6a). Almost all motifs that were found significant in at least one pairwise comparison were detected as significant in the all-versus-all comparisons (compare Fig. 5g and Fig. 6a). Applying the motif score or the motif existence in the LMM produced almost the same results, with some motifs differing. The motif existence approach should be used with caution since adjusting the threshold that defines a sequence as motif can have large impacts on the results. Therefore, the all-versus-all comparison is able to confirm motifs significantly associated with binding of CEBP, FOXA1, or HNF4A in whole mouse liver previously identified by MARGE's pairwise comparisons. To make sure that the all-versus-all comparison is sensitive, we shuffled the strain order and repeated the analysis (Fig. 6b). To assess how much the results are influenced when very similar strains are shuffled, AJ and C57 were switched, but CAST and SPRET were kept at the same position. To further assess robustness of the results, the more diverse strains were shuffled with the more similar strains. Furthermore, we used completely different mouse genomes (NOD, DBA, PWK, and WSB). The color bar in Figure 6b shows the number of differences between the strains. When two very similar strains were changed (AJ with C57) the results are almost the same and the data sets are clustered together. However, as soon as more different strains are switched, the results changed dramatically. Motifs that are significant in all comparisons (e.g. NF1) should be counted as false positive results. This analysis shows that changing very similar data sets with each other does not affect the results, probably because most of the informative loci are found between these two strains and the two more diverse strains.

#### **All-versus-all analysis of heterozygous human data**

To show that MARGE is also able to analyze data from several human individuals with a low number of mutations, 34 PU.1 ChIP-seq datasets from Waszak et al (25) were analyzed with MARGE (listed in Sup. Table 1). The VCF files were downloaded from the 1000 Genomes Project (30) and the individual MARGE files and genomes were generated. A bowtie2 (27) index was created for each genome (two indices per genotype – one for the complete genome containing mutations on allele 1 and one for mutations on allele 2) and the ChIP-seq reads were mapped against both indices of the corresponding genotype. Only data sets with an overall mappability of 80% were considered in the downstream analysis (22 individuals). Peaks were called on all perfectly aligned reads and all peaks were merged and annotated allele-specific (320,146 peaks). To see how noise influences the MARGE results, MARGE was applied to an unfiltered peak file, as well as a peak file only containing reliable peaks with at least 8 reads in at least on individual (16, respectively). The dataset used in this analysis

was based on lymphoblastoid cell lines, human B cell lines infected with an Epstein-Barr virus to immortalize them.

Because the dataset is based on B cells it is not expected that any macrophage specific LDTFs are significant, instead B cell specific LDTFs (like PRDM1 also known as BLIMP-1, E2A etc.) would be expected to show a significant association with PU.1 binding (60). Figure 6c shows a UCSC genome browser session for one locus in three different individuals where one SNP that causes a loss of a PRDM1 motif close to an ETS factor motif is associated with loss of binding of PU.1. Applying the mutation approach systematically to all loci in all individuals and then summarizing the motifs, MARGE identified the B cell LDTF PRDM1 as highly significant, as well as a motif belonging to the IRF family of transcription factors known to play a role in B cells (Fig 6d) and an ETS motif, important for PU.1 binding. DUX4 has been previously associated with acute lymphoblastic leukemia (ALL) (61) which is coherent with the cancer-like cell type used in this experiment. MARGE was able to identify many other important transcription factors for B cells including NUR77 and a KLF binding motif (associated with B cell development (62,63)). The more stringent the filtering, the less significant motifs could be found. Filtering by 8 reads, about half of the significant motifs could be found. But filtering by 16 reads only found PRDM1 as significant. This highlights the importance of a good quality data set, because a lot of difference is found in lower bound peaks rather than the top peaks. Overall, MARGE was able to find significant motifs associated with PU.1 binding in human lymphoblastoid cell lines taking advantage of allele-specific binding in many individuals.

## Discussion

We developed a powerful tool to efficiently analyze ChIP-seq and other NGS data to understand the impact of transcription factor motifs on collaborative binding of transcription factors. MARGE is the first publicly available suite of software tools to integrate natural genetic variation (including InDels) and NGS binding data and provides complementary algorithms to analyze data from different genetic backgrounds in a pairwise manner as well as by utilizing a linear mixed model. It further provides many useful tools to directly look at genetic differences between different genetic backgrounds. By simulating a dataset and also applying MARGE to real world data, we could show that the algorithm works correctly in identifying motifs significantly associated with the binding of a measured transcription factor. Here, we applied MARGE to ChIP-seq data, which requires a well-working antibody for the reference transcription factor. However, MARGE can also be applied to ATAC-seq data or DNase I hypersensitivity data, which does not require any previous knowledge. In this case, rather than collaborative binding partners for a reference transcription factor, analysis of open chromatin would be expected to recover the dominant collaborative factors needed to establish open chromatin regions. Therefore, MARGE can potentially be applied to identify key regulatory factors in any cell type as long as parallel datasets from genetically diverse strains or individuals are available.

The algorithm assumes that the binding of the measured factor is only affected by local mutations in transcription factor binding motifs. As a result, sequence changes that influence binding on a global or long-distance scale in *trans* will not be detected and introduce noise to this. Furthermore, MARGE only analyzes one motif at a time. More complex relationships between transcription factors (e.g. the requirement for binding of three factors simultaneously) are not considered in the analysis. As in every analysis based on statistical tests, the power of discovery is dependent on the number of observations. A greater number of genetic variations between two individuals provides a better analysis result and will detect more significant motifs. For comparisons with low numbers of genetic variations MARGE offers a linear mixed model to increase the power of detection by merging all genetic variation between all individuals. This, however, requires substantially more experiments. Furthermore, the software is dependent on a list of position-weight matrices for the detection of TF binding sites. It is known that TF can bind to very weak motifs that cannot be detected by a motif-scanning algorithm but play important roles in regulating gene expression (64). However, MARGE is dependent on finding motifs based on scanning the DNA for the consensus sequence provided by the PWM. This limits the sensitivity of MARGE. Improvements in our understanding how to detect motifs in sequence will therefore improve the power of MARGE. Similar to de-novo motif finding, also MARGE only detects TF motifs. There are sometimes many similar transcription factors capable of binding the same consensus motif, which MARGE cannot discriminate. As more TFs and their motifs are characterized, these types of analysis will surely improve.

Genome-wide association studies (GWAS) evaluating common sequence variants associated with diverse phenotypes consistently demonstrate that the majority of variants reside in non-coding regions of the genome (20,65,66). These findings suggest that such variants impose risk by altering promoter and enhancer elements that regulate gene expression. Interpretation of such variants is currently limited because the genomic location of the regulatory elements at which they could potentially exert their effects varies according to cell type. By identifying important motif mutations, MARGE can provide a new and unique way to analyze transcription factor binding and detect the major collaborative factors involved in the establishment of cell-specific enhancer landscapes. With the advances in sequencing technology and availability of human samples, MARGE can facilitate the analysis of datasets that provide insights into the relationship between non-coding genetic variation and gene expression in humans.

### **Availability**

The MARGE source code and installation package are freely available on GitHub:

[https://github.com/vlink/marge/blob/master/MARGE\\_v1.0.tar.gz](https://github.com/vlink/marge/blob/master/MARGE_v1.0.tar.gz).

The mouse LPM dataset from (21) was downloaded from the GEO database under accession number GSE62826. The data is available at [http://genome.ucsc.edu/cgi-bin/hgTracks?hgS\\_doOtherUser=submit&hgS\\_otherUserName=vlink&hgS\\_otherUserSession](http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=vlink&hgS_otherUserSession)

Name=MARGE\_LPM\_data. The mouse liver data set from (46) was downloaded from ArrayExpress Archive (<http://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-1414. The data is available at [http://genome.ucsc.edu/cgi-bin/hgTracks?hgS\\_doOtherUser=submit&hgS\\_otherUserName=vlink&hgS\\_otherUserSessionName=MARGE\\_Liver\\_data](http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=vlink&hgS_otherUserSessionName=MARGE_Liver_data). The human data set from (25) was downloaded from the ArrayExpress Archive under accession number E-MTAB-3657. The data is accessible at [http://genome.ucsc.edu/cgi-bin/hgTracks?hgS\\_doOtherUser=submit&hgS\\_otherUserName=vlink&hgS\\_otherUserSessionName=MARGE\\_human\\_data](http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=vlink&hgS_otherUserSessionName=MARGE_human_data).

MARGE is implemented in Perl and R (44). It has been tested on several UNIX systems, including CentOS and Debian with Perl version 5.20 and higher and R version 3.3 and higher. We provide a script that installs MARGE and allows download of pre-processed mutation data from the mouse genome project (29) and the genomes from the 1000 Genome Project used in this manuscript. MARGE requires the Perl core modules POSIX, Getopt::Long, Storable and threads, as well as the modules Set::IntervalTree (67), and Statistics-Basic (68). It further requires the R packages SeqLogo (69), gridBase (70), lme4 (71), and gplots (72). It also requires an installed version of gzip. For the motif mutation analysis MARGE requires HOMER (1) (<http://homer.ucsd.edu/homer/>) to be installed and executable. Without a working installation of HOMER, MARGE's functionality is limited to only visualization and annotation of the data.

## Funding

This work was supported by NIH grants CA173903, GM085764, and DK091183. CER was supported by NIH-NHLBI grant R00123485.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

We thank Leslie van Ael for support with figures. We further thank Ty Troutman, Jenhan Tao, and Inge Holtman for running MARGE and their feedback. We thank Chris Benner, Ty Troutman, Dylan Skola, Jenhan Tao, and Zhengyu Ouyang for critically reading the manuscript.

## References

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

2. Barish, G.D., Yu, R.T., Karunasiri, M., Ocampo, C.B., Dixon, J., Benner, C., Dent, A.L., Tangirala, R.K. and Evans, R.M. (2010) Bcl-6 and NF-kappaB cistromes mediate opposing regulation of the innate immune response. *Genes Dev*, **24**, 2760-2765.
3. Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F. *et al.* (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat Genet*, **38**, 1289-1297.
4. John, S., Sabo, P.J., Thurman, R.E., Sung, M.H., Biddie, S.C., Johnson, T.A., Hager, G.L. and Stamatoyannopoulos, J.A. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet*, **43**, 264-268.
5. Lefterova, M.I., Steger, D.J., Zhuo, D., Qatanani, M., Mullican, S.E., Tuteja, G., Manduchi, E., Grant, G.R. and Lazar, M.A. (2010) Cell-specific determinants of peroxisome proliferator-activated receptor gamma function in adipocytes and macrophages. *Mol Cell Biol*, **30**, 2078-2089.
6. Nielsen, R., Pedersen, T.A., Hagenbeek, D., Moulos, P., Siersbaek, R., Megens, E., Denissof, S., Borgesen, M., Francoijs, K.J., Mandrup, S. *et al.* (2008) Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev*, **22**, 2953-2967.
7. 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.
8. Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D. *et al.* (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, **444**, 499-502.
9. Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K. *et al.* (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, **3**, e7.
10. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **459**, 108-112.
11. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75-82.
12. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854-858.
13. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.

14. Bossard, P. and Zaret, K.S. (1998) GATA transcription factors as potentiators of gut endoderm differentiation. *Development*, **125**, 4909-4917.
15. Lee, C.S., Friedman, J.R., Fulmer, J.T. and Kaestner, K.H. (2005) The initiation of liver development is dependent on Foxa transcription factors. *Nature*, **435**, 944-947.
16. McPherson, C.E., Shim, E.Y., Friedman, D.S. and Zaret, K.S. (1993) An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell*, **75**, 387-398.
17. 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.
18. Link, V.M., Gosselin, D. and Glass, C.K. (2015) Mechanisms Underlying the Selection and Function of Macrophage-Specific Enhancers. *Cold Spring Harb Symp Quant Biol*, **80**, 213-221.
19. Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orozco, L.D. and Glass, C.K. (2013) Effect of natural genetic variation on enhancer selection and function. *Nature*, **503**, 487-492.
20. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, **106**, 9362-9367.
21. Gosselin, D., Link, V.M., Romanoski, C.E., Fonseca, G.J., Eichenfield, D.Z., Spann, N.J., Stender, J.D., Chun, H.B., Garner, H., Geissmann, F. *et al.* (2014) Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell*, **159**, 1327-1340.
22. Hogan, N.T., Whalen, M.B., Stolze, L.K., Hadeli, N.K., Lam, M.T., Springstead, J.R., Glass, C.K. and Romanoski, C.E. (2017) Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. *Elife*, **6**.
23. Fisher, R. (1919) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, **2**.
24. Pinheiro, J.B., Douglas. (2000) *Mixed-Effects Models in S and S-PLUS*. Springer.
25. Waszak, S.M., Delaneau, O., Gschwind, A.R., Kilpinen, H., Raghav, S.K., Witwicki, R.M., Orioli, A., Wiederkehr, M., Panousis, N.I., Yurovsky, A. *et al.* (2015) Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell*, **162**, 1039-1050.
26. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.
27. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.
28. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
29. Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M. *et al.* (2011) Mouse genomic

- variation and its effect on phenotypes and gene regulation. *Nature*, **477**, 289-294.
30. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68-74.
  31. Browning, S.R. and Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*, **81**, 1084-1097.
  32. Delaneau, O., Marchini, J. and Zagury, J.F. (2011) A linear complexity phasing method for thousands of genomes. *Nat Methods*, **9**, 179-181.
  33. Degner, J.F., Marioni, J.C., Pai, A.A., Pickrell, J.K., Nkadori, E., Gilad, Y. and Pritchard, J.K. (2009) Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, **25**, 3207-3212.
  34. Satya, R.V., Zavaljevski, N. and Reifman, J. (2012) A new strategy to reduce allelic bias in RNA-Seq readmapping. *Nucleic Acids Res*, **40**, e127.
  35. Stevenson, K.R., Coolon, J.D. and Wittkopp, P.J. (2013) Sources of bias in measures of allele-specific expression derived from RNA-sequence data aligned to a single reference genome. *BMC Genomics*, **14**, 536.
  36. van de Geijn, B., McVicker, G., Gilad, Y. and Pritchard, J.K. (2015) WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*, **12**, 1061-1063.
  37. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res*, **12**, 996-1006.
  38. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol*, **29**, 24-26.
  39. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, **14**, 178-192.
  40. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.
  41. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, **44**, D733-745.
  42. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R. *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*, **44**, D110-115.
  43. Douglas, B.M., Mächler; Ben, Bolker; Steve, Walker. (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*.
  44. R Development Core Team. (2016), *R Foundation for Statistical Computing*. <https://www.R-project.org/>, Vienna, Austria.
  45. Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*.

46. Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C. *et al.* (2013) Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell*, **154**, 530-540.
47. Dent, A.L., Vasanwala, F.H. and Toney, L.M. (2002) Regulation of gene expression by the proto-oncogene BCL-6. *Crit Rev Oncol Hematol*, **41**, 1-9.
48. Gao, Z., Ure, K., Ables, J.L., Lagace, D.C., Nave, K.A., Goebbels, S., Eisch, A.J. and Hsieh, J. (2009) Neurod1 is essential for the survival and maturation of adult-born neurons. *Nat Neurosci*, **12**, 1090-1092.
49. Kanatsuka, A., Tokuyama, Y., Nozaki, O., Matsui, K. and Egashira, T. (2002) Beta-cell dysfunction in late-onset diabetic subjects carrying homozygous mutation in transcription factors NeuroD1 and Pax4. *Metabolism*, **51**, 1161-1165.
50. Kurebayashi, S., Ueda, E., Sakaue, M., Patel, D.D., Medvedev, A., Zhang, F. and Jetten, A.M. (2000) Retinoid-related orphan receptor gamma (RORgamma) is essential for lymphoid organogenesis and controls apoptosis during thymopoiesis. *Proc Natl Acad Sci U S A*, **97**, 10132-10137.
51. Sun, Z., Unutmaz, D., Zou, Y.R., Sunshine, M.J., Pierani, A., Brenner-Morton, S., Mebius, R.E. and Littman, D.R. (2000) Requirement for RORgamma in thymocyte survival and lymphoid organ development. *Science*, **288**, 2369-2373.
52. Anguita, E., Candel, F.J., Chaparro, A. and Roldan-Etcheverry, J.J. (2017) Transcription Factor GFI1B in Health and Disease. *Front Oncol*, **7**, 54.
53. Ong, C.T. and Corces, V.G. (2014) CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, **15**, 234-246.
54. Jakobsen, J.S., Waage, J., Rapin, N., Bisgaard, H.C., Larsen, F.S. and Porse, B.T. (2013) Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries. *Genome Res*, **23**, 592-603.
55. Wang, N.D., Finegold, M.J., Bradley, A., Ou, C.N., Abdelsayed, S.V., Wilde, M.D., Taylor, L.R., Wilson, D.R. and Darlington, G.J. (1995) Impaired energy homeostasis in C/EBP alpha knockout mice. *Science*, **269**, 1108-1112.
56. Racanelli, V. and Rehmann, B. (2006) The liver as an immunological organ. *Hepatology*, **43**, S54-62.
57. Moya, M., Benet, M., Guzman, C., Tolosa, L., Garcia-Monzon, C., Pareja, E., Castell, J.V. and Jover, R. (2012) Foxa1 reduces lipid accumulation in human hepatocytes and is down-regulated in nonalcoholic fatty liver. *PLoS One*, **7**, e30014.
58. Babeu, J.P. and Boudreau, F. (2014) Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal inflammatory networks. *World J Gastroenterol*, **20**, 22-30.
59. Wagner, M., Zollner, G. and Trauner, M. (2011) Nuclear receptors in liver disease. *Hepatology*, **53**, 1023-1034.
60. Matthias, P. and Rolink, A.G. (2005) Transcriptional networks in developing and mature B cells. *Nat Rev Immunol*, **5**, 497-508.
61. Izraeli, S. (2016) Deciphering "B-others": Novel fusion genes driving B-cell acute lymphoblastic leukemia. *EBioMedicine*, **8**, 8-9.

62. Vu, T.T., Gatto, D., Turner, V., Funnell, A.P., Mak, K.S., Norton, L.J., Kaplan, W., Cowley, M.J., Agenes, F., Kirberg, J. *et al.* (2011) Impaired B cell development in the absence of Kruppel-like factor 3. *J Immunol*, **187**, 5032-5042.
63. Zikherman, J., Parameswaran, R. and Weiss, A. (2012) Endogenous antigen tunes the responsiveness of naive B cells but not T cells. *Nature*, **489**, 160-164.
64. Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S. and Levine, M.S. (2015) Suboptimization of developmental enhancers. *Science*, **350**, 325-328.
65. Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E. *et al.* (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*, **482**, 390-394.
66. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S. and Raychaudhuri, S. (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet*, **45**, 124-130.
67. Booth, B. (2014), <http://search.cpan.org/~benbooth/Set-IntervalTree/>.
68. Miller, P. (2014), <http://search.cpan.org/~jettero/Statistics-Basic-1.6611/>.
69. Bembom, O. (2016).
70. Murrell, P. (2014), <https://CRAN.R-project.org/package=gridBase>.
71. Walker, D.B.a.M.M.a.c.a.B.B.a.S. (2015) Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, **67**, 1-48.
72. Marc, G.R.W.a.B.B.a.L.B.a.R.G.a.W.H.A.L.a.T.L.a.M.M.a.A.M.a.S.M.a. (2016), <https://CRAN.R-project.org/package=gplots>.

## Tables and Figures Legends

**Figure 1: Overview of the MARGE pipeline.** (A) MARGE merges VCF files for SNPs and InDels, offers some basic filtering and split the merged VCF file into separate genotype-specific mutation files. (B) It then generates individual genomes by inserting the annotated mutations in the reference genome per genotype and (C) allows mapping of the experimental data sets to the individualized genomes. (D) The data mapped to the individualized genomes is then shifted back to the reference coordinates. (E) In case of heterozygous data additional processing is necessary. MARGE offers (F) scripts for data visualization including BED files for genetic variation per genotype. It further offers (G) de-novo motif analysis for the individual genomes to make sure the enrichment analysis is performed on the correct sequence instead of the reference. MARGE also offers a new algorithm (H) to associated TF binding motifs with genotype-specific binding for pairwise comparisons, as well as comparisons for many different individuals (all-versus-all comparison). Taken all of that together MARGE is able to identify TF binding motifs that are functionally associated with TF binding.

**Figure 2: Details of pipeline:** (A) MARGE merges SNP and InDel VCF files and then splits the merged file. It finds the shortest annotation for each mutation, changing the original annotation from the VCF file. (B) Comparison of the overall mapping efficiency. There is a small decrease in overall mappability when data is mapped to the reference. (C) Comparison of mapping efficiency for uniquely mapped reads after mapping to different genomes. There is an increase in mapping performance when mapped to individualized genomes. (D) Percentage of peaks uniquely called to dataset mapped to one genotype versus another. Up to 12% of peaks are unique to one genotype. (E) Pipeline for processing heterozygous data: Data is mapped to both alleles and shifted back to the reference coordinates. Reads that do not uniquely align to the genome are filtered out. Perfectly aligned reads, as well as perfectly aligned reads overlapping mutations are filtered out and peaks are called on perfectly aligned reads. For each locus without any mutations, the peaks for both alleles are annotated with half the reads that mapped to this locus. For each locus with mutations a ratio is calculated based on the reads overlapping mutations and then the locus is annotated with the number of perfectly aligned reads multiplied by the corresponding to the ratio. (F) Schematic of the shifting process: Genomic coordinates of the individual genomes do not concur with the reference due to InDels. MARGE shifts the individual coordinates to the reference without changing the length of the sequence. (G) Shifting peak coordinates leads to minor loss of peaks. 34 PU.1 ChIP-seq data sets were mapped and peaks were called before and after shifting. Even with 2 million genetic variants between the reference and the individualized

genomes only up to 11 peaks are different. **(H)** UCSC genome browser shot showing PU.1 ChIP-seq data in large peritoneal macrophages in 3 different inbred strains of mice (C57, NOD, and SPRET). Bed graphs generated by MARGE show genetic differences between the strains. The red rectangle shows a zoomed-in area of the UCSC genome browser.

**Figure 3: Schematic showing the algorithm for the motif mutation analysis for pairwise comparisons or comparisons of a big group of individuals.** **(A)** Pairwise comparison:

Data is mapped to individual genomes and shifted to reference coordinates. Peaks are called per genotype and are subsequently merged and annotated with the tag counts from the tag directories with HOMER. The merged file is iteratively scanned for the TF binding motifs of interest. For all peaks containing the current TF motif of interest (marked in green) the binding difference between the two genotypes is calculated (fold change). For each TF the fold change distribution of all peaks is plotted (more information Sup. Fig. 1a) and a Student's t-test is performed on the fold change distribution of all peaks versus all peaks containing a mutation in genotype1 (red) (genotype2 (blue), respectively). Further a t-test is performed comparing the fold change distribution of all peaks missing the motif of interest in genotype1 versus genotype2 (purple) and corrected for multiple testing. **(B)** Motif mutation analysis on more than two genotypes: Data is mapped to the individual genomes, shifted back to the reference coordinates and peaks are called on each genotype separately and subsequently merged and annotated. Heterozygous data should be annotated with MARGE's annotation function. The merged file is iteratively scanned for the TF motif of interest (marked in green). Per TF an output file is generated containing the locus, the binary existence of a motif, the motif score and the read counts. This output file is then inserted into a linear mixed model (LMM) implemented in R with the package lme4 modeling the binding as dependency of the motif score (or motif existence) with random factors Strain and Locus. A p-value is generated using the R command drop1 and corrected for multiple testing.

**Figure 4: Analysis of a simulated dataset** **(A)** Motifs were defined as important collaborative binding (Tead3, Ventx, Zic1), somewhat collaborative (Rora, Znf354a, Plag1) and not collaborative (Pax6, Nr4a2, lin54, Bhlha15). Peak files were generated for all loci where PU.1 and one of the TF are within 200bp to each other for three mouse strains (C57, BALB, and SPRET) and consecutively merged between two strains. For highly collaborative TF 85% of the strain specific peaks show strain specific binding (somewhat collaborative: 50%, not collaborative: 10%). Fold change was randomly chosen to be between 2 and 10 fold for differently and to be between 1 and 1.5 fold for similarly bound peaks. Read counts were randomly chosen to be between 0 and 500. **(B)** MARGE correctly identifies the association between motif and binding data. Motif mutation distribution plot (Sup. Fig. 2a) for one collaborative motif (Tead3) shows a highly significant association between motif mutation and binding data (medium significance for Plage1 (somewhat collaborative), no significant for Nr4a2 (not collaborative)). **(C)** Summary heatmap for all analysis on the simulated datasets.

MARGE showed high significance for the collaborative TF and less or no significance for non-collaborative TF binding motifs. **(D)** Motif mutation position plot for Tead3, showing which positions are mutated and associated with different binding (more information Sup. Fig. 3a). It furthermore shows that in most cases InDels and multiple SNPs cause significant change in binding. **(E)** TF binding motif distribution of PU.1 and Ventx. Motifs for Ventx are closely distributed around the PU.1 binding site (more information Sup. Fig 3b).

**Figure 5: Analysis using MARGE's pairwise-comparison** **(A)** Motif mutation plot for PU.1 data in LPMs analyzing the impact of mutations in the CEBP binding motif on PU.1 binding. Red ticks show mutations in the CEBP motif in C57 (blue for SPRET). Loss of the CEBP motifs is significantly associated with strain-specific PU.1 binding. **(B)** Motif position mutation plot for CEBP motif showing the position and effect of mutations in the CEBP motif in relation to PU.1 binding. The most conserved positions in the CEBP motif are associated with a loss of PU.1 binding. **(C)** The CEBP motif is distributed closely around the PU.1 motif with a depletion of the CEBP motif at the PU.1 binding site. **(D)** UCSC genome browser shot - Left panel: The gain of a PU.1 motif in SPRET adjacent to a CEBP motif results in PU.1 binding only in SPRET, but not in C57 or NOD. Right panel: The gain of a CEBP motif in NOD in close vicinity to a PU.1 motif results in PU.1 binding only in NOD. **(E)** Summary heat map of multiple testing corrected p-values for TF motifs associated with PU.1 binding. The heat map includes some negative control motifs that are not associated with macrophage biology which were not identified as significant. **(F)** UCSC genome browser shot - The loss of a RORA TF motif in SPRET causes loss of binding of CEBPa, FOXA1, and HNF4A in SPRET, but not in AJ, C57 and CAST. **(G)** Summary heat map of multiple testing corrected p-values of TF binding motifs associated with CEBPa, FOXA1, and HNF4A binding in whole liver. All factors reached significance in every pairwise comparison. Nuclear receptors were significantly associated with binding of the different factors.

**Figure 6: Results of all-versus-all analysis.** **(A)** Summary heatmap of multiple testing corrected p-values of all-versus-all analysis of CEBP, FOXA1, and HNF4A ChIP-seq data sets from whole liver in AJ, C57, CAST, and SPRET. The analysis confirms the results from the pairwise analysis performed in Fig. 5g. The same motifs are highly significance with slight variations independent of considering motif score (MS) or motif existence (ME). **(B)** Summary heatmap of multiple testing corrected p-value of the all-versus-all analysis for CEBP, FOXA1, and HNF4A ChIP-seq data sets with the original order of the strains and shuffled order of the strains to assess sensitivity of MARGE. The color of the boxes correlates to the number of mutations (from 0 – white to 50 million – brown). When very similar strains are switched (AJ and C57) the MARGE results are clustered together. As soon as more diverse strains are switched or different strains are used, the results cluster as outliers to the original data and almost all motifs lose significance. **(C)** UCSC genome browser shot visualizing three human PU.1 datasets. The allele-specific loss of a PRDM1 motif close to an ETS motif causes allele-

specific loss of PU.1 binding. **(D)** Summary heat map of multiple testing corrected p-values of transcription factor motifs significantly associated with PU.1 binding in human lymphoblastoid cell. Many TF motifs found to be significantly associated with PU.1 binding are either known to play important roles in B cell development and maintenance or cancer. By increasing the stringency of peaks included in the analysis (and decreasing the number of observations) the number of significant motifs decreases. Only PRDM1 is found as significant when using a filter of 16 reads.

**Supplemental Figure 1: Effect of mapping ChIP-seq data for human lymphoblastoid cell lines.** **(A)** Percentage of reads mapped to the same locus after mapping to individualized genotype and reference. Only about 90% of reads mapped to the same locus when comparing mapping results for individualized genomes versus the reference. **(B)** Percentage of peaks unique to either the individualized genome of the reference genome hg19. Up to 4% of the peaks were called uniquely in either the PU.1 ChIP-Seq dataset mapped to hg19 or the individualized genome.

**Supplemental Figure 2: Detailed description of the output plots MARGE generates for the motif mutation analysis.** **(A)** For each TF of interest MARGE generates this plot. The peaks are rank-ordered by most genotype-specific bound in genotype1 to most genotype-specific bound in genotype2. Genotype1 is color-coded red, whereas genotype2 is color-coded blue. The left upper corner shows how many mutations could be found in the motifs for genotype1 and genotype2. A red tick marks each peak without a TF binding motif of the TF of interest, if the motif is missing in genotype1 and a blue tick marks the motif is missing in genotype2. All data on the left bottom of the plot show peaks where the binding is very specific to genotype1. The right upper corner shows peaks where binding is very specific to genotype2. The box plot on the right summarizes the fold change distribution. Grey shows the fold change distribution for all peaks having the TF binding motif of interest. Red shows the fold change distribution for all peaks missing the TF binding motif of interest in genotype1, whereas blue shows the distribution for all peaks missing the TF binding motif in genotype2. A Student's t-test is performed comparing these distributions. The p-value is shown below the box plots. Red shows the comparison of the background (grey box) versus the distribution for genotype1 (red box), blue shows the comparison of the background (grey box) versus the distribution for genotype2 (blue box) and purple shows the comparison between genotype1 (red) and genotype2 (blue). **(B)** Kernel density plot for data shown in the motif mutation distribution plot. A Gaussian kernel is applied to the fold change distributions for all peaks with the motif of interest (black – background), all peaks with missing motifs in genotype1 (red) and all peaks with missing motifs in genotype2 (blue) and plotted. The p-values are the corresponding p-values from the t-test explained above.

**Supplemental Figure 3: Detailed description of additional output plots provided by MARGE.** (A) Motif position mutation plot shows the consensus logo of the position-weighted matrix (PWM) from the motif. Mutations that affect the TF binding between both genotypes are marked by dots, whereas a star marks mutations with no effect. All mutations resulting in an adenosine are color-coded green (cysteine are color-coded blue, guanine are yellow and thymine are red). The number of InDels and multiple SNPs are reported separately for significant changes in binding and no changes. The y-axis shows the frequency for the mutations. (B) Motif distance distribution plot. All peaks are centered on their anchor TF binding motif and the distribution of the TF of interest is plotted around the TF. Further, a genome-wide background is plotted. The right y-axis shows the motif frequency in the genotypes, the left y-axis shows the frequency in the background. The distribution of the anchor TF is plotted for the background and the genotypes in the order of background (grey), genotype1 (dark green), and genotype2 (light green). They should overlap, so only the distribution of genotype2 should be visible. For the motif of interest the genome wide background is plotted in grey, whereas the distribution for genotype1 is blue and for genotype2 is red. The distribution should overlap and only the distribution of genotype2 should be visible. The x-axis shows the distance of the motif to the center.

**Table 1:** Overview of all natural genetic variation found in all strain-wise comparisons

**Table 2:** Overview of all private genetic variation found in in this strain versus all other strains

**Supplemental Table 1:** Summary of peak numbers before and after shifting for all 34 individuals from a human PU1 ChIP-Seq dataset. Only up to 11 peaks are lost after shifting, which is less than 0.1% of all peaks.

**Supplemental Table 2:** Summary of peak numbers before and after shifting for mouse data. The number of genetic variations is up to 40 million between two strains, and less than 0.1% of the peaks are lost after shifting.

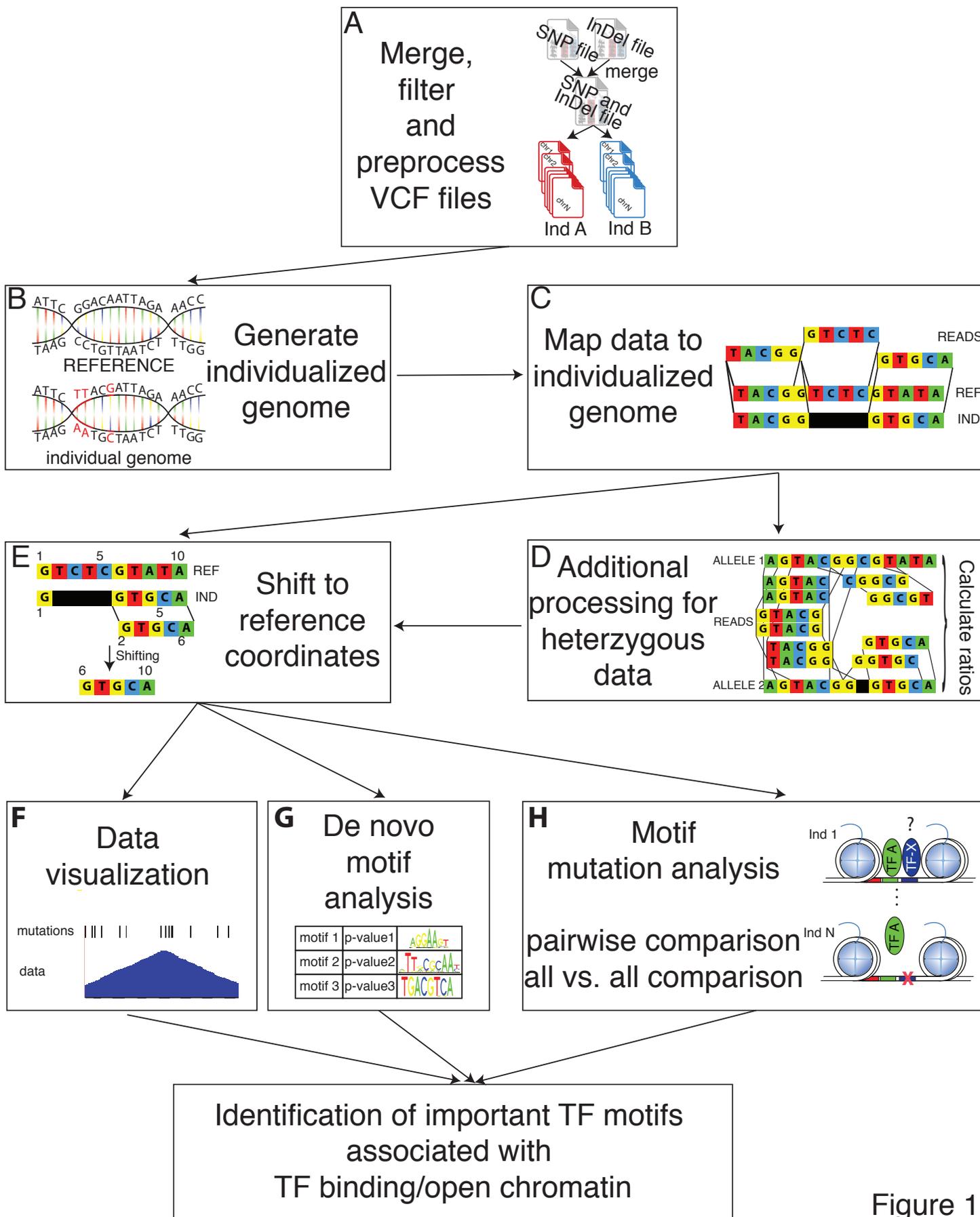


Figure 1

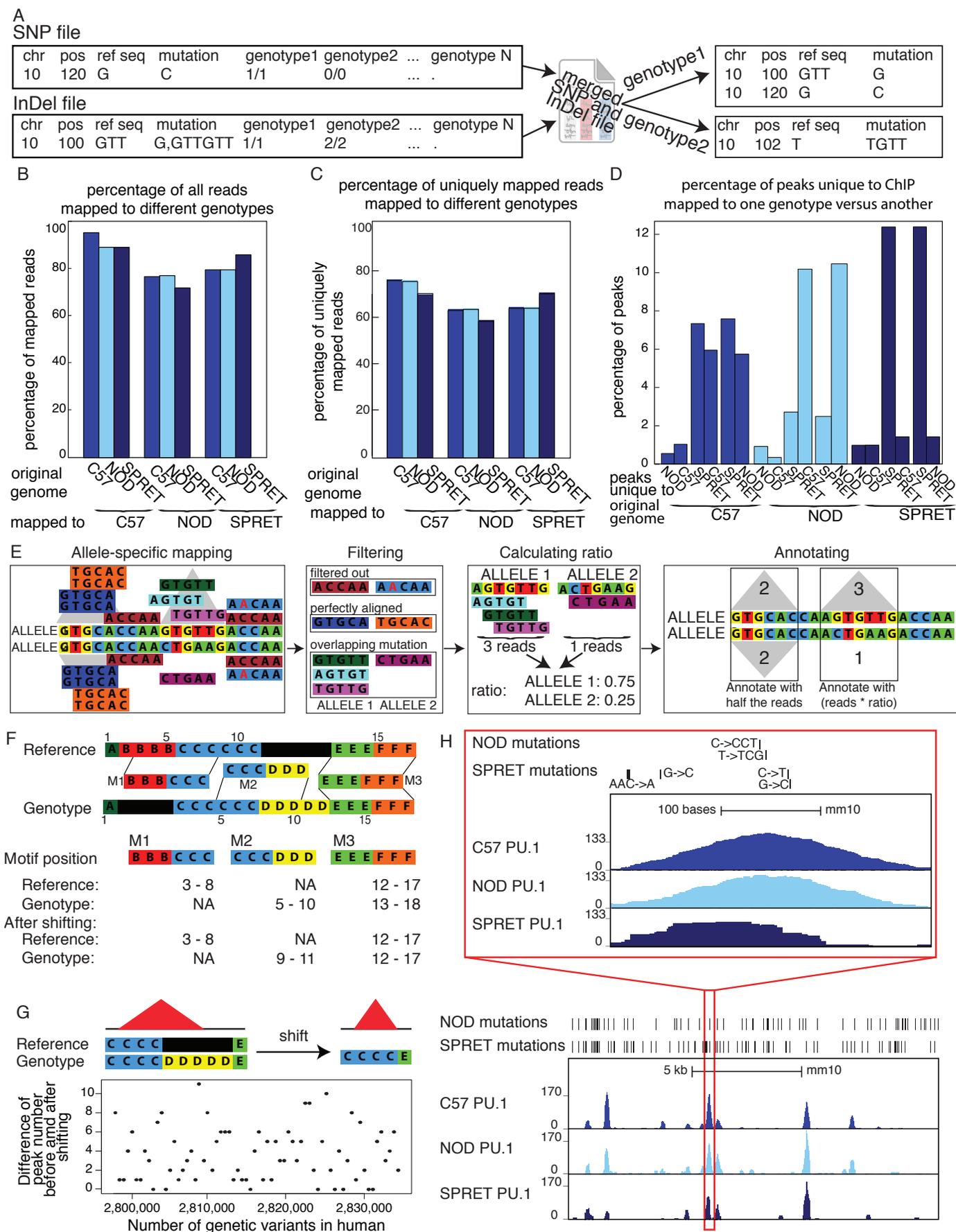


Figure 2

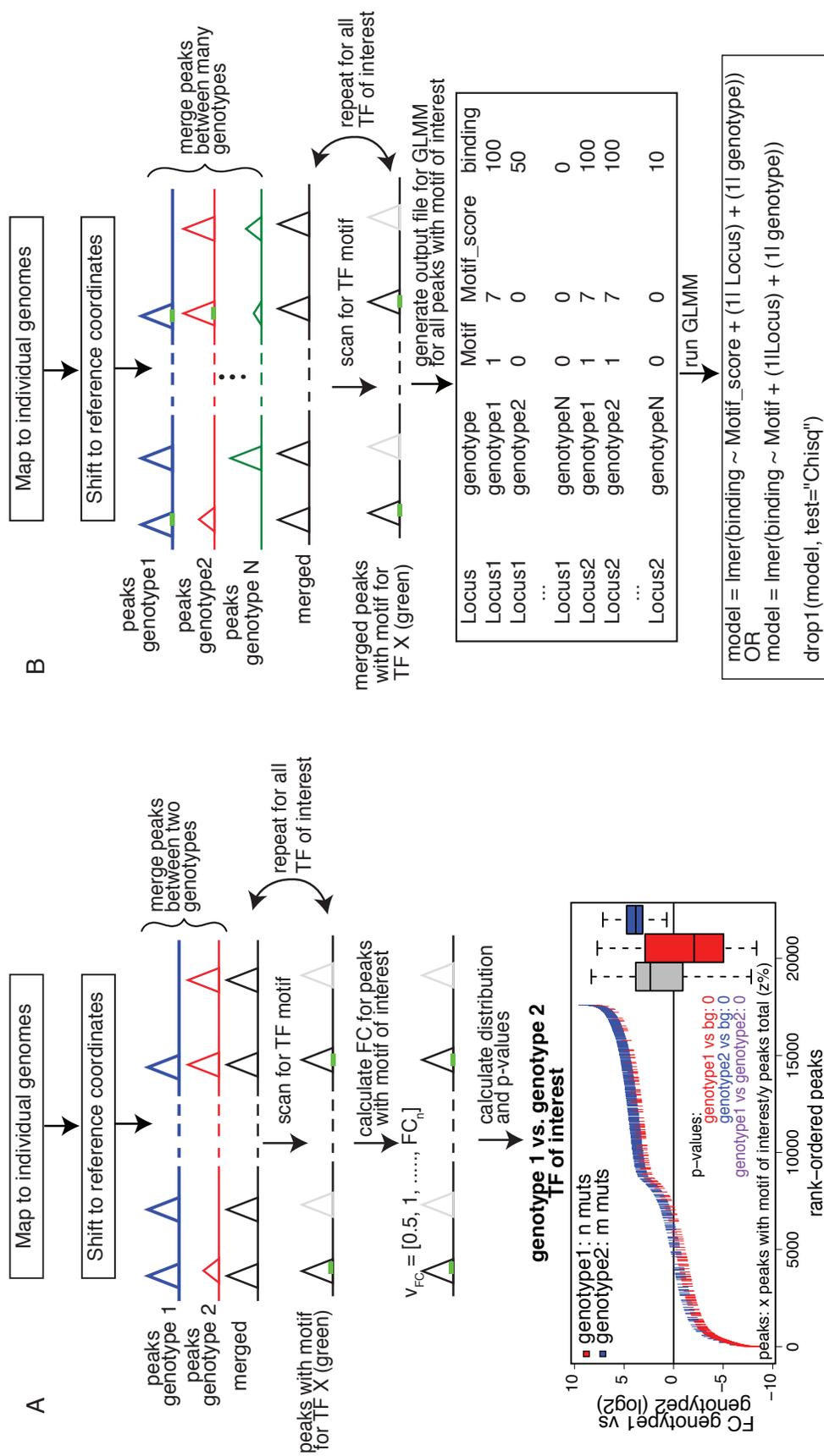


Figure 3

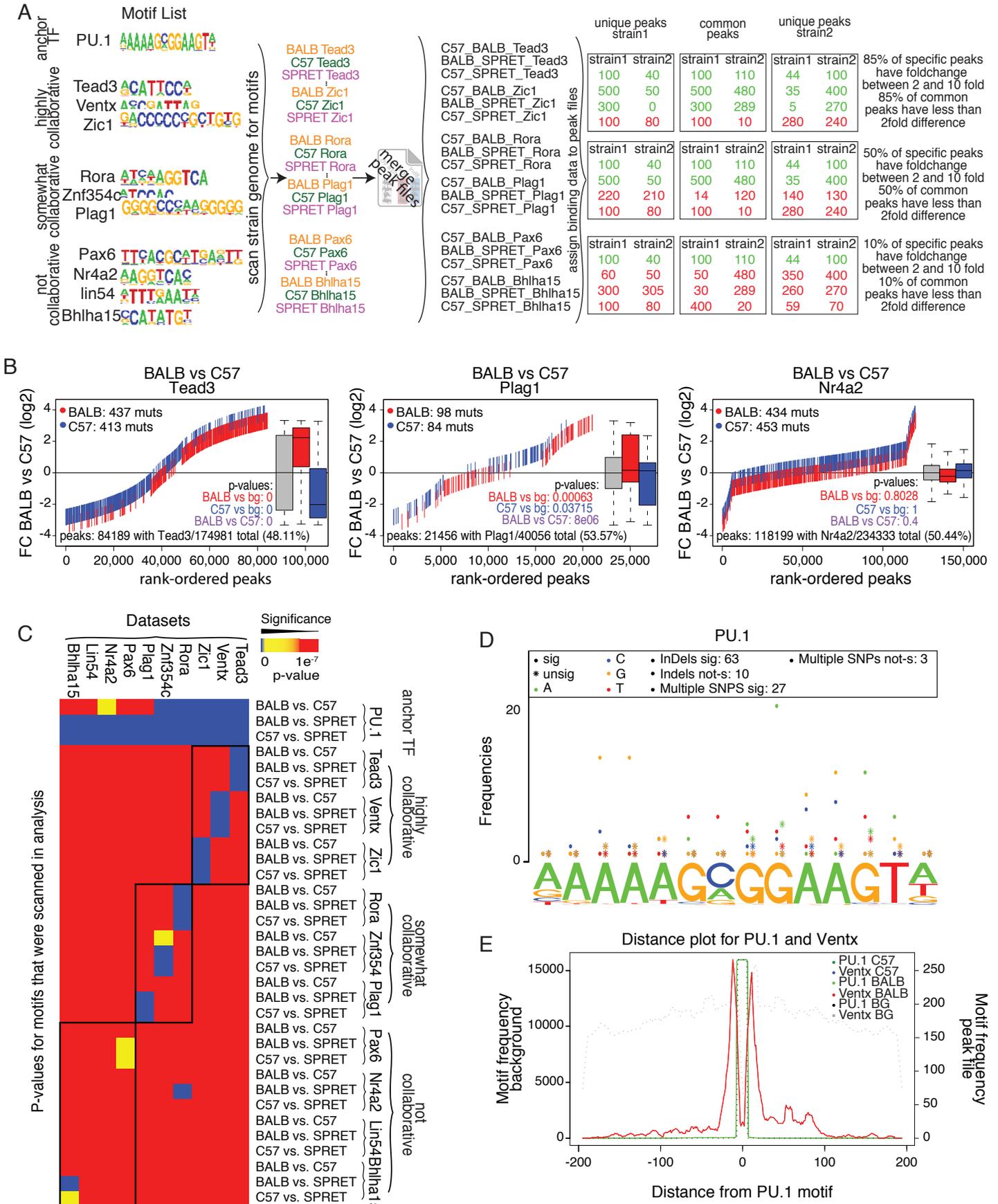


Figure 4

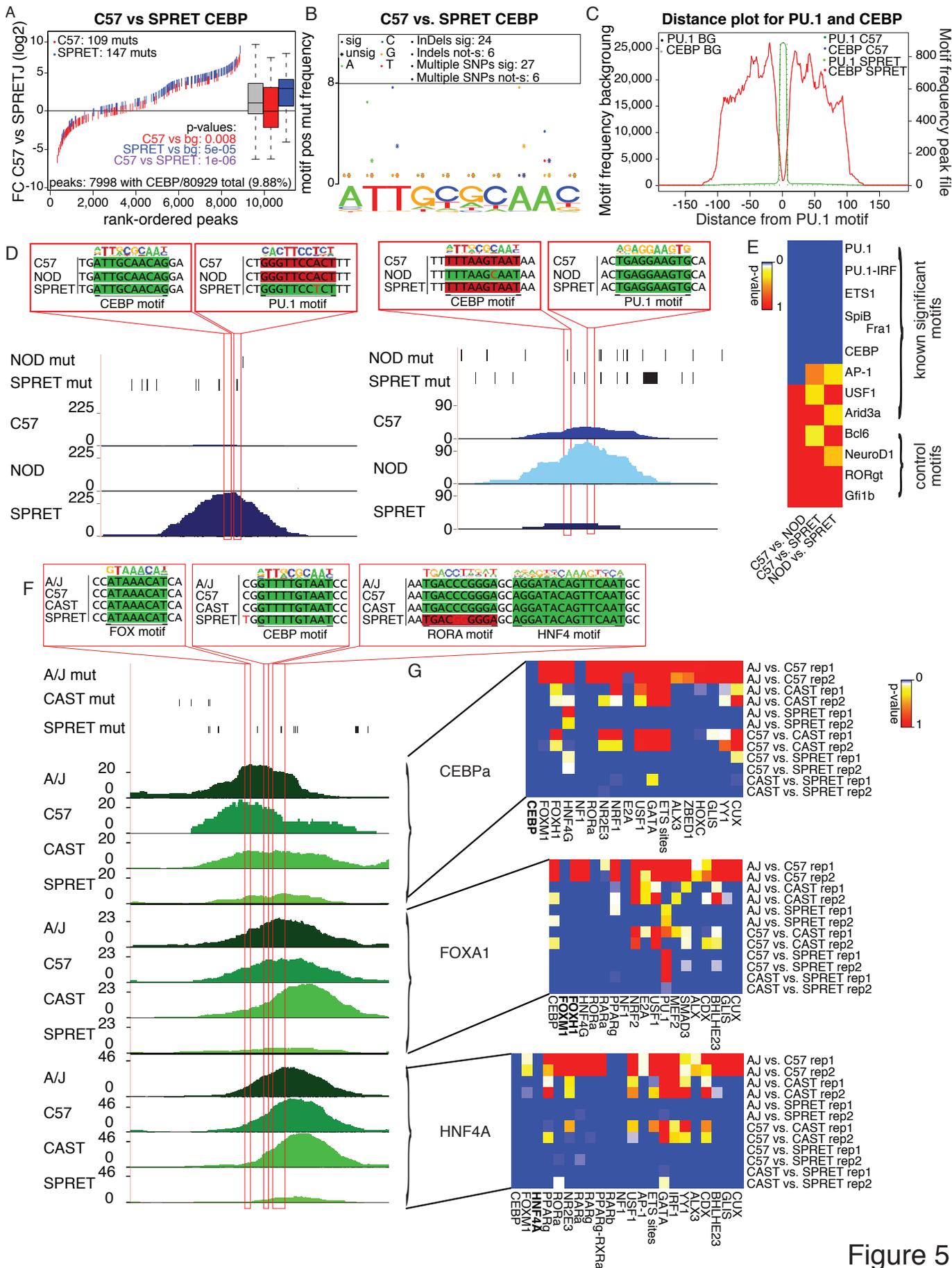
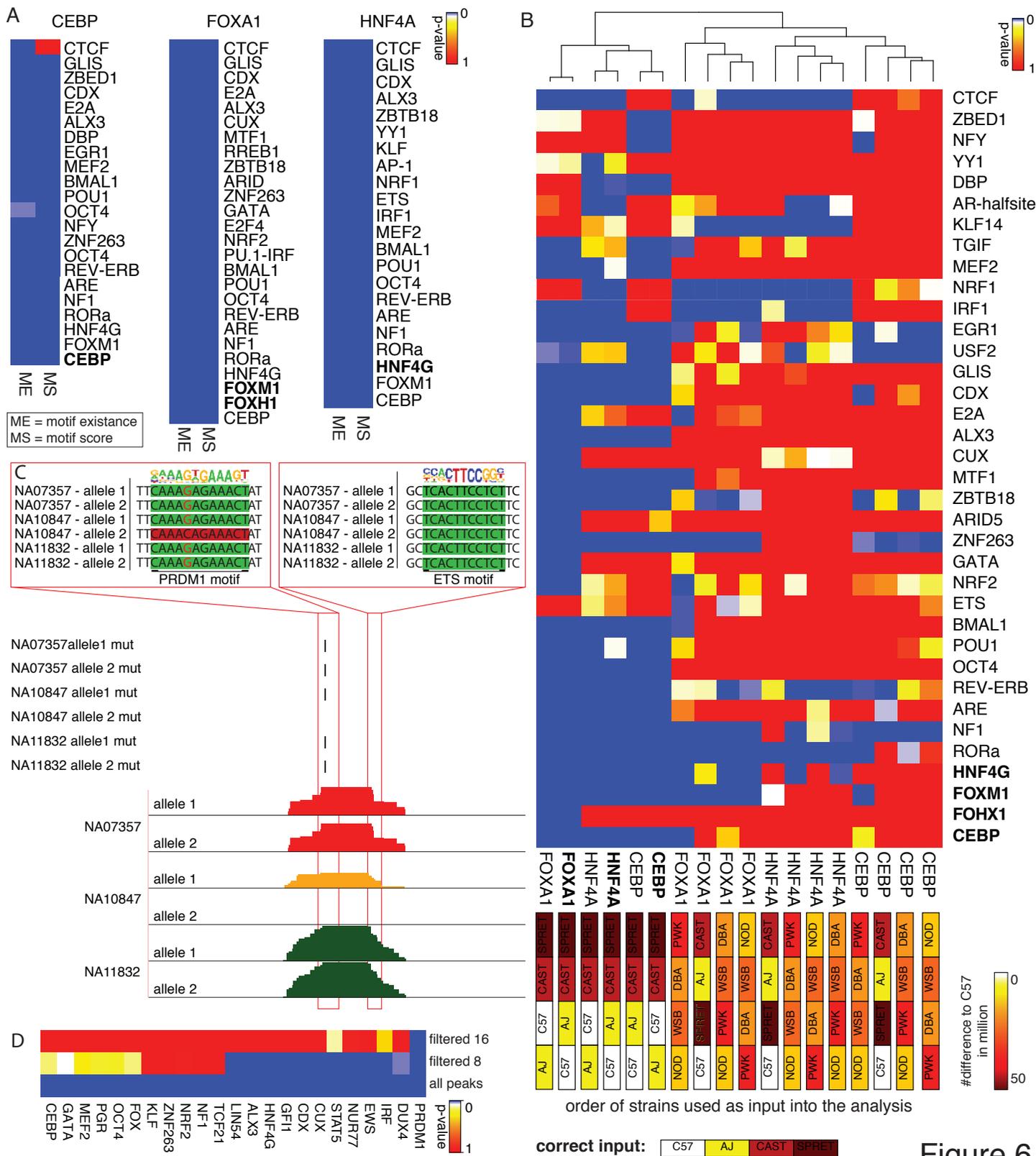


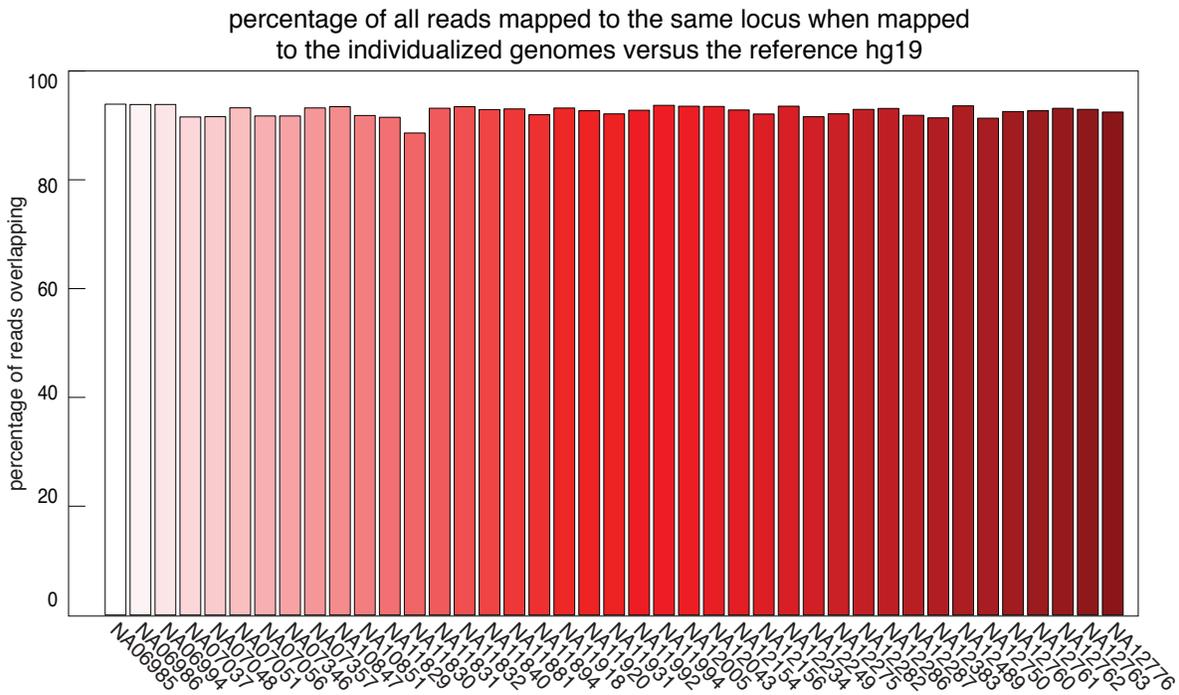
Figure 5



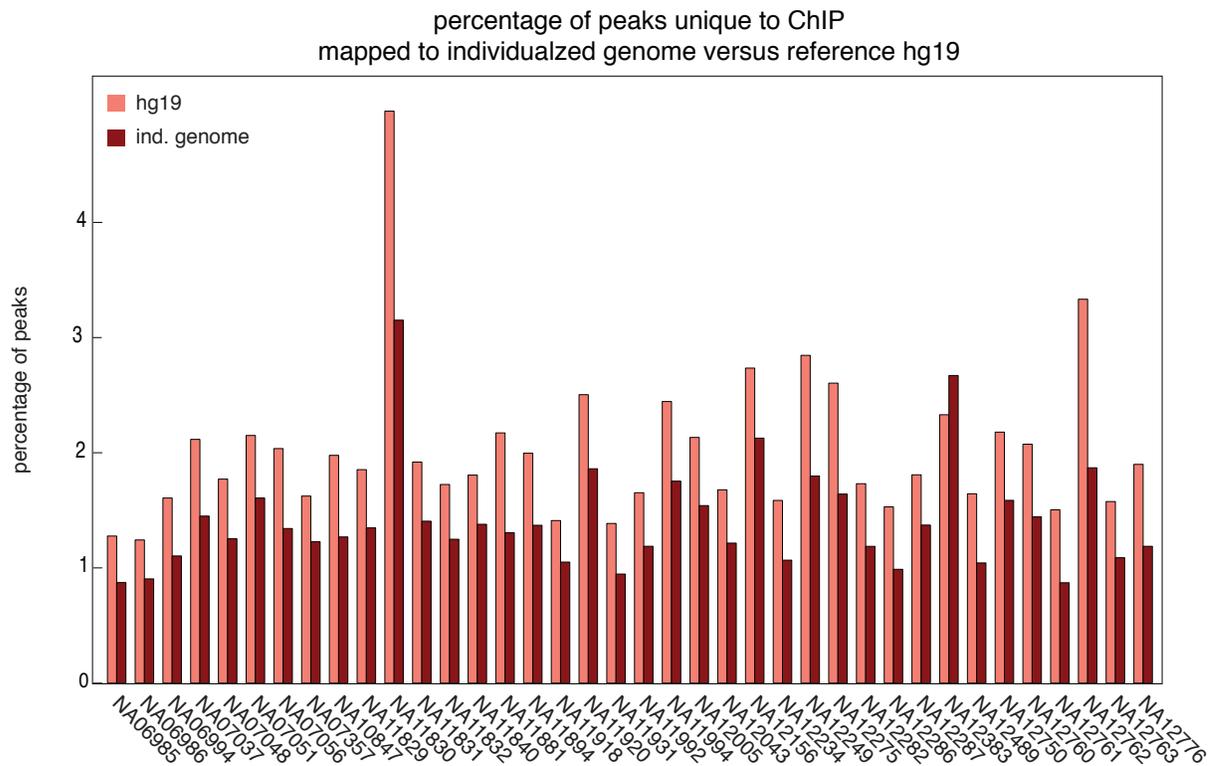
| Strain comparison        | #SNPs      | #InDels   |
|--------------------------|------------|-----------|
| C57BL/6J vs. NOD/ShiLtJ  | 4,734,324  | 272,463   |
| C57BL/6J vs. SPRET/EiJ   | 40,757,582 | 2,206,269 |
| NOD/ShiLtJ vs. SPRET/EiJ | 41,033,145 | 2,302,767 |

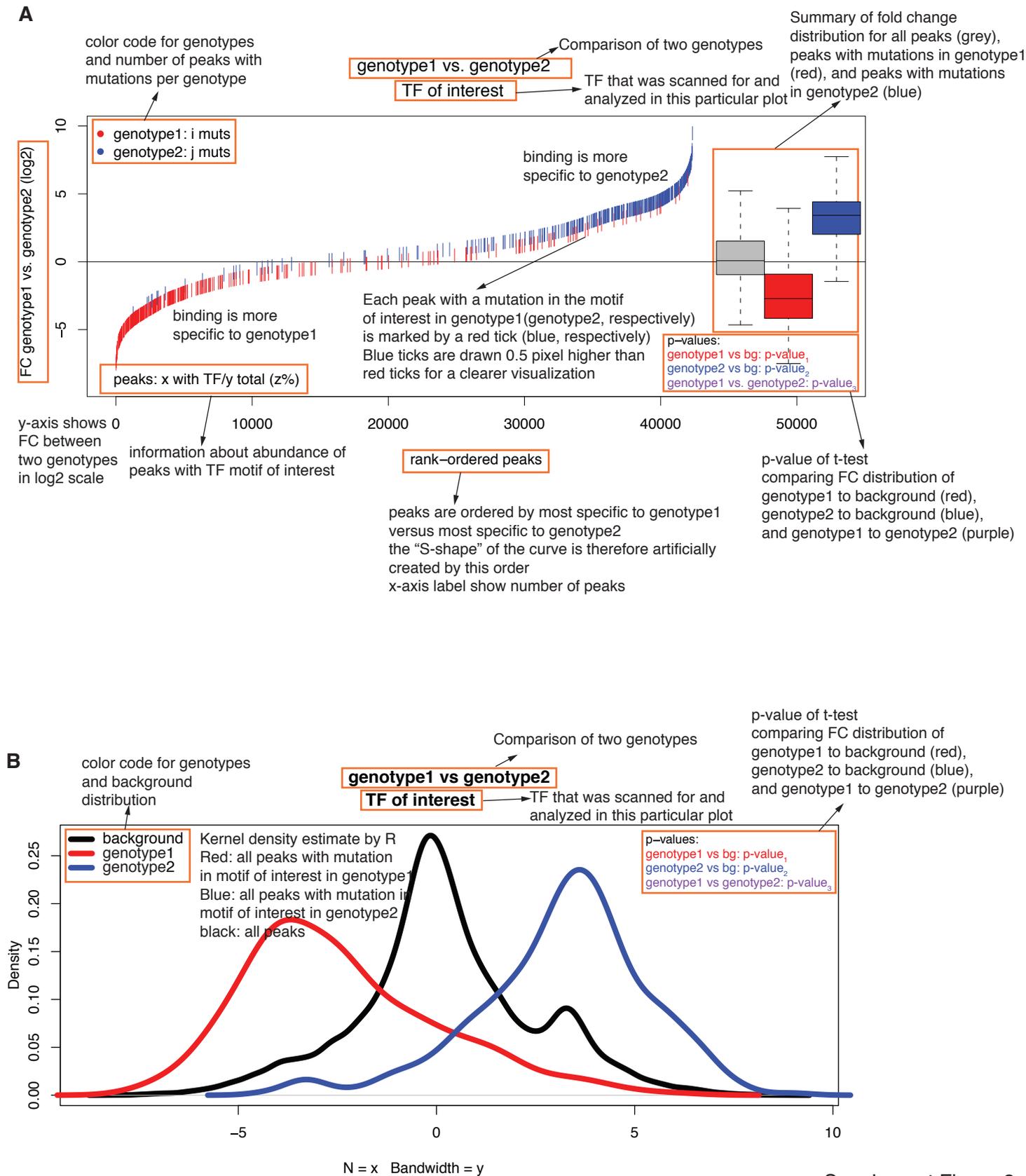
| Private variation per strain | #SNPs      | #InDels   |
|------------------------------|------------|-----------|
| NOD/ShiLtj                   | 2,474,126  | 160,882   |
| SPRET/Eij                    | 38,490,407 | 2,101,665 |

**A**

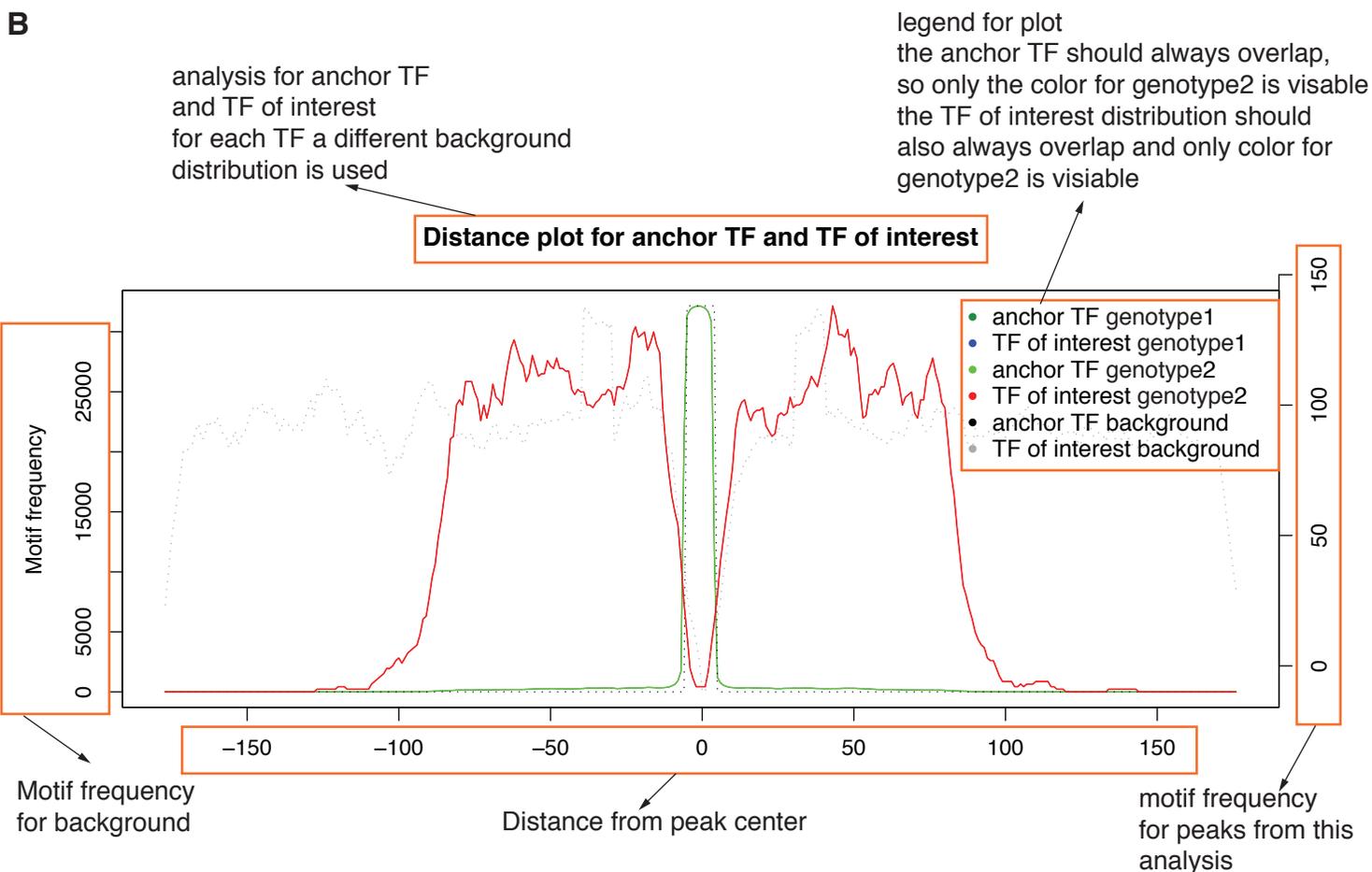
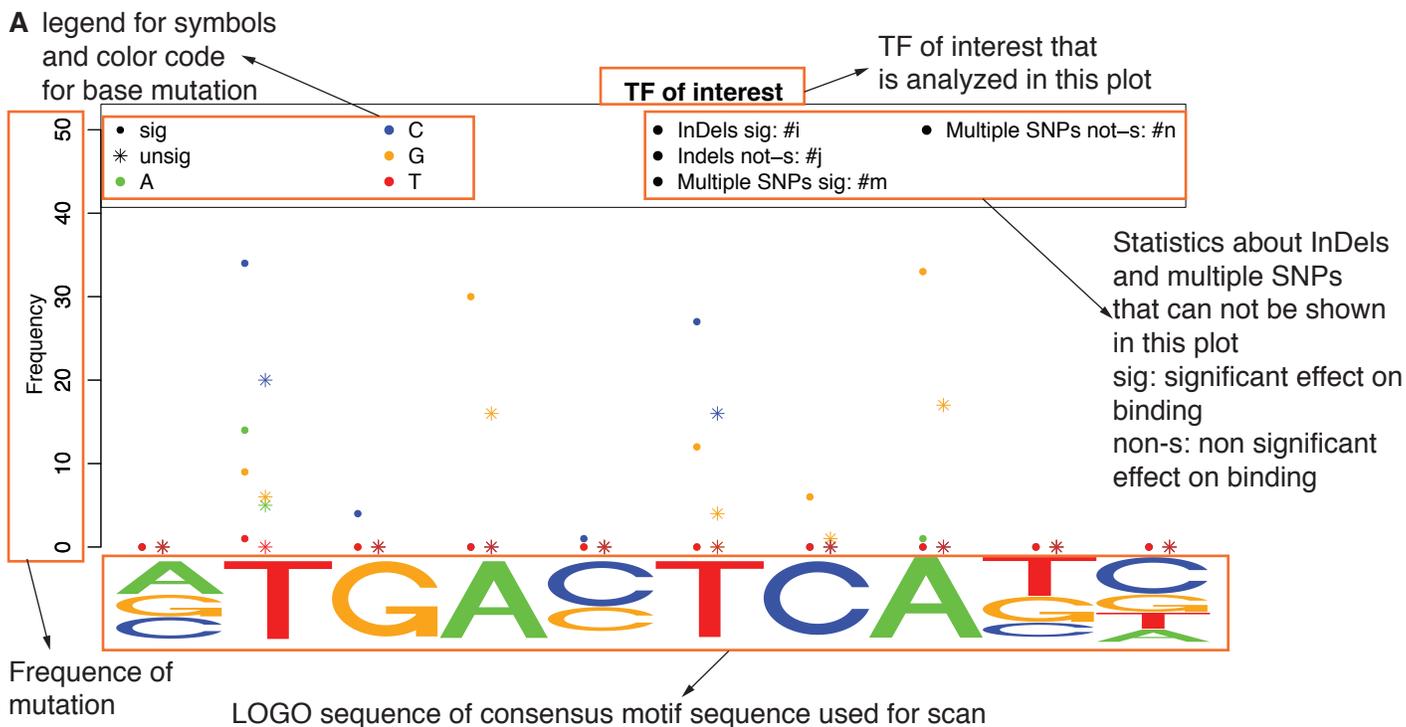


**B**





Supplement Figure 2





# Chapter 4

## Transcription Factor Landscapes in Macrophages from Genetically Diverse Mice Reveal Connected Regulatory Domains

**Verena M. Link**, Sascha H. Duttke, Hyun B. Chun, Inge R. Holtman, Emma Westin, Marten A. Hoeksema, Yohei Abe, Dylan Skola, Casey E. Romanoski, Jenhan Tao, Gregory J. Fonseca, Ty D. Troutman, Nathanael J. Spann, Tobias Strid, Mashito Sakai, Miao Yu, Rong Hu, Rongxin Fang, Dirk Metzler, Bing Ren, and Christopher K. Glass

manuscript under review in *Cell*



# Transcription Factor Landscapes in Macrophages from Genetically Diverse Mice Reveal Extensive Connected Regulatory Domains

Verena M. Link<sup>1,2</sup>, Sascha H. Duttke<sup>1,\*</sup>, Hyun B. Chun<sup>1,\*</sup>, Inge R. Holtman<sup>1,3</sup>, Emma Westin<sup>1</sup>, Marten A. Hoeksema<sup>1</sup>, Yohei Abe<sup>1</sup>, Dylan Skola<sup>1</sup>, Casey E. Romanoski<sup>4</sup>, Jenhan Tao<sup>1</sup>, Greg Fonseca<sup>1</sup>, Ty D. Troutman<sup>1</sup>, Nathanael Spann<sup>1</sup>, Tobias Strid<sup>1</sup>, Mashito Sakai<sup>1</sup>, Miao Yu<sup>5</sup>, Rong Hu<sup>5</sup>, Rongxin Fang<sup>5</sup>, Dirk Metzler<sup>2</sup>, Bing Ren<sup>1,5</sup>, and Christopher K. Glass<sup>1,6,7</sup>

<sup>1</sup> *Department of Cellular and Molecular Medicine, School of Medicine, University of California, San Diego, La Jolla, CA, USA*

<sup>2</sup> *Faculty of Biology, Division of Evolutionary Biology, Ludwig-Maximilian University of Munich, Germany*

<sup>3</sup> *Department of Neuroscience, Section Medical Physiology, University of Groningen, University Medical Center Groningen, Netherlands*

<sup>4</sup> *Department of Cellular and Molecular Medicine, University of Arizona College of Medicine, Tucson, AZ, USA*

<sup>5</sup> *Ludwig Institute for Cancer Research, La Jolla, CA, USA*

<sup>6</sup> *Department of Medicine, University of California, San Diego, La Jolla, CA, USA*

<sup>7</sup> *Corresponding author: [ckg@ucsd.edu](mailto:ckg@ucsd.edu)*

*\* Equivalent second co-author contributions*

## **Abstract**

**Non-coding genetic variation is a major driver of phenotypic diversity and allows investigation of mechanisms that control gene expression. Here, we systematically investigate effects of >50 million variations from five strains of mice on mRNA, nascent transcription, transcription start sites and transcription factor binding in resting and activated macrophages. We observe substantial differences in gene expression that are associated with distinct molecular pathways. Evaluation of genetic variation provides evidence for roles of ~100 transcription factors in shaping the binding of macrophage lineage-determining factors. Unexpectedly, a substantial fraction of the strain-specific binding of these factors cannot be explained by local mutations. Integration of genomic features with assays of chromatin interactions provides evidence for hundreds of connected *cis*-regulatory domains that are associated with differences in transcription factor binding and gene expression. This system and the >250 data sets establish a substantial new resource for investigation of how genetic variation affects cellular phenotypes.**

## **Introduction**

Mammalian organisms are composed of several hundred cell types that share a common genome. The development and function of each cell thus requires appropriate selection of promoter and enhancer elements that regulate their specific programs of gene expression (Heinz et al., 2015; Levine, 2010; Shlyueva et al., 2014). Genome-wide assessment of chromatin features specific to enhancers and promoters across many cell types and tissues in humans and mice revealed hundreds of thousands of enhancer-like regions, with any particular cell type exhibiting on the order of 20,000-30,000 such elements (Andersson et al., 2014; Roadmap Epigenomics et al., 2015). The general question of how each cell type selects its particular repertoire of transcriptional regulatory elements is therefore central to understanding the corresponding cell's development and functions.

Investigation of mechanisms underlying the selection of cell-specific enhancers indicate key roles of so-called pioneering factors that have the potential to recognize their binding motifs in

the context of closed chromatin (Soufi et al., 2015) and can therefore function as lineage determining transcription factors (Heinz et al., 2015; Iwafuchi-Doi and Zaret, 2014). However, such factors only bind to a small fraction of their corresponding recognition motifs that are present within the genome, and the same pioneering factor can bind to different genomic regions in different cell types (Heinz et al., 2010; Jin et al., 2011; Roadmap Epigenomics et al., 2015). Therefore, additional mechanisms are required to specify their DNA binding patterns in each cell type.

Studies in macrophages provided evidence for a collaborative/hierarchical model for enhancer selection driven by macrophage-restricted combinations of lineage-determining factors that include PU.1 and C/EBP $\alpha/\beta$  (Heinz et al., 2010). In these studies, the pioneering functions of PU.1 and C/EBPs were suggested to be dependent on collaborative interactions at sites in the genome containing closely spaced (i.e. < ~150 bp) binding motifs for each factor. Collaborative binding of PU.1 and C/EBPs was supported by studies investigating effects of genetic variation in macrophages derived from C57BL/6J and BALB/cJ mice, in which hundreds of strain-specific binding sites for each factor were observed (Heinz et al., 2013). Strain-specific mutations in the recognition motif for PU.1 that resulted in loss of PU.1 binding in that strain also resulted in loss of nearby C/EBP $\beta$  binding, despite intact C/EBP recognition motifs, and vice versa. While these studies provided support for a collaborative model of enhancer selection, they also indicated that the majority of strain-specific binding of PU.1, C/EBP and P65 could not be explained by mutations in their respective binding motifs (Heinz et al., 2013). This discrepancy raised new questions regarding the determinants of the pioneering functions of factors such as PU.1 and C/EBP and the extent to which the binding of these factors is influenced by nearby versus distant genomic elements.

Additional motivation for elucidating mechanisms underlying general and cell-specific gene expression derives from the importance of non-coding genetic variation as a determinant of phenotypic diversity. While mutations in amino acid coding sequences that result in altered or loss of function of essential proteins are well-established causes of monogenic diseases, they

account for a small fraction of overall genetic variation and inherited risk of disease. Genome-Wide Association Studies (GWAS) of diverse phenotypic traits, including disease risk, consistently identify between 80 to 90% of the significant variants to reside in non-coding regions of the genome, implying regulatory functions (Farh et al., 2015; Genomes Project et al., 2015; Hindorff et al., 2009; Maurano et al., 2012). Consistent with this, genetic variation has been demonstrated to directly affect transcription factor binding as an underlying determinant of altered chromatin states and gene expression (Farh et al., 2015; Heinz et al., 2013; Kasowski et al., 2010; Kilpinen et al., 2013; McDaniell et al., 2010; McVicker et al., 2013; Reddy et al., 2012). Despite this substantial progress, interpretation of non-coding genetic variation remains challenging for most loci and it is not yet possible to accurately predict gene expression from genotype.

Here we exploit genetic variation provided by five diverse inbred strains of mice to query mechanisms underlying transcription factor binding and function. To eliminate confounding effects of strain-specific differences in tissue environments that are known to influence macrophage phenotypes *in vivo* (Gosselin et al., 2014; Lavin et al., 2014), we performed studies in bone marrow derived macrophages (BMDMs), in which each strain-specific population of macrophages was established using an identical M-CSF-dependent differentiation protocol. To assess the impact of genetic variation on signal-dependent transcription factor binding and function, BMDMs were activated with Kdo2 lipid A (KLA), a highly specific TLR4 agonist (Raetz et al., 2006). This system has several experimental strengths. BMDMs are derived from readily available inbred strains of mice in which all loci are homozygous and whole genome sequences are available (Keane et al., 2011). The five strains selected provide genetic variation ranging between ~4.5 million SNPs + InDels, similar to differences between any two individuals, to ~50 million SNPs + InDels, on the order of all such common variants in the human population (Genomes Project et al., 2015). The selected strains of mice have been extensively phenotyped and exhibit marked phenotypic diversity (Bogue et al., 2017; Lusic et al., 2016). Sufficient cells can be obtained for a broad range of genomic, proteomic, lipidomic and functional assays and can be readily derived from crosses between strains and genetically modified mice. Using this

experimental system, we systematically evaluate the effects of SNPs and InDels on gene expression, nascent transcription, open chromatin, transcription factor binding and histone modifications associated with primed and/or active regulatory elements in resting and activated primary macrophages (Figure 1A), generating more than 250 genome-wide data sets. In parallel, we develop a general computational pipeline for assessing the significance of motif mutations on transcription factor binding.

Given the diverse roles of macrophages in immunity, tissue homeostasis and disease (Hirsch et al., 2012; Malm et al., 2015; Moore and Tabas, 2011; Nguyen et al., 2006; Williams et al., 2016), the experimental system, accompanying data and analytical pipeline provide a significant new resource for investigation of the transcriptional mechanisms underlying macrophage gene expression and their context-specific functions. We observe striking effects of genetic variation on nascent and mature mRNA expression that predict distinct macrophage phenotypes in each mouse strain. Differences in gene expression are associated with order of magnitude greater differences in transcription factor binding. Leveraging these differences, we provide evidence supporting the hypothesis that the genomic binding patterns of macrophage lineage determining transcription factors are influenced by a large fraction of other transcription factors expressed in these cells. Although most variation in nascent transcription and transcription factor binding is consistent with *cis*-regulation, thousands of strain-specific differences in transcription factor binding and chromatin features cannot be explained by local mutations. Integration of strain-specific genomic features indicate that they frequently reside in highly interconnected clusters that are associated with strain-specific gene expression, suggesting a domain-wide regulatory environment that influences transcription factor binding and function.

## Results

### Genetic variation and mRNA expression

The effect of genetic variation on polyadenylated (polyA) RNA expression was assessed in BMDMs derived from C57BL/6J (C57), BALB/cJ (BALB), NOD/ShiLtJ (NOD), PWK/PhJ (PWK) and SPRET/EiJ (SPRET) (Figure S1A) by RNA-seq under basal conditions (notx) and following stimulation with KLA for 1 hour. A minimum of two biological replicates was performed for each mouse strain and condition, with replicates being highly correlated (Figure S1B, S1C). Pairwise comparisons of BALB, NOD, PWK and SPRET BMDMs to C57 BMDMs indicate a progressive increase in differential gene expression in resting cells (Figure 1B). Using a 4-fold cutoff and false discovery rate (FDR) of 0.01, divergent gene expression ranged from 112 RNA transcripts in the BALB x C57 comparison (4,158,340 SNPs and 240,320 InDels) to 1,438 RNA transcripts in the SPRET x C57 comparison (40,757,582 SNPs and 2,206,269 InDels) (Table S1). Shared and private differently regulated genes are shown in Figure 1C. These data show that BMDMs, despite being maintained in identical environments, exhibit a remarkable diversity of gene expression. For example, ~10% of the expressed transcriptomes vary at least 4-fold (FDR < 0.01) without perturbation by KLA. Upon KLA treatment a total of 129 genes were regulated in any strain by greater than 4-fold at a FDR threshold of 0.01 (Table S1). Although differences in the expression response across strains increased with genetic variation, there were very few qualitative differences in the response to KLA at 1h. This suggests strong conservation of the initial TLR4 response between BMDMs of all strains (Figure 1C, D). The relationship of transcriptional variation as a function of SNPs + InDels for all ten pairwise comparisons is indicated in Figure 1E. Notably, differential gene expression exhibits a sharp rise for pairwise comparisons between C57, BALB and NOD and comparisons of these strains with PWK. Addition of genes differentially regulated in SPRET leads to a further but non-linear increase in divergent gene expression.

Clustering the RNA-seq data segregates the samples by strain, with KLA treatment being a secondary determinant (Figure S1C). WGCNA analysis (Langfelder and Horvath, 2008) identifies numerous differentially expressed gene modules, many of which are significantly enriched for

genes associated with specific biological functions including autophagy, metabolism, cell cycle and interferon signaling (Figure 1C, Figure S1D). To validate one such module, we tested the prediction that macrophages derived from SPRET mice would exhibit defects in the type I interferon response following TLR4 ligation in comparison to C57 macrophages. C57 and SPRET BMDMs in replicates were treated with KLA for 6 hours, at which time a robust Type I interferon response is observed in C57 BMDMs at the level of mRNAs. This is shown by an expression heat map of 46 known interferon stimulated genes (ISGs) exhibiting >4-fold induction in C57 BMDMs (Figure 1F). A subset of these genes, exemplified by *Mx1* and *Mx2*, are fully induced in SPRET BMDMs. However, 37 of the 46 ISGs exhibit >2-fold less expression in SPRET BMDMs following KLA treatment, exemplified by *Ccl5* and *Ccl2*, and 40 of these ISGs exhibit > 2-fold less expression in SPRET BMDMs under basal conditions. These findings suggest that differences in the basal level of activity of the Type I IFN pathway in these macrophages determine its overall responsiveness to stimulation.

### **Effect of genetic variation on nascent transcription**

We performed whole genome run-on analysis coupled to deep sequencing (GRO-seq) (Core et al., 2008) to more directly assess the effects of genetic variation on nascent transcription in BMDMs from each strain under control conditions and 1h after treatment with KLA. A minimum of two biological replicates was performed for each mouse strain and condition (Table S2). As in the case of polyA RNA transcripts, pairwise comparisons of nascent gene body transcripts exhibited increasing but non-linear strain-specific differences with increasing degrees of genetic variation (Figure 2A). Strain-specific GRO-seq signal is exemplified for *Igf1* in Figure 2B. KLA treatment induced GRO-seq signal at 939 genes >4-fold with an FDR <0.01 in BMDMs from at least one strain, and repressed 452 genes. The larger effect of KLA signaling on nascent transcripts is consistent with the 1 hour time point being relatively early in the overall response to TLR4 signaling, such that many induced or repressed mRNAs have not yet reached changes required to meet the stringent cutoffs for selection. Clustering the GRO-seq data sets indicates that the KLA treatment response is the dominant variable, in contrast to strain background for polyA transcripts (Figure S2A). As in the case of the RNA response, very few genes exhibited

divergent responses to KLA at the one hour time point, exemplified for comparisons of BALB and SPRET to C57 in Figure 2C.

To define sites of transcription initiation, we performed 5'GRO-seq, which selects for the capped ends of nascent transcripts and enables base pair resolution of RNA polymerase (Pol) II start sites (Lam et al., 2013). The relationship of 5'GRO-seq to GRO-seq and H3K27ac at the *Igf1* locus is illustrated in Figure 2B. In addition to genic start sites, GRO-seq and 5'GRO-seq also quantify RNA generated at enhancers (eRNAs) (Hah et al., 2011; Kaikkonen et al., 2013; Lam et al., 2013), observed upstream of the *Igf1* transcription start site (TSS) in Figure 2B. Because polyA RNAs can be initiated from different promoters in a cell-specific manner (Noguchi et al., 2017), methods such as 5'GRO-seq are required to annotate genic start sites within a given cell type. Using 5'GRO-seq, we find about 30% of mRNAs to be initiated further than 50 bp from RefSeq annotated start sites, suggesting utilization of alternative core promoter elements in macrophages and/or technical differences with respect to prior methods used for start site annotation (Figure S2B, S2C). Using 5'GRO-seq annotated TSS, we investigated the extent to which differences in nascent and polyA RNA levels could be explained by mutations within the core promoter (-30 to +20 bp from the TSS), the proximal promoter region (-300 to +50 bp from the TSS) and more distal elements. Figure 2D shows the percentage of core promoter regions containing mutations that exhibit >4-fold differences in gene expression against the percentage of core promoter regions with mutations that do not exhibit strain-specific gene expression. For comparisons of C57 versus PWK or SPRET, 20 - 40% of core promoters contain mutations regardless of variation in gene expression. For comparisons of C57 versus BALB or NOD, the mutation frequency in core promoters of differentially expressed genes is 10-15%, in comparison to 5-10% in core promoters of similarly expressed genes. Using the 2-sample test for equality of proportions with continuity correction, these values are significantly similar ( $p$ -value  $< 1e-12$  for C57 vs. BALB,  $p$ -value  $< 1e-3$  for C57 vs. NOD). For the proximal promoter, ~40% of differently regulated genes in C57, BALB and NOD BMDMs contain sequence variants in comparison to a 25-30% mutation frequency in the proximal promoters of similarly expressed genes ( $p$ -value  $< 2.2e-16$  for C57 vs. BALB and  $p$ -value =  $6.57e-07$  for C57 vs. NOD, 2-sample test

for equality of proportions with continuity correction) (Figure S2D). Therefore, the majority of differences in gene expression cannot be explained through variation within the core or proximal promoter sequences.

We next established the relative contributions of local versus distal genetic variation on differential expression of nascent and polyA RNA transcripts by analyzing BMDMs derived from F1 crosses of C57, PWK and SPRET mice. In this context, differences between inbred parental strains are likely a *cis* effect if the difference between the parental alleles is maintained within the F1 hybrid (where there is one copy of each parental allele). In contrast, if alleles that are differentially expressed in the parental strains become similarly expressed in the F1 animal, we consider differential regulation in the parental strains to be mainly due to *trans* effects. A plot of fold difference in allele-specific reads of nascent gene body transcripts for a cross of C57 and SPRET mice versus the fold difference in the parental strains is illustrated in Figure 2E. A corresponding plot for a cross of C57 and PWK mice is shown in Figure S2E. These comparisons indicate that about 80% of the differences in nascent transcripts are determined in *cis* (Figure S2F). When relaxing the threshold for *cis*-regulation to a difference of 1.5-fold, 90% of genes are considered *cis*-regulated. Examples illustrating primarily *cis*-regulation are provided for *Npy* and *Plag2g7* for C57 versus SPRET BMDMs in Figure 2F. Collectively, these findings indicate that strain-specific gene expression primarily results from *cis*-variation that is distal from core and proximal promoter elements.

### **Effect of genetic variation on LDTF binding and chromatin signatures**

ChIP-seq experiments for PU.1, C/EBP $\beta$ , CJUN and the P65 component of NF $\kappa$ B under resting and KLA-stimulated conditions were performed in BMDMs derived from each mouse strain. Regions of open chromatin were assessed using the Assay for Transposase-Accessible Chromatin, or ATAC-seq (Buenrostro et al., 2013), and ChIP-seq of dimethylation of lysine 4 on histone 3 (H3K4me2) and acetylation of lysine 27 on histone 3 (H3K27ac) were used as surrogates of primed/active and active regulatory regions, respectively (Creyghton et al., 2010; He et al., 2010). The Irreproducible Discovery Rate (IDR) method (Li et al., 2011) was used to

define highly reproducible peaks across replicates for the CHIP-seq experiments evaluating binding of PU.1, C/EBP $\beta$ , CJUN, P65 and the ATAC-seq data. Because IDR is not applicable to histone modification CHIP-seq experiments, DESeq2 was used (Love et al., 2014) and regions similar between replicates ( $p$ -value < 0.001) were kept. Examples of biological replicates and correlation heat maps are provided in Figure S3A and S3B. The numbers of features identified for each of these assays in the five strains of BMDMs under control and KLA-treated conditions are provided in Table S3.

The effect of genetic variation on H3K27ac CHIP-seq regions is illustrated for comparisons of BALB and SPRET BMDMs to C57 BMDMs in Figure 3A and corresponding comparisons for H3K4me2 are shown in Figure S3C. As in the case of polyA and nascent gene body RNA, variation in these features scale with genetic diversity, but to a greater degree. Extension of these comparisons for ATAC-seq defined regions is illustrated in Figure 3B. Strikingly, variation in IDR-defined open chromatin regions occurs to an order of magnitude greater extent than polyA or nascent gene body RNA expression. Genomic regions exhibiting at least 4-fold differences in ATAC-seq tag counts range from ~1650 for the comparison of C57 and BALB to ~19,700 for the comparison of C57 to SPRET. We performed de novo motif analysis of distal ATAC-seq peaks (>3000 bp from a TSS) associated with H3K27ac, corresponding to potential enhancer elements, in resting BMDMs from each strain. This analysis returned a consistent pattern of motifs for PU.1, AP-1 and C/EBP as the most highly enriched motifs, followed by motifs for USF, RUNX and a composite PU.1-IRF motif (Figure 3C). We then defined the intersections of CHIP-seq peaks for PU.1, CJUN and C/EBP $\beta$  with distal ATAC-seq peaks associated with H3K27ac. These three factors, alone or in combination, were found to occupy ~85% of the putative distal regulatory regions of BMDMs in each strain, exemplified for C57 BMDMs in Figure 3D. H3K27ac CHIP-seq data was also used to define super enhancers, which are genomic regions that are occupied by a high density of active transcriptional regulatory elements and are associated with genes required for cellular identity (Whyte et al., 2013). In comparison to H3K27ac patterns as a whole, H3K27ac signal at super enhancers was generally concordant across strains (Figure 3E).

The impact of genetic variation on transcription factor binding is illustrated for PU.1 in Figure 3F. As in the case of H3K27ac regions and ATAC-seq peaks, we observed a striking graded progression of strain-specific binding as a function of extent of SNPs and InDels. Strain-specific binding of PU.1 ranged from ~3,800 peaks comparing BALB and C57 to more than 23,000 peaks comparing SPRET to C57, the latter number representing nearly one quarter of the IDR-defined PU.1 binding sites. Similar patterns were observed for C/EBP $\beta$ , CJUN and P65 (Figure S3D-S3F). Thus, variation in transcription factor binding greatly exceeds variation in gene expression. To quantify the extent of *cis* versus *trans*-regulation of binding for transcription factors, PU.1 ChIP-seq experiments were performed in two F1 strains (PWK x C57 and SPRET x C57) in resting BMDMs and after KLA stimulation for 1 hour. Directly comparing fold change of allele-specific reads between parents and their corresponding fold change in the F1 strains indicates that more than 70% of the peaks follow the parental pattern and are therefore considered to be *cis*-regulated (Figure 3G, Figure S3G, S3H), consistent with findings for nascent gene body RNA.

We next assessed the extent to which strain-similar and strain-specific ATAC-seq and ChIP-seq peaks exhibit local genetic variation ( $\pm$  150 bp of the peak center). We considered homozygous variants that passed a stringent quality filter as well as sequence variants meeting lower stringency criteria (see extended methods for details). For comparisons of C57 to BALB and NOD BMDMs, 20-22% of the strain-similar peaks contain SNPs and/or InDels (Figure 3H) (22% - 25% for the more lenient definition, respectively). These frequencies increase to 71-96% in comparisons involving PWK and SPRET BMDMs, indicating that the great majority of local sequence variants are silent. For strain-specific peaks defined by a >4-fold cut-off, comparisons of C57 to BALB and NOD BMDMs indicated that only 52-59% of the ATAC-seq peaks and 64-78% of the PU.1 ChIP-seq peaks contain local variants. Extending the strain-specific cut-off to >8 fold, 76-77% of ATAC-seq peaks and 76-81% of PU.1 ChIP-seq peaks contain local SNPs and/or InDels (Figure 3H). Similar relationships are observed for C/EBP $\beta$ , CJUN and P65, although the fractions of strain-specific peaks containing mutations are somewhat lower (Figure S3I). While these findings are consistent with the expected effects of genetic variation on transcription factor

binding and open chromatin, they also indicate that substantial fractions of strain-specific differences in these features cannot be explained by local SNPs or InDels. Further examination of structural variants indicates that these regions frequently explain extreme (i.e., all or none) strain-specific differences (Figure S3J, exemplified for C57 versus SPRET) for regions without mutations, but overall account for only about 2% - 4% of the features with mutations exhibiting >4-fold differences between strains (Figure S3K).

### **Inference of an extensive network of collaborative transcription factors**

Prior observations that mutations in PU.1 motifs alter the binding of nearby C/EBP $\beta$  and vice versa (Heinz et al., 2013) provide the basis for a general approach for discovery of collaborative binding partners by systematic analysis of effects of local motif mutations. Here, we qualitatively advance this strategy by leveraging the diversity of five strains of mice, simultaneously assessing four transcription factors under basal and stimulated conditions, and developing a new Mutational Analysis of Regulatory Genomic Elements (MARGE) (Link et al., 2018) software pipeline to comprehensively evaluate the relationship of motif mutations with transcription factor binding (see extended methods for details). Using an input of normalized ChIP-seq or ATAC-seq data from genetically diverse samples, a library of motifs to query, and corresponding genomic sequence for each sample, MARGE utilizes a general linear mixed model to calculate a *p*-value for whether mutations in a particular motif are significantly associated with differential transcription factor binding, chromatin accessibility or transcription initiation. Because many motifs in existing databases are highly redundant and are recognized by the same factor or family of factors, we also generated a non-redundant motif library by clustering all motifs currently resident in the JASPER 2016 non-redundant database and combining motifs with a Pearson correlation coefficient of greater than 0.9. After further manual curation this exercise yielded 230 motifs that were used for analysis.

We applied MARGE to systematically identify motifs for which disruptions due to mutations were highly correlated with strain-specific binding of PU.1, C/EBP $\beta$ , CJUN or P65 under control and KLA treatment conditions. A heat map for a subset of the most highly significant motifs is

illustrated in Figure 4A, along with the corresponding motif name. The complete set of significant motifs is illustrated in Figure S4A. In total, mutations in 80 motifs were found to be associated with strain-specific binding of PU.1, C/EBP $\beta$ , CJUN and/or P65 at a  $p$ -value of less than  $1e^{-10}$  (Table S4). These motifs could in turn be associated with more than 100 transcription factors expressed  $> 1$  transcript per million (TPM) (Table S5), providing genetic evidence for functional roles of a large fraction of the TFs expressed in BMDMs as collaborative partners that drive the selection of potential regulatory elements.

In addition to consensus PU.1, C/EBP, and AP-1 motifs, motifs for related factors were also amongst the most highly significant motifs identified. These included motifs for several ETS factors (e.g., GABPA, ELK, ELF), motifs for factors related to C/EBP (e.g. DBP) and motifs for factors related to CJUN (e.g., ATF, MAF). These motifs are more difficult to interpret because in addition to being high affinity binding sites for the corresponding factors, they are also lower affinity binding sites for PU.1, C/EBP $\beta$  and CJUN, respectively. Analysis of PU.1 peaks containing a single ETS motif that is not a consensus PU.1 motif indicated that mutations within these motifs significantly impact PU.1 binding, consistent with direct interactions with low affinity sites (Heinz et al., 2013). However, many binding sites for PU.1, C/EBP $\beta$  and CJUN contain multiple iterations of ETS, C/EBP and AP-1-like motifs. For example, about 40% of PU.1 peaks contain multiple iterations of related ETS motifs. When the motif with the highest score for a PU.1 motif is masked, mutations in the remaining ETS motifs remain significant as determinants of PU.1 binding. Similar relationships are observed for DBP motifs associated with C/EBP $\beta$  peaks and ATF/MAF motifs associated with CJUN binding. These findings are consistent with both homotypic and heterotypic interactions and underscore the potential complexity of combinatorial interactions between members of transcription factor families that recognize related binding motifs.

Single factor interaction networks for PU.1 under basal conditions and P65 under KLA-treatment conditions are illustrated in Figure 4B and Figure 4C, respectively. Corresponding networks for C/EBP $\beta$  and CJUN are illustrated in Figure S4B and Figure S4C. In these networks,

the node sizes represent the fraction of PU.1 or P65 binding sites that contain the corresponding motifs, and the edge thicknesses corresponds to the effect size of motif mutations. The PU.1 network illustrates 14 of the 48 motifs exhibiting  $p$ -value  $<1e^{-10}$  significance for PU.1 binding under no treatment conditions, while the P65 network illustrates 15 of the 60 motifs exhibiting  $p$ -value  $<1e^{-10}$  significance for P65 binding following KLA treatment. For highly related motifs (e.g., ETS factor motifs), the motif with the largest effect size is illustrated. In panel B, the PU.1 node in red represents the fraction of all PU.1 peaks (green) that contain motifs preferentially matching to the PU.1 motif consensus sequence. As expected, mutations in the PU.1 motif and related ETS motifs have the strongest effect sizes on PU.1 binding. Mutations in AP-1 motifs and C/EBP motifs have the next strongest effect sizes, consistent with prior studies (Heinz et al., 2013). In addition, mutations in motifs for more than a dozen different classes of transcription factors were significantly associated with strain-specific binding of PU.1, including RUNX, USF, DR2, DBP, MAF, MYB, NRF, and E2A motifs (Figure 4B and Table S4). Mutations in NF $\kappa$ B motifs have the strongest effect size on P65 binding, but the NF $\kappa$ B node size indicates that the majority of IDR-defined P65 peaks lack consensus NF $\kappa$ B motifs (Figure 4C). PU.1/ETS, AP-1 and C/EBP motifs exhibit the next strongest effects on P65 binding, but mutations in motifs for many other factors have significant effects, particularly NRF2, RUNX and DBP. While most significant motif mutations are associated with decreased PU.1 and P65 binding, mutations in E2A and ZEB motifs have the opposite effect (Figure 4B, C).

An integrated interaction network for PU.1, C/EBP $\beta$  and CJUN under no treatment conditions is illustrated in Figure 4D. In this network, node sizes are the average fractional overlap of the indicated motif with PU.1, C/EBP $\beta$  or CJUN peaks and edges are factor-specific effect sizes. This network illustrates largely dominant effect sizes of mutations in motifs for PU.1, C/EBP and AP-1 factors on each other's binding. In most cases, motif mutations affecting one factor affected all factors, although often with different effect sizes, exemplified for USF. However, some motifs exhibited specific correlations, such as the unique association of mutations in the KLF motif with CJUN binding (Figure 4D). Mutations in E2A motifs were associated with increased

binding of PU.1, C/EBP $\beta$  and CJUN, whereas mutations in ZEB motifs were associated with selective increases in binding of PU.1 and CJUN (Figure 4D).

From these analyses, it is possible to estimate the fraction of strain-specific binding of PU.1, C/EBP $\beta$ , CJUN and P65 attributable to local mutations. Taking the binding sites exhibiting > 4-fold differences in one or more strains, mutations in the motifs for PU.1, C/EBP, AP-1 and NF $\kappa$ B motifs were associated with 22, 10, 8 and 9%, of strain-specific variation, respectively (Figure 4E). Incorporating the additional motifs found significant by the MARGE analysis increases the percentage of strain-specific binding explained to 70, 50, 65 and 60%, respectively. The fraction of binding sites explained by MARGE motifs is further increased considering the 500 most differential binding sites for each factor (Figure S4D). However, even at these sites, more than 20% of strain specific binding of each factor remains unexplained.

The interaction networks inferred from motif mutations are based on genomic regions in which genetic variation results in strain-specific differences in transcription factor binding. To investigate the potential of RUNX, USF and NRF factors (all found to be highly significant in the MARGE analysis) to function as collaborative binding factors for PU.1, C/EBP $\beta$  and CJUN, we performed ChIP-seq assays for RUNX1 and USF2 in C57 BMDMs. In addition, we analyzed previous ChIP-seq data for NRF2 in C57 BMDMs (Eichenfield et al., 2016). In each case we observed broad overlap with the genomic binding locations of PU.1, C/EBP $\beta$  and CJUN, exemplified for overlaps between PU.1, RUNX1 and USF2 in Figure 4F. We further investigated the potential of these factors to play roles in driving the selection of the 20% of open regions of chromatin marked by H3K27ac that are not occupied by PU.1, C/EBP $\beta$ , or CJUN (Figure 3D). These analyses indicated that USF2 occupied more than 80% of these locations, primarily in association with RUNX1 (Figure 4G). Thus, nearly all putative enhancer elements in macrophages are marked by combinations of only four transcription factors: PU.1, C/EBP $\beta$ , CJUN and RUNX1. The relationship of mutations in RUNX1 binding motifs to binding of PU.1 and RUNX1 is exemplified in Figure 4H (See Figure S4E and S4F for NRF2 and USF2), supporting a role of RUNX1 (as well as NRF2 and USF2) as a collaborative binding partners of PU.1.

### Regional correlation of transcription factor binding

The observations that the majority of strain-specific differences in gene expression and transcription factor binding are in *cis* but that substantial fractions of strain-specific transcription factor binding sites cannot be explained by local mutations led us to investigate the possibility of cooperativity between distinct *cis*-regulatory elements. To investigate this, we calculated Pearson correlation coefficients (PCC) of normalized tag counts for transcription factor and ATAC-seq peaks across the 5 strains under basal and KLA treatment conditions. Heat maps of these values for ATAC-seq, PU.1, and C/EBP $\beta$  peaks along ~6 MB regions of chromosome 18 are illustrated in Figure 5A, B and C. In these plots, the axes represent sequential locations associated with the indicated feature, with the matrix values corresponding to correlation coefficients defined by the accompanying scale. Blocks of highly correlated peaks are observed to diverge from the diagonal. An example of a genomic region in the vicinity of the *Collec12* gene exhibiting strain-specific correlation of ATAC-seq, PU.1, C/EBP $\beta$ , CJUN, H3K27ac and GRO-seq features is illustrated in Figure 5D. Within the shaded region more than a dozen high confidence PU.1 peaks exhibit concordant changes in normalized tag counts across the five strains. Similar concordant changes are observed for ATAC-seq, C/EBP $\beta$ , CJUN, H3K27ac and GRO-seq data. For reasons described further below, we refer to these correlated regions as connected regulatory domains (CRDs).

As one approach to define CRDs, we considered the minimum number of consecutive features (e.g., ChIP-seq peaks) that exceed a specific PCC. CRDs for different minimal feature number and correlation coefficient are illustrated in Figure 5E. No sharp inflections are observed that provide a basis for optimization of parameters. CRDs defined by arbitrary cutoffs of a PCC of >0.8 for a minimum of 4 peaks result in 400 - 800 locally correlated regions for PU.1, C/EBP $\beta$ , CJUN and P65 ChIP-seq peaks and ~4000 locally correlated ATAC-seq regions, which contain various combinations of CRDs of the other factors (Table S6). For CRDs defined by an *n* of >4 peaks and correlation coefficient >0.8, the mean number of peaks within a correlated cluster was 6 for PU.1/ C/EBP $\beta$ , cJUN and P65, and 8 for ATAC-seq (Figure S5A) and the size of genomic

regions containing connected peaks ranged from 1kb to 5Mb with a mean length of 70kb – 200kb (Figure S5B). Based on the analysis of F1 hybrids, nearly all strain-specific PU.1 peaks associated with CRDs retain their parental allelic pattern in F1 BMDMs, indicating that such sites are primarily regulated in *cis* (Figure S5C).

CRDs capture clusters of peaks that are highly similar across all strain comparisons, as well as clusters of peaks that exhibit coordinated differences between strains, such as the example shown for *Colec12*. Correlated PU.1 peaks that exhibit preferential binding in C57 or SPRET BMDMs are associated with corresponding differences in expression of the nearest gene in comparison to CRDs that are strain-similar (Figure 5F). Similarly, correlated PU.1 peaks that exhibit significantly different binding in C57 or SPRET BMDMs ( $p$ -value  $< 2.2e-16$ ) are associated with corresponding significant differences in local 5'GRO-seq signal ( $p$ -value =  $2.7e-10$  for C57 specific set,  $p$ -value  $< 2.2e-16$  for SPRET specific set) in comparison to CRDs that are strain-similar ( $p$ -value = 0.72) (Figure 5F and Figure 5G), as well as significantly different gene expression of the nearest expressed gene ( $p$ -value =  $3.9e-4$  for C57 specific set,  $p$ -value =  $4.1e-4$  for SPRET specific set,  $p$ -value = 0.78 for common set). The same was observed for ATAC-seq peaks (Figure S5E). Overall, strain-specific CRDs were highly correlated with strain-specific patterns of 5'GRO-seq signal, exemplified for PU.1 CRDs in Figure 5G. In contrast, we observed almost no overlap between strain-specific CRDs and super enhancers (Figure S5D).

### **Identification of connected regulatory domains**

The observation of regionally correlated transcription factor binding, open chromatin, histone signatures and gene expression raised the question of the relationships of these regions to DNA methylation and chromatin organization. We determined the patterns of DNA cytosine methylation in BMDMs from C57 and SPRET mice by performing bisulfite sequencing (Hajkova et al., 2002). We observed differentially methylated regions at promoters and regulatory elements, as expected, but the overall patterns of DNA methylation were very similar between the two strains (Figure S6A), including at promoters of differentially expressed genes (Figure S6B, S6C) such as *Colec12*, *Npy* and *Igf1* (Figure S6D). As these BMDMs are derived from the

most divergent strains, differences in DNA methylation are unlikely to be major drivers or consequences of CRDs.

We next performed in situ Hi-C assays (Rao et al., 2014) in C57 and SPRET BMDMs to define maps of DNA interactions. A Hi-C contact matrix depicting normalized contact frequencies for C57 and SPRET BMDMs for chromosome 18 is indicated in Figure 6A (left). Overall there was a high degree of similarity, with the Eigenvalue of first principle component (PC1), correlated with active or inactive regions of chromatin, being nearly identical between strains. Strain-similar contact frequencies and Eigenvalues are illustrated in the vicinity of the *Spi1* locus, encoding PU.1 (Figure 6B, left). However, genomic regions were also observed exhibiting markedly different contact frequencies that correlated with strain-specific gene expression, exemplified by the *Colec12* locus (Figure 6B, right). The overall relationship of Eigenvalues for PC1 calculated for 100 kb windows (Figure S6E) indicates a few hundred regions of the genome in which the Eigenvalue reverses sign, as in the case of *Colec12* (Figure 6B right). A small fraction of these regions are associated with strain-specific differences in gene expression (Figure S6F). However, the great majority of differentially expressed genes reside in compartments associated with similar PC1 values in both strains.

To further investigate the relationship between correlated chromatin features and chromatin organization we identified topological associating domains (TADs) from these Hi-C data sets (Dixon et al., 2012). In brief, TADs are chromatin regions that are typically constrained by CTCF boundaries such that elements within the TADs have coordinated *cis*-regulatory activity. TADs were also highly similar in BMDMs of C57 and SPRET mice (Figure S6G), although there were frequently subtle differences in how adjacent domains were parsed (e.g., Figure 6A, right). Comparison with ATAC-seq and CHIP-seq data indicated that nearly all locally correlated chromatin features reside within TADs, exemplified by the grid lines in Figures 5A, B and C.

To further investigate spatial relationships of locally correlated features, we performed proximity ligation-assisted CHIP-seq (PLAC-seq) in C57, BALB, NOD, and SPRET-derived BMDMs

using H3K4me3 as anchor (Fang et al., 2016). From these data sets we defined all significant interactions observed in at least two strains as a consensus set. We investigated the overlap of significant interactions between and within ATAC-seq CRDs (Figure 6C). Almost 50% of all significant PLAC-seq consensus interactions are either within or between these features, with an additional 28% of interactions connecting correlated ATAC-seq peaks with other regions in the genome. One example of a highly connected region of CRDs is exemplified in Figure 6D. Only about 20% of all significant PLAC-seq interactions are not connected to ATAC-seq CRDs, even though these features only cover about 11% of the whole genome (Figure S6H). Almost 40% of the correlated ATAC-seq peaks have 11 or more significant interactions (Figure S6I). Comparing the number of PLAC-seq interactions connected to ATAC-seq CRDs to a size-match background showed a highly significant enrichment of interactions associated with CRDs ( $p$ -value  $< 2.2e^{-16}$ ) (Figure S6J).

## Discussion

These studies report systematic analyses of the effects of natural genetic variation on transcription factor binding, epigenetic state and gene expression in resting and activated macrophages. We observe striking levels of variation across BMDMs isolated from different strains at each level of analysis, with diversity of transcription factor binding greatly exceeding that of active histone modifications, nascent RNA production and mature transcript levels. In view of the diverse roles of macrophages in immunity, tissue homeostasis and diseases including atherosclerosis, diabetes, cancer and neurodegeneration (Hirsch et al., 2012; Malm et al., 2015; Moore and Tabas, 2011; Nguyen et al., 2006; Williams et al., 2016) the data sets provided by these studies will be substantial resources for advancing understanding of transcriptional mechanisms and effects of genetic variation. Differences in mRNA gene expression across the five strains are substantial and segregate into distinct biological processes, suggesting substantial differences with respect to immune and tissue homeostatic functions. Consistent with this, we validated the prediction that SPRET mice would exhibit altered Type I interferon responses following TLR4 stimulation. It will be of interest to determine whether the degree of variation in gene expression observed here in macrophages is a common feature of

other cell types. Host responses to pathogens are powerful drivers of evolution of the immune system, and the unique histories of pathogen exposure in the five mouse strains used for analysis may have resulted in more substantial differences in immune cells than other cell types. As the general approach described here can be applied to any cell type, it will be of interest to apply these methods to parenchymal and other cell types of various organs.

The measurement of nascent RNA by GRO-seq and sites of transcription initiation by 5'GRO-seq enabled estimates of the locations of variation underlying differences in genic transcription. By using 5'GRO-seq to annotate transcription start sites, we find that mutations affecting the core promoter element and/or adjacent upstream elements account for less than 20% of divergent nascent gene expression among the most closely related strains. Mutations in core promoter elements are expected to be associated with large effect sizes and therefore more likely to be subject to negative selection. Conversely, analysis of parental alleles in F1 crosses of C57 to PWK or SPRET mice indicates that more than 70% of divergent gene transcription is due to *cis* variation. Thus, differential regulation of gene expression observed in these studies is mostly due to effects on distal regulatory elements, consistent with recent studies across human tissues (Consortium et al., 2017). The availability of data sets for mature and nascent RNA levels across five diverse strains of mice will be of value for future exploration of how genetic variation influences diverse aspects of RNA processing, such as splicing and polyadenylation.

A primary motivation of the proposed studies was to systematically explore functional roles of the hundreds of transcription factors that are expressed in macrophages in establishing the *cis*-regulatory landscape of BMDMs by determining the consequences of mutations in their respective DNA recognition motifs on the binding of a subset of macrophage lineage determining factors. We observed that while mutations in the motifs for PU.1, C/EBP, CJUN and P65 are significantly associated with strain-specific binding, the majority of these differences are explained by mutations in nearby collaborative factors. As previously demonstrated, motif mutations in PU.1 affect nearby binding of C/EBP $\beta$  and vice versa. The present analyses of motif mutations now extend these relationships to a transcription factor interaction network

dominated by collaborative interactions between PU.1, C/EBPs, AP-1, RUNX and USF. Altogether, these studies provide evidence for significant roles of >80 motifs, which can be assigned to 106 of the 248 transcription factors confidently expressed in macrophages. These findings therefore support the hypothesis that a large fraction of the transcription factors expressed in BMDMs shape the DNA binding patterns and functions of macrophage lineage determining factors. In general, motif importance is correlated with expression levels of the factor or factors that recognize it, although there are notable exceptions. For example, E2F family members are among the most highly expressed transcription factors in BMDMs, but mutations in their recognition motifs are not associated with strain-specific binding of PU.1, C/EBP $\beta$  or CJUN.

Beyond the understanding of mechanisms that account for the genomic binding patterns of sequence specific transcription factors, the ability to predict whether such binding results in an increase, decrease or no change in the activity state of a *cis*-regulatory element and a corresponding change in nearby gene expression remains an elusive goal. The data sets generated by these studies provide quantitative estimates of strain-specific enhancer activities based on H3K27ac and nascent RNA levels. It will therefore be of interest to explore the relationships between sequence variants and enhancer activity states to better understand mechanisms underlying functional consequences of transcription factor binding.

Notably, substantial fractions of strain-specific DNA binding and ATAC-seq sites are not associated with or cannot otherwise be explained by local DNA variants. This discrepancy led to the finding of regions in the genome in which transcription factor binding, open chromatin, histone modifications and gene expression are highly correlated. Further investigation of these regions using Hi-C and PLAC-seq analyses indicated that they primarily reside within TADs and are highly intra- and inter-connected. These studies thereby confirm and extend prior studies in human lymphoblastoid cell lines demonstrating both local and distal control of chromatin state (Grubert et al., 2015; Waszak et al., 2015) In particular, several independent lines of evidence suggest that regional interactions between connected *cis*-regulatory elements influence

transcription factor binding independently of local DNA variants. An important future goal will be to determine underlying mechanisms. Given the observation that many connected regulatory domains are associated with coding or non-coding transcripts, one possible model is that initiation of transcription from a 'bootstrap' enhancer or promoter enables RNA polymerase II to function as a chromatin remodeling factor that overcomes local barriers to transcription factor binding. In this model, genetic variants that disable the initiating transcriptional start site would compromise transcription factor binding to the entire CRD. Importantly, BMDMs from different strains of mice provide a highly tractable model system determining molecular mechanisms underlying regional control of transcription factor binding and function.

In concert, the present studies reveal complex relationships between genetic variation, transcription factor binding, epigenetic state and gene expression, only some of which can be currently explained. Elucidation of the underlying mechanisms will be necessary for a better understanding of how non-coding genetic variation influences cellular phenotype. Although macrophages from only five strains of mice were evaluated, they collectively provided more than 50 million SNPs and InDels for analysis. While many of the major conclusions derived from these studies can be achieved based on the 5-6 million variants provided by comparisons of C57, BALB and NOD strains, the inclusion of PWK and SPRET BMDMs greatly increased statistical power to detect both local and regional transcription factor interactions. This degree of genetic variation and the number of complementary genomic assays are not typically achievable in studies of primary human cells. The model system and associated data sets provided by these studies thus represent a powerful new resource for investigation of the influence of genetic variation on gene expression and cellular phenotype. While the positions of *cis*-regulatory elements controlling gene expression in mice and humans are poorly conserved, the mechanisms driving cell specific gene expression are very similar (Cheng et al., 2014; Stergachis et al., 2014). For example, a recent direct comparison of the transcription factor networks driving mouse and human microglia-specific gene expression indicated that they were nearly identical (Gosselin et al., 2017). It is therefore likely that general principles derived from studies

of the influence of genetic variation on gene expression in mice will substantially advance efforts to understand the relationship of non-coding genetic variation and phenotype in humans.

### **Acknowledgements**

We thank L. Van Ael for assistance with manuscript preparation and J. Collier for technical assistance. These studies were supported by NIH grants DK091183, CA17390 and GM085764 and a Leducq Transatlantic Network grant to C. K. G. DNA sequencing was supported by NIH grant DK063491. S.H.D. is a CRI-Irvington Postdoctoral Fellow. D.M. was supported by the German Research Council (DFG) within the Collaborative Research Center (SFB) 1243 (project A16). D.S. was supported by an UCSD institutional Predoctoral training grant (T32DK007541). T.S. was supported by the Swedish Society for Medical Research. H.C.B. was supported by the Sarnoff Cardiovascular Research Foundation Fellowship. T. D. T. was supported by NIH under Award Number T32DK007044 and T32CA009523. I. R. H. was supported the Dutch MS Research Foundation, and the Gemmy and Mibeth Tichelaar Foundation. C. E. R. was supported by NIH-NHLBI grant R00123485.

### **Author Contributions**

V.M.L., D.M, B.R., and C.K.G conceived the study. V.M.L., S.H.D, H.B.C., E.W., M.H., Y.A., T.D.T., M.S., G.J.F, T.S. performed experiments. M.Y. and R.H. performed PLAC-seq experiments. R.F. shared PLAC-seq data processing pipeline. V.M.L., I.R.H., D.S., J.T., and C.E.R analyzed sequencing data. V.M.L. and C.K.G. wrote the manuscript with contributions from D.M. and B.R.

### **Declaration of Interests**

The authors declare no conflict of interests.

### **References**

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory.

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

Barton, K. (2017). MuMIn: Multi-Model Inference.

Bogue, M.A., Grubb, S.C., Walton, D.O., Philip, V.M., Kolishovski, G., Stearns, T., Dunn, M.H., Skelly, D.A., Kadakkuzha, B., TeHennepe, G., *et al.* (2017). Mouse Phenome Database: an integrative database and analysis suite for curated empirical phenotype data from laboratory mice. *Nucleic acids research*.

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. *Nature methods* 10, 1213-1218.

Cheng, Y., Ma, Z., Kim, B.H., Wu, W., Cayting, P., Boyle, A.P., Sundaram, V., Xing, X., Dogan, N., Li, J., *et al.* (2014). Principles of regulatory information conservation between mouse and human. *Nature* 515, 371-375.

Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., *et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422-1423.

Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, *et al.* (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204-213.

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

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* 107, 21931-21936.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376-380.

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

Douglas, B.M., Mächler; Ben, Bolker; Steve, Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* 3, 99-101.

Duttke, S.H., Lacadie, S.A., Ibrahim, M.M., Glass, C.K., Corcoran, D.L., Benner, C., Heinz, S., Kadonaga, J.T., and Ohler, U. (2015). Human promoters are intrinsically directional. *Molecular cell* 57, 674-684.

Eichenfield, D.Z., Troutman, T.D., Link, V.M., Lam, M.T., Cho, H., Gosselin, D., Spann, N.J., Lesch, H.P., Tao, J., Muto, J., *et al.* (2016). Tissue damage drives co-localization of NF-kappaB, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages. *Elife* 5.

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A.D., and Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* 26, 1345-1348.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., *et al.* (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337-343.

Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015). A global reference for human genetic variation. *Nature* 526, 68-74.

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

Gosselin, D., Skola, D., Coufal, N.G., Holtman, I.R., Schlachetzki, J.C.M., Sajti, E., Jaeger, B.N., O'Connor, C., Fitzpatrick, C., Pasillas, M.P., *et al.* (2017). An environment-dependent transcriptional network specifies human microglia identity. *Science (New York, NY)* 356.

Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., *et al.* (2015). Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell* 162, 1051-1065.

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

Hajkova, P., el-Maarri, O., Engemann, S., Oswald, J., Olek, A., and Walter, J. (2002). DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol Biol* 200, 143-154.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., *et al.* (2010). Nucleosome dynamics define transcriptional enhancers. *Nature genetics* 42, 343-347.

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.

Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orozco, L.D., and Glass, C.K. (2013). Effect of natural genetic variation on enhancer selection and function. *Nature*.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nature reviews Molecular cell biology*.

Hetzl, J., Duttke, S.H., Benner, C., and Chory, J. (2016). Nascent RNA sequencing reveals distinct features in plant transcription. *Proceedings of the National Academy of Sciences of the United States of America* 113, 12316-12321.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 106, 9362-9367.

Hirsch, E.C., Vyas, S., and Hunot, S. (2012). Neuroinflammation in Parkinson's disease. *Parkinsonism Relat Disord* 18 Suppl 1, S210-212.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. *Genes & development* 28, 2679-2692.

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. *Proceedings of the National Academy of Sciences of the United States of America* 108, 5290-5295.

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. *Molecular cell* 51, 310-325.

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E., *et al.* (2010). Variation in transcription factor binding among humans. *Science (New York, NY)* 328, 232-235.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., *et al.* (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289-294.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome research* 12, 996-1006.

Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N.I., *et al.* (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science (New York, NY)* 342, 744-747.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571-1572.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., *et al.* (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* 498, 511-515.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24, 719-720.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* 9, 357-359.

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

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Current biology : CB* 20, R754-763.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

Li, Q., Brown, J.B., Huang, H., and Bickel, J. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* 5, 1752-1779.

Link, V.M., Romanoski, C.E., Metzler, D., and Glass, C.K. (2018). MARGE: Mutation analysis of Regulatory Genomic Elements. *bioRxiv*.

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

Lusis, A.J., Seldin, M.M., Allayee, H., Bennett, B.J., Civelek, M., Davis, R.C., Eskin, E., Farber, C.R., Hui, S., Mehrabian, M., *et al.* (2016). The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J Lipid Res* 57, 925-942.

Mahony, S., and Benos, P.V. (2007). STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic acids research* 35, W253-258.

Malm, T.M., Jay, T.R., and Landreth, G.E. (2015). The evolving biology of microglia in Alzheimer's disease. *Neurotherapeutics* 12, 81-93.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., *et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, NY)* 337, 1190-1195.

McDaniell, R., Lee, B.K., Song, L., Liu, Z., Boyle, A.P., Erdos, M.R., Scott, L.J., Morken, M.A., Kucera, K.S., Battenhouse, A., *et al.* (2010). Heritable individual-specific and allele-specific chromatin signatures in humans. *Science (New York, NY)* 328, 235-239.

McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science (New York, NY)* 342, 747-749.

Moore, K.J., and Tabas, I. (2011). Macrophages in the pathogenesis of atherosclerosis. *Cell* 145, 341-355.

Neuwirt, E. (2014). RColorBrewer: ColorBrewer Palettes.

Nguyen, D., Ping, F., Mu, W., Hill, P., Atkins, R.C., and Chadban, S.J. (2006). Macrophage accumulation in human progressive diabetic nephropathy. *Nephrology (Carlton)* 11, 226-231.

Noguchi, S., Arakawa, T., Fukuda, S., Furuno, M., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Kaida, K., Kaiho, A., Kanamori-Katayama, M., *et al.* (2017). FANTOM5 CAGE profiles of human and mouse samples. *Sci Data* 4, 170112.

Oishi, Y., Spann, N.J., Link, V.M., Muse, E.D., Strid, T., Edillor, C., Kolar, M.J., Matsuzaka, T., Hayakawa, S., Tao, J., *et al.* (2017). SREBP1 Contributes to Resolution of Pro-inflammatory TLR4 Signaling by Reprogramming Fatty Acid Metabolism. *Cell Metab* 25, 412-427.

R Development Core Team (2016). R: A Language and Environment for Statistical Computing. In R Foundation for Statistical Computing (Vienna, Austria: <https://www.R-project.org/>).

Raetz, C.R., Garrett, T.A., Reynolds, C.M., Shaw, W.A., Moore, J.D., Smith, D.C., Jr., Ribeiro, A.A., Murphy, R.C., Ulevitch, R.J., Fearn, C., *et al.* (2006). Kdo2-Lipid A of *Escherichia coli*, a defined endotoxin that activates macrophages via TLR-4. *J Lipid Res* 47, 1097-1111.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., *et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665-1680.

Reddy, T.E., Gertz, J., Pauli, F., Kucera, K.S., Varley, K.E., Newberry, K.M., Marinov, G.K., Mortazavi, A., Williams, B.A., Song, L., *et al.* (2012). Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome research* 22, 860-869.

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al.* (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317-330.

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.

Serra, F., Bau, D., Goodstadt, M., Castillo, D., Fillion, G.J., and Marti-Renom, M.A. (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput Biol* *13*, e1005665.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* *13*, 2498-2504.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nature reviews Genetics* *15*, 272-286.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. *J Mol Biol* *147*, 195-197.

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*.

Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelting-Sun, S., Lee, K., *et al.* (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* *515*, 365-370.

Tripathi, S., Pohl, M.O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D.A., Moulton, H.M., DeJesus, P., Che, J., Mulder, L.C., *et al.* (2015). Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* *18*, 723-735.

Urich, M.A., Nery, J.R., Lister, R., Schmitz, R.J., and Ecker, J.R. (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat Protoc* *10*, 475-483.

Waszak, S.M., Delaneau, O., Gschwind, A.R., Kilpinen, H., Raghav, S.K., Witwicki, R.M., Orioli, A., Wiederkehr, M., Panousis, N.I., Yurovsky, A., *et al.* (2015). Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell* *162*, 1039-1050.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* *153*, 307-319.

Williams, C.B., Yeh, E.S., and Soloff, A.C. (2016). Tumor-associated macrophages: unwitting accomplices in breast cancer malignancy. *NPJ Breast Cancer* *2*.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* *9*, R137.

Zhou, X., and Wang, T. (2012). Using the Wash U Epigenome Browser to examine genome-wide sequencing data. *Curr Protoc Bioinformatics Chapter 10*, Unit10 10.

## Figure legends

**Figure 1. Effect of genetic variation on mRNA expression.** **A.** Overview of experimental design and main data sets. (notx = no treatment, KLA = KLA treatment for 1 hour) **B.** Comparison of RNA-seq for polyadenylated (polyA) transcripts in BMDMs derived from the indicated mouse strains under no treatment (notx) conditions. Log<sub>2</sub>(TPM+1) values are plotted for BALB, NOD, PWK and SPRET vs. C57 (TPM = transcripts per kilobase million). The number of SNPs and InDels for each comparison is shown at the top. Transcripts exhibiting >2-fold or >4-fold changes at an FDR < 0.01 are color coded light blue and dark blue respectively. **C.** Clustering of differentially expressed genes determined by WGCNA. Differences are indicated by z-score. The top functional annotations for each cluster are determined by Metascape (Tripathi et al., 2015) and are illustrated on the right with the log q-value. (See Figure S1 for modules). **D.** Ratio-ratio plots of the fold response to KLA in BALB vs. C57 BMDMs (top) and in SPRET vs. C57 BMDMs (bottom). Each point represents one polyA RNA transcript. Dark blue dots show genes that are 4-fold reciprocal regulated, whereas green dots show a 4-fold stronger response to the KLA stimulus in one strain over the other. **E.** Relationship of differentially expressed genes to number of genetic variation in million. **F.** Comparison of expression of 46 primary interferon stimulated genes (ISGs) in C57 and SPRET BMDMs under no treatment conditions and following KLA stimulation for 6h. The right column represents the ratio of SPRET/C57 gene expression following KLA treatment.

**Figure 2. Effect of genetic variation on nascent transcription.** **A.** Comparison of GRO-seq gene body tag counts in BMDMs derived from the indicated mouse strains under no treatment conditions. Log<sub>2</sub>(tag counts+1) values are plotted for BALB, NOD, PWK and SPRET versus. C57. Nascent transcriptions exhibiting >2-fold or >4-fold changes at an FDR < 0.01 are color coded light blue and dark blue respectively. **B.** Comparison of GRO-seq, 5'GRO-seq and H3K27ac signal at the *Igf1* locus in BMDMs derived from each strain under no treatment conditions. **C.** Ratio-ratio plots of GRO-seq tag counts for KLA/no treatment conditions, comparing BALB vs. C57 in the left panel and SPRET vs. C57 in the right panel. Dark blue dots show genes that are 4-fold reciprocal regulated, whereas green dots show a 4-fold stronger response to the KLA stimulus

in one strain over the other. **D.** Relationship of differential RNA-seq expression as a function of mutations between -30 and +20 bp of the TSS defined by 5'GRO-seq signal. **E.** Ratio ratio plot of gene body GRO-seq tag counts in BMDMs derived from C57 and SPRET mice versus allele-specific tag counts in BMDMs derived from SPRET x C57 F1 mice. **F.** GRO-seq expression for *Npy* and *Pla2g7* in BMDMs derived from C57 and SPRET mice and allele-specific tag counts in BMDMs derived from SPRET x C57 F1 mice.

**Figure 3. Effect of genetic variation on cis-regulatory landscapes.** **A.** Scatter plots of log<sub>2</sub> tag counts for H3K27ac ChIP-seq regions comparing BALB and C57 (left) and SPRET and C57 (right). Regions exhibiting >2-fold or >4-fold different binding are colored light blue and dark blue respectively. **B.** Scatter plots of log<sub>2</sub> tag counts for ATAC-seq peaks passing IDR comparing BALB and C57 (left) and SPRET and C57 (right). Peaks exhibiting >2-fold or >4-fold changes are colored light blue and dark blue respectively. **C.** De novo motif analysis of distal (>3000 bp from TSS) ATAC-seq peaks associated with H3K27ac signal in each strain. Boxes display negative log<sub>10</sub> *p*-values for enrichment of the corresponding motif and its rank order in parentheses. **D.** Pie chart indicating fractions of distal H3K27ac-positive regions of open chromatin occupied by PU.1, C/EBPβ and/or CJUN. **E.** Heat map of H3K27ac tag density at genomic regions defined as super enhancers. **F.** Comparison of log<sub>2</sub> ChIP-seq tag counts for PU.1 in BMDMs derived from the indicated mouse strains under no treatment conditions. Features exhibiting >2-fold or >4-fold changes are colored light blue and dark blue respectively. **G.** Ratio ratio plot of PU.1 ChIP-seq tag counts in BMDMs derived from C57 and SPRET mice versus allele-specific tag counts in BMDMs derived from C57 x SPRET F1 mice. **H.** SNPs and InDels frequencies in ATAC-seq and PU.1 peaks within 300bp of the peak center for the indicated strain comparisons for stringent VCF filter criteria and more relaxed criteria.

**Figure 4. Transcription factor interaction networks inferred from effects of motif mutations.** **A.** Heat map of a subset of significant motifs after application of MARGE under control and KLA treatment conditions. For a complete listing, see Table S4. **B.** Top 14 of 60 motifs correlated with binding of PU.1 under no treatment conditions as determined by motif mutation analysis.

Node size is fraction of PU.1 peaks containing the indicated motif and edge thickness is proportional to the effect size of motif mutations. Nodes in red indicate motifs in which mutations result in reduced PU.1 binding. Nodes in blue indicate motifs in which mutations result in increased PU.1 binding. **C.** Top motifs correlated with binding of P65 under KLA treatment conditions as determined by motif mutation analysis. Node size and edge thickness are as defined in Panel B. **D.** Integrated network of collaborative transcription factors. The top 15 of 80 motifs for which motif mutations affected binding of at least one of the three factors are shown. Node sizes are average of all three analyses and edge thickness is as defined in Panel B. **E.** Fraction of strain specific binding of PU.1, C/EBP $\beta$ , CJUN and P65 explained by mutations in their respective recognition motifs and by all mutations considered by MARGE analysis. Values are for peaks exhibiting >4-fold differences in at least one comparison in genomic regions containing local variants. **F.** Overlap of binding of PU.1, RUNX1 and USF2 under no treatment conditions as determined by CHIP-seq for each factor. **G.** Fraction of open chromatin marked by H3K27ac and not bound by PU.1, CJUN or C/EBP occupied by RUNX1, USF2 and/or NRF2. **H.** Relationship of mutations in RUNX motifs on binding of RUNX1 and PU.1 in C57 and SPRET BMDMs.

**Figure 5. Regional correlation of ATAC-seq and ChIP-seq peaks. A-C.** Heat maps of Pearson correlation coefficients (PCC) of PU.1, C/EBP $\beta$  and CJUN peaks, respectively, across the five strains under control conditions in a 5 mega base window from chromosome 18. Vertical and horizontal lines represent TAD boundaries as defined by C57 Hi-C assays presented in Figure 6. **D.** Illustration of regional correlation of GRO-seq, ATAC-seq, PU.1, C/EBP $\beta$ , CJUN, and H3K27ac signal in the vicinity of the *Collec12* gene. **E.** Percentages of PU.1 connected regulatory domains (CRDs) based on minimum peak number and minimum Pearson correlation coefficient (PCC). **F.** Relationship of strain-specific PU.1 CRDs to enhancer activity measured by 5'GRO-seq and expression of nearest gene measured by RNA-seq. Significance was calculated using a two-sided t-test. **G.** Heat maps for relative binding and 5'GRO-seq signal at PU.1 CRDs. The ordering of PU.1 signal and corresponding 5'GRO-seq signal is the same for the two plots.

**Figure 6. Connected regulatory domains. A.** Left: Hi-C contact frequency maps for chromosome 18 in BMDMs derived from C57 (lower left) and SPRET (upper right) mice. The values for the PC1 eigenvector are shown at the bottom. Right: Zoomed-in view of Hi-C contact frequency for chromosome 18 visualizing TAD boundaries in SPRET and C57. **B.** RNA-seq, H3K27ac, PC1, and Hi-C contact loops in the vicinity of the *Spi1* locus (left) and the *Colec12* locus (right). **C** Fraction of significant consensus PLAC-seq interactions within ATAC-seq CRDs, between ATAC-seq CRDs and outside of CRDs. **D.** Example of ATAC-seq notx CRDs highly connected by PLAC-seq consensus interactions.

## Supplemental Figures

**Figure S1. A.** (Left) Bones were extracted from 8 – 12 week old female mice and 20 million cells were plated per 15 cm dish in Media with M-CSF for 7 days. (Right) Average number of cells extracted per mouse (C57: 14 harvest (n=53), BALB: 11 harvest (n=54), NOD: 10 harvest (n=47), PWK: 12 harvest (n=99), SPRET: 9 harvest (n=66)). **B.** Comparison of polyA RNA-seq replicates for C57 in no treatment (left) and after KLA treatment for 1h (right) (TPM = transcripts per kilobase million). **C.** Clustered spearman correlation matrix for different RNA-seq replicates for no treatment and KLA 1h. **D.** Gene dendrogram and module colors from WGCNA analysis.

**Figure S2. A.** Clustered spearman correlation matrix for GRO-seq replicates using tag counts along gene bodies for no treatment and KLA 1h. **B.** Distance of RefSeq annotated transcription start sites (TSS) and TSS defined by 5'GRO-seq data for C57. **C.** Bar plot showing the percentage of overlap between RefSeq annotated TSS and TSS defined by 5'GRO-seq signal for no treatment and KLA 1h. **D.** Relationship of differential RNA-seq expression as a function of mutations between -30 and +20 bp of the TSS defined by 5'GRO-seq signal. **E.** Ratio ratio plot of gene body GRO-seq tag counts in BMDMs derived from C57 and PWK mice versus allele-specific tag counts in BMDMs derived from PWK x C57 F1 mice. **F.** Boxplot showing average percentage of *cis*-regulated genes between parental and F1 alleles (FPC = PWK x C57 F1, FSC = SPRET x C57 F1).

**Figure S3. A.** Comparison of replicates for PU.1 notx ChIP-seq for C57 (left) and SPRET (right). Grey dots are peaks called in either replicate, green dots are peaks passing IDR. **B.** Clustered spearman correlation matrix for PU.1 ChIP-seq replicates (top left), C/EBP $\beta$  ChIP-seq replicates (top right), CJUN ChIP-seq replicates (bottom left), and P65 ChIP-seq replicates (bottom right) for all strains in no treatment and KLA 1h conditions (KLA 1h condition only for P65). **C-E.** Scatter plots of log<sub>2</sub> tag counts for H3K4me<sub>2</sub> ChIP-seq regions (**C**), C/EBP $\beta$  ChIP-seq peaks (**D**), CJUN ChIP-seq peaks (**E**) and P65 ChIP-seq peaks (**F**) comparing BALB and C57 (left) and SPRET and C57 (right). Regions exhibiting >2-fold or >4-fold different binding are colored light blue and dark blue respectively. **G.** Ratio ratio plot of normalized PU.1 ChIP-seq tag counts in BMDMs

derived from C57 and PWK mice versus allele-specific tag counts in BMDMs derived from PWK x C57 F1 mice. **H.** Boxplot showing average percentage of *cis*-regulated transcription factor binding sites for all loci with fold change going in the same direction, for all loci with 2-fold change (4-fold change, respectively) in parental and F1 alleles. **I.** SNP+InDel frequencies in CEBP $\beta$ , CJUN and P65 peaks for the indicated strain comparisons. **J.** Average minimum and maximum tag counts per peak for ATAC-seq (left) and PU.1 (right) with >4-fold difference in binding between C57 versus SPRET without mutations that overlap or do not overlap structural variances. **K.** Percentage of differently bound ATAC-seq and PU.1 peaks without mutations overlapping structural variances (left panel) and peaks with mutations overlapping structural variances (right panel).

**Figure S4. A.** Heat map of all significant motifs ( $p$ -value <  $1e^{-10}$ ) after application of MARGE under control and KLA treatment conditions. Dendrogram on right shows clustering of consensus motifs by similarity based on Pearson correlation coefficient (PCC). On the left consensus motif logos are shown. **B.** Top motifs correlated with binding of C/EBP $\beta$  under no treatment conditions as determined by MARGE's motif mutation analysis. Node size is fraction of C/EBP $\beta$  peaks containing the indicated motif and edge thickness is proportional to the effect size of motif mutations. Nodes in red indicate motifs in which mutations result in reduced C/EBP $\beta$  binding. Nodes in blue indicate motifs in which mutations result in increased C/EBP $\beta$  binding. **C.** Top motifs correlated with binding of CJUN under no treatment conditions as determined by MARGE's motif mutation analysis. Node size and edge thickness are as defined in Panel B. **D.** Fraction of strain specific binding of PU.1, C/EBP $\beta$ , CJUN and P65 explained by mutations in their respective recognition motifs and by all motifs found significant by MARGE analysis. Values are for the 500 most differently bound peaks containing local variants. **E.** Relationship of mutations in NRF2 motifs on binding of NRF2 and PU.1 in C57 and SPRET BMDMs with intact PU.1, C/EBP, and AP-1 binding motifs. **F.** Relationship of mutations in USF motifs on binding of USF2 and PU.1 in C57 and SPRET BMDMs with intact PU.1, C/EBP and AP-1 binding motifs.

**Figure S5. A.** Bar plot showing the average number of peaks within CRDs for different ChIP-seq data sets and treatments **B.** Bar plot showing the average length of CRDs in kilo bases (kb) for different ChIP-seq data sets and treatments. **C.** Ratio ratio plot of normalized PU.1 ChIP-seq tag counts summed for all peaks within CRDs in BMDM derived from C57 and SPRET mice versus allele-specific tag counts in BMDMs derived from SPRET x C57 F1 mice. **D.** Overlap of CRDs and super enhancers in percentage. CRDs and super enhancers were merged for different assays and percentage of CRDs only, super enhancers only and overlap of both was calculated. **E.** Relationship of strain-specific ATAC-seq notx CRDs to enhancer activity measured by 5'GRO-seq and expression of nearest gene measured by RNA-seq. Significance was calculated using a two-sided t-test.

**Figure S6. A.** Percentage of CpGs methylated (mCpG) at promoters (defined by RefSeq) of all genes for C57 (green) and SPRET (purple). **B.** Percentage of CpGs methylated at promoters of genes 2-fold differently expressed in C57 vs. SPRET. **C.** Percentage of CpGs methylated at promoters of genes 4-fold differently expressed in C57 vs. SPRET. **D.** UCSC genome browser shot showing promoter methylation for highly differently expressed genes *Colce12* (left), *Npy* (middle) and *Igf1* (right). **E.** Scatter plot of PC1 Eigenvector values comparing C57 and SPRET derived from the respective Hi-C contact maps. **F.** Expression values of genes associated with Eigenvector values of the same sign (black) in C57 and SPRET BMDMs and genes associated with Eigenvector values of opposite sign (red). **G.** General conservation of TAD boundaries and contact frequency within TADs measured by inclusion ratio in BMDMs derived from C57 and SPRET mice. **H.** Fraction of genome covered by ATAC-seq CRDs. **I.** Distribution of significant PLAC-seq consensus interactions in ATAC-seq CRDs. **J.** Comparison of overlap of significant PLAC-seq interactions in ATAC-seq CRDs in comparison to size-matched background shows significant enrichment of PLAC-seq interactions in CRDs.

## Supplemental Tables

**Table S1. Summary of mRNA expression** Gene expression measured by RNA-seq normalized to log<sub>2</sub> transcripts per million (TPM) in replicates for all expressed genes (TPM > 1) for BALB, C57, NOD, PWK, and SPRET BMDMs under control conditions (notx) and after KLA 1h treatment.

**Table S2. Summary of nascent transcription.** GRO-seq tag counts along gene bodies normalized to 10 million sequenced reads in replicates for BALB, C57, NOD, PWK, and SPRET BMDMs under control conditions (notx) and after KLA 1h treatment.

**Table S3. Summary of sequencing assays.** Number of features (peaks or regions) identified for each of assay (ATAC-seq, PU.1 ChIP-seq, CEBPb ChIP-seq, CJUN ChIP-seq, P65 ChIP-seq, H3K27ac ChIP-seq and H3K4me2 ChIP-seq) in the five strains of BMDMs under control (notx) and KLA-treated conditions for 1 hour after IDR, as well as the union of features per experiment.

**Table S4. MARGE analysis results.** List of all transcription factor binding motifs found significant in MARGE analysis for at least one factor and treatment condition.

**Table S5. Candidate transcription factors for MARGE analysis results.** List of all expressed transcription factors (TPM > 1) which can bind motifs found as significant by MARGE analysis including the transcription factor, its transcription factor family, as well as RNA-seq expression of this factor in all five strains and treatment (in TPM).

**Table S6. Summary of CRDs.** List of all CRDs for all assays and treatments, as well as number of CRDs that are strain-specific and strain-similar.

## **STAR Methods**

### **Key Resources Table**

*see separate file (use template)*

### **Contact for Reagents and Resource Sharing**

Further information and requests for resources and reagents should be directed and will be fulfilled by the Lead Contact, Christopher K. Glass (ckg@ucsd.edu)

### **Experimental Model and Subject Details**

#### **Mice**

Female and male breeder mice for C57BL/6J, BALB/cJ, NOD/ShiLtJ, PWK/PhJ, and SPRET/EiJ mice were purchased from Jackson Laboratory. Mice were housed at the UCSD animal facility on a 12 h/12 h light/dark cycle with free access to food and water. All animal procedures were in accordance with University of California San Diego research guidelines for the care and use of laboratory animals. 8 – 12 week old female mice were used for experiments. For F1 crosses female C57BL/6J mice were crossed with male PWK/PhJ and SPRET/EiJ and 8 -12 week old female F1 mice were used for experiments.

#### **Bone marrow-derived macrophage (BMDM) culture**

Femur, tibia and iliac bones from the different mouse strains were flushed with DMEM high glucose (Corning) and red blood cells were lysed using red blood cell lysis buffer (eBioscience). After counting, 20 million bone marrow cells were seeded per 15cm non-tissue culture plates in DMEM high glucose (50%) with 20% FBS (Omega Biosciences), 30% L929-cell conditioned media (as source of M-CSF), 100 U/ml penicillin/streptomycin+L-glutamine (Gibco) and 2.5µg/ml Amphotericin B (HyClone). After 4 days of differentiation, 16.7 ng/ml mouse M-CSF (Shenandoah Biotechnology) was added to the media. After an additional 2 days of culture,

non-adherent cells were washed off with room temperature DMEM and macrophages were obtained as a homogeneous population of adherent cells which were scraped and subsequently seeded onto tissue culture-treated petri dishes overnight in DMEM containing 10% FBS, 100 U/ml penicillin/streptomycin+L-glutamine, 2.5µg/ml Amphotericin B and 16.7 ng/ml M-CSF. For KLA activation, macrophages were treated with 10 ng/mL KLA (Avanti Polar Lipids) for 1 or 6 hours.

## **Method Details**

### **RNA-seq library preparation**

Total RNA was isolated from cells and purified using Quick RNA mini prep columns and RNase-free DNase digestion according to the manufacturer's instructions (Zymo Research). RNA-seq libraries were prepared from poly(A)-enriched mRNA, as previously described (Oishi et al., 2017). Libraries were sequenced on HiSeq 4000 or NextSeq 500.

### **Crosslinking for ChIP-seq**

For PU.1, C/EBPβ, H3K4me2, and H3K27ac ChIP-seq, culture media was removed and plates were washed once with PBS and then fixed for 10 minutes with 1% formaldehyde (Fischer Scientific) in PBS at room temperature and reaction was then quenched by adding glycine (Sigma) to 0.125M.

For CJUN, P65, USF2 and RUNX1 ChIP-seq, cells were cross-linked for 30 minutes with 2mM DSG (Pierce) in PBS at room temperature. Subsequently cells were fixed for 10 minutes with 1% formaldehyde at room temperature and the reaction was quenched with 0.125M glycine. After fixation, cells were washed once with cold PBS and then scraped into supernatant using a rubber policeman, pelleted for 8 minutes at 400xG at 4°C. Cells were transferred to DNA lobind tubes and pelleted at 700xG for 5 minutes at 4°C and snap-frozen in liquid nitrogen and stored at -80°C until ready for ChIP-seq protocol preparation.

### **ChIP-seq library preparation**

Chromatin immunoprecipitation (ChIP) was performed as described previously (Oishi et al., 2017). In brief, cells were resuspended in swelling buffer (10mM HEPES/KOH (pH7.9), 85mM KCl, 1mM EDTA, 0.5% IGEPAL CA-630) with protease inhibitors for 5min and then spun down and resuspended in 500µl lysis buffer (50mM Tris/HCl (pH7.4), 1% SDS, 0.5% Empigen BB, 10mM EDTA) with protease inhibitors, and chromatin was sheared using the Bioruptor (Diagenode). Lysate was diluted with 750µl dilution buffer (20mM Tris/HCl, 100mM NaCl, 0.5% TritonX-100, 2mM EDTA), 1% was taken as input DNA, and immunoprecipitation was carried out overnight with Dynabeads protein G bound to specific antibodies for PU.1 (Santa Cruz, sc-352X), C/EBPβ (Santa Cruz, sc-150), H3K4me2 (Millipore, 07-030), H3K27ac (Active Motif, 39135), CJUN (Santa Cruz, sc-1694), P65 (Santa Cruz, sc-372X), USF2 (Santa Cruz, sc-862X) and RUNX1 (Santa Cruz, sc-365644). Beads were washed twice each with wash buffer I (20mM Tris/HCl, 150mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA), wash buffer II (10mM Tris/HCl, 250mM LiCl, 1% IGEPAL CA-630, 0.7% Na-deoxycholate, 1mM EDTA), TE 0.2% Triton X-100 and TE 50mM NaCl and subsequently eluted with elution buffer (TE, 2% SDS). DNA was reverse-crosslinked and purified using CHIP DNA Clean & Concentrator (Zymo Research) according to the manufacturer's instructions. Sequencing libraries were prepared from eluted DNA by blunting, A-tailing, adaptor ligation as previously described (Heinz et al., 2010) using NextFlex barcodes (Bioo Scientific). Libraries were PCR-amplified for 12-15 cycles, size selected using PAGE/TBE gels for 200-400bp fragments by gel extraction and single-end sequenced HiSeq 4000 or NextSeq 500.

### **5'GRO-seq and GRO-seq**

Nascent transcription was captured by global nuclear run-on sequencing (GRO-seq) (Core et al., 2008) and nascent transcription start sites by 5'GRO-seq (Lam et al., 2013). Nuclei were isolated from BMDMs using hypotonic lysis [10 mM Tris-HCl (pH 7.5), 2 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>; 0.1% IGEPAL CA-630] and flash frozen in GRO-freezing buffer [50 mM Tris-HCl (pH 7.8), 5 mM MgCl<sub>2</sub>, 40% Glycerol]. 3-5 x 10<sup>6</sup> BMDM nuclei were run-on with BrUTP-labelled NTPs as described (Duttke et al., 2015) with 3x NRO buffer [15mM Tris-Cl (pH 8.0), 7.5 mM MgCl<sub>2</sub>, 1.5 mM DTT, 450 mM KCl, 0.3 U/µL of SUPERase In, 1.5% Sarkosyl, 366 µM ATP, GTP (Roche), Br-UTP (Sigma

Aldrich) and 1.2  $\mu\text{M}$  CTP (Roche, to limit run-on length to  $\sim 40$  nt)]. Reactions were stopped after five minutes by addition of 500  $\mu\text{L}$  Trizol LS reagent (Invitrogen), vortexed for 5 minutes and RNA extracted and precipitated as described by the manufacturer.

### **GRO-seq library preparation**

For GRO-seq, RNA pellets were resuspended in 18  $\mu\text{L}$  ddH<sub>2</sub>O + 0.05% Tween (dH<sub>2</sub>O+T) and after addition of 2  $\mu\text{L}$  fragmentation mix [100 mM ZnCl<sub>2</sub>, 10 mM Tris-HCl (pH 7.5)], incubated at 70°C for 15 minutes. Fragmentation was stopped by addition of 2.5  $\mu\text{L}$  100 mM EDTA. BrdU enrichment was performed using BrdU Antibody (IIB5) AC beads (Santa Cruz, sc-32323 AC, lot #A0215 and #C1716), as described in detail by Hetzel et al (Hetzel et al., 2016). Beads were washed once with GRO binding buffer [0.25 $\times$ saline-sodium-phosphate-EDTA buffer (SSPE), 0.05% (vol/vol) Tween, 37.5 mM NaCl, 1 mM EDTA] + 300 mM NaCl followed by three washes in GRO binding buffer and resuspend as 25% (vol/vol) slurry with 0.1 U/ $\mu\text{L}$  SUPERase-in. To fragmented RNA, 500  $\mu\text{L}$  cold GRO binding buffer and 40  $\mu\text{L}$  equilibrated BrdU antibody beads were added and samples slowly rotated at 4°C for 80 minutes. Beads were subsequently spun down at 1000xG for 15 seconds, supernatant removed and the beads transferred to a Millipore Ultrafree MC column (UFC30HVN; Millipore) in about 2 $\times$  200  $\mu\text{L}$  GRO binding buffer. The IP reaction was washed twice with 400  $\mu\text{L}$  GRO binding buffer before RNA was eluted by incubation in 200  $\mu\text{L}$  Trizol LS (Thermo Fisher) under gentle agitation for 3 minutes. The elution was repeated a second time, 120  $\mu\text{L}$  of dH<sub>2</sub>O+T added to increase the supernatant and extracted as described by the manufacturer.

For end-repair and decapping, RNA pellets were dissolved in 8  $\mu\text{L}$  TET [10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.05 % Tween20] by vigorous vortexing, heated to 70°C for 2 minutes and placed on ice. After a quick spin, 22  $\mu\text{L}$  Repair MM [3  $\mu\text{L}$  10 $\times$  PNK buffer, 15.5  $\mu\text{L}$  dH<sub>2</sub>O+T, 0.5  $\mu\text{L}$  SUPERase-In RNase Inhibitor (10 U), 2  $\mu\text{L}$  PNK (20U), 1  $\mu\text{L}$  RppH (5U)] was added, mixed by flicking and incubated at 37°C for 1 hour. To phosphorylate the 5' end, 0.5  $\mu\text{L}$  100 mM ATP was subsequently added and the reactions were incubated for another 45 minutes at 37°C (the high ATP concentration quenches RppH activity). Following end repair, 2.5  $\mu\text{L}$  50 mM EDTA was added, reactions mixed and then heated to 70°C for 2 minutes before being placed on ice. A

second BrdU enrichment was performed as detailed above. For library preparation, RNA pellets were dissolved in 2.75  $\mu$ l TET + 0.25  $\mu$ l Illumina TruSeq 3' Adapter (10  $\mu$ M), heated to 70°C for 2 minutes and placed on ice. 7  $\mu$ l of 3' MM [4.75  $\mu$ l 50% PEG8000, 1  $\mu$ l 10x T4 RNA ligase buffer, 0.25  $\mu$ l SUPERase-In, 1  $\mu$ l T4 RNA Ligase 2 truncated (200U; NEB)] was added, mixed well by flicking and reactions incubated at 20°C for 1 hour. Reactions were diluted by addition of 10  $\mu$ l TET + 2  $\mu$ l 50 mM EDTA, heated to 70°C for 2 minutes, placed on ice and a third round of BrdUTP enrichment was performed. RNA pellets were transferred to PCR strips during the 75% ethanol wash and dried. Samples were dissolved in 4  $\mu$ l TET [10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 0.05% Tween 20] + 1  $\mu$ l 10  $\mu$ M reverse transcription (RT) primer. To anneal the RT-primer, the mixture was incubated at 75°C for 5 minutes, 37°C for 15 minutes and 25°C for 10 minutes. To ligate the 5' Illumina TruSeq adapter, 10  $\mu$ l 5' MM [1.5  $\mu$ l ddH<sub>2</sub>O + 0.2% Tween20, 0.25  $\mu$ l denatured 5' TruSeq adapter (10  $\mu$ M), 1.5  $\mu$ l 10x T4 RNA ligase buffer, 0.25  $\mu$ l SUPERase-In, 0.2  $\mu$ l 10 mM ATP, 5.8  $\mu$ l 50% PEG8000, 0.5  $\mu$ l T4 RNA ligase 1 (5U; NEB)] was added and reactions were incubated at 25°C for 1 hour. Reverse transcription was performed using Protoscript II (NEB) [4  $\mu$ l 5x NEB FirstStrand buffer (NEB; E7421AA), 0.25  $\mu$ l SUPERase-In, 0.75  $\mu$ l Protoscript II (150U; NEB)] at 50°C for 1 hour. After addition of 30  $\mu$ l PCR MM [25  $\mu$ l 2X LongAmp Taq 2X Master Mix (NEB), 0.2  $\mu$ l 100  $\mu$ M forward primer, 2.8  $\mu$ l 5M Betaine and 2  $\mu$ l 10  $\mu$ M individual barcoding primer], mixtures were amplified (95°C for 3 minutes, [95°C for 60 seconds, 62°C for 30 seconds, 72°C for 15 seconds] x13, 72°C for 3 minutes). PCR reactions were cleaned up using 1.5 volumes of SpeedBeads™ (GE Healthcare) in 2.5M NaCl/20% PEG8000 and libraries size selected on a PAGE/TBE gels to 160–225 base pairs. Gel slices were shredded by spinning through a 0.5 ml perforated PCR tube placed on top of a 1.5 ml tube. 150  $\mu$ l Gel EB [0.1% LDS, 1M LiCl, 10 mM Tris-HCl (pH 7.8)] was added and the slurry incubate under agitation overnight. To purify the eluted DNA, 700  $\mu$ l Zymogen ChIP DNA binding buffer was added into the 1.5 ml tube containing the shredded gel slice and the Gel EB, mixed by pipetting and the slurry transferred to a ZymoMiniElute column. Samples were first spun at 1000xG for 3 minutes, then 10,000xG for 30 seconds. Flow through was removed, and samples washed with 200  $\mu$ l Zymo WashBuffer (with EtOH). Gel remainders were removed by flicking and columns washed by addition of another 200  $\mu$ l Zymo WashBuffer (with EtOH). Flow through was

removed, columns spun dry by centrifugation at 14,000xG for 1 minute and DNA eluted by addition of 20  $\mu$ l pre-warmed Sequencing TET [10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.05% Tween 20]. Libraries were sequenced on an Illumina NextSeq 500.

### **5'GRO-seq library preparation**

RNA pellets were resuspended in 10  $\mu$ l TET, heated to 70°C for 2 minutes and placed on ice. 10  $\mu$ l of dephosphorylation MM [2  $\mu$ l 10x CutSmart, 6.75  $\mu$ l dH<sub>2</sub>O+T, 1  $\mu$ l Calf Intestinal alkaline Phosphatase (10 U; CIP, NEB), 0.25  $\mu$ l SUPERase In (5U)] was added. Following incubation at 37°C for 45 minutes, 2  $\mu$ l 50 mM EDTA was added, reactions mixed, heated to 70°C for 2 minutes and placed on ice. BrdU enrichment was performed as described for GRO-seq. RNA pellets were dissolved in 10  $\mu$ l TET and a second round of dephosphorylation and BrdU enrichment was performed. Libraries were prepared as described in (Hetzl et al., 2016). Briefly, libraries were done as described for GRO-seq with exception of the 3'Adapter ligation step. Here, prior to 3'Adapter ligation, samples were dissolved in 3.75  $\mu$ l TET heated to 70°C for 2 minutes and placed on ice. RNAs were decapped by addition of 6.25  $\mu$ l RppH MM [1  $\mu$ l 10x T4 RNA ligase buffer, 4  $\mu$ l 50% PEG8000, 0.25  $\mu$ l SUPERase-In, 1  $\mu$ l RNA 5' Pyrophosphohydrolase (5U; RppH, NEB)] and incubated at 37°C for 1 hour. Afterwards, to ligate the 3' Illumina TruSeq adapter 10  $\mu$ l of 3'MM was added [1  $\mu$ l 10x T4 RNA ligase buffer, 6  $\mu$ l 50% PEG8000, 1.5  $\mu$ l ddH<sub>2</sub>O+T, 0.25  $\mu$ l heat-denatured Illumina TruSeq 3'Adapter, 0.25  $\mu$ l SUPERase-In, 1  $\mu$ l T4 RNA Ligase 2 truncated K227Q (200U; NEB)], mixed well by flicking and reactions incubated at 20°C for 1 hour. Reactions were diluted by addition of 10  $\mu$ l TET + 2  $\mu$ l 50 mM EDTA, heated to 70°C for 2 minutes, placed on ice and a third round of BrdUTP enrichment was performed. 5' adapter ligation, reverse transcription and library size selection were performed as described for GRO-seq. Samples were amplified for 14 cycles, size selected for 160–250 bp and sequenced on an Illumina NextSeq 500.

### **ATAC-seq library preparation**

To approximately 150k nuclei in 22.5  $\mu$ l GRO freezing buffer, isolated as described for GRO-seq above, 25  $\mu$ l DNA Tagmentation buffer was added, reaction mixed and 2.5  $\mu$ l DNA

Tagmentation Enzyme mix (Nextera DNA Library Preparation Kit, Illumina) added. Mixture was incubated at 37°C for 30 minutes and subsequently purified using the Zymogen ChIP DNA purification kit as described by the manufacturer. DNA was amplified using the Nextera Primer Ad1 and a unique Ad2.n barcoding primers using NEBNext High-Fidelity 2X PCR MM for 10 cycles. PCR reactions were purified using 1.5 volumes of SpeedBeads in 2.5M NaCl, 20% PEG8000, size selected using PAGE/TBE gels for 160 – 280 bp and DNA eluted as described for GRO-seq.

### **PLAC-seq library preparation**

PLAC-seq libraries were prepared as described in (Fang et al., 2016). In brief, cells were cross-linked for 15 minutes at room temperature with 1% formaldehyde and quenched for 5 minutes at room temperature with 0.2M glycine. The cross linked cells were centrifuged at 2500xG for 5 minutes. To isolate nuclei, cross-linked cells were resuspended in 200 µl lysis buffer (10mM Tris-HCl (pH 8.0), 10mM NaCl, 0.2% IPEGAL CA-630) and incubated on ice for 15 minutes. The suspension was then centrifuged at 2500xG for 5 minutes and the pellet washed by resuspending in 300 µl lysis buffer and centrifuging at 2500xG for 5 minutes. The pellet was resuspended in 50 µl 0.5% SDS and incubated for 10 minutes at 62°C. 170 µl 1.47% TritonX-100 was added to the suspension and incubated for 15 minutes at 37°C. 25 µl of 10X NEBuffer 2 and 100U Mbol was added to digest chromatin for 2 hours at 37°C at 1000rpm rotation. Enzymes were inactivated by heating for 20 minutes at 62°C. Fragmented ends were biotin labeled by adding 50 µl of a mix containing 0.3mM biotin-14-dATP, 0.3mM dATP, 0.3mM dTTP, 0.3mM dGTP, and 0.8U/µl Klenow and incubated for 60 minutes at 23°C with rotation (500rpm). Ends were subsequently ligated by adding a 900 µl master mix containing 120 µl 10X T4 DNA ligase buffer (NEB), 100 µl 10% TritonX-100, 12 µl 10mg/mL BSA, 10 µl 400U/µl T4 DNA Ligase (NEB, high concentration formula) and 658 µl H<sub>2</sub>O and incubated for 240 minutes at 23°C with 300rpm slow rotation. Nuclei were pelleted for 5 minutes at room temperature at 2500XG. For the ChIP, nuclei were resuspended in RIPA Buffer (10mM Tris (pH 8.0), 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate) with proteinase inhibitors and incubated on ice for 10 minutes. Sonication was performed using a Covaris M220 instrument

(Power 75W, duty factor 10%, cycle per bust 200, time 10 minutes, temperature 7°C) and nuclei were spun for 20 minutes at 14000rpm at 4°C. For pre-cleaning, protein G sepharose beads were added to the supernatant and rotated for 3 hours at 4°C. 5% of supernatant was taken as input DNA, and to the remaining volume 2.5µg of anti-H3K4me3 antibody (04-745, Millipore) was added and rotated at 4°C overnight for immunoprecipitation. 0.5% BSA-blocked protein G sepharose beads was added and the sample was rotated for 3 hours at 4°C. The sample was centrifuged at 2000rpm for 1 minute and the beads were washed three times with RIPA buffer, two times with high-salt RIPA buffer (10mM Tris pH 8.0, 300mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% deoxycholate), one time with LiCl buffer (10mM Tris (pH 8.0), 250mM LiCl, 1mM EDTA, 0.5% IGEPAL CA-630, 0.1% sodium deoxycholate) and finally two times with TE buffer (10mM Tris (pH 8.0), 0.1mM EDTA). Washed beads were treated with 10µg RNase A in extraction buffer (10mM Tris (pH 8.0), 350mM NaCl, 0.1mM EDTA, 1%SDS) for 1 hour at 37°C, and subsequently 20µg proteinase K was added at 65°C overnight. ChIP DNA was purified by Phenol/Chloroform/Isoamyl Alcohol (25:24:1) extraction and then ethanol purification with final elution volume. For Biotin pull down, 20 µL of 10mg/mL Dynabeads My One T1 Streptavidin beads washed with 400 µl of 1X Tween Wash Buffer (5mM Tris-HCl (pH 7.5), 0.5mM EDTA, 1M NaCl, 0.05% Tween) supernatant removed after separation on a magnet. Beads were resuspended with 2X Binding Buffer (10mM Tris-HCl (pH 7.5), 1mM EDTA, 2M NaCl), added to the sample and incubated for 15 minutes at room temperature. Beads were subsequently washed twice with 1X Tween Wash Buffer and in between heated on a thermomixer for 2 minutes at 55°C with mixing and once washed once with 1X NEB T4 DNA ligase buffer. To repair fragmented ends and remove biotin from unligated ends, beads were resuspended in 88 µl 1X NEB T4 DNA ligase buffer, 2 µl 25mM dNTP mix, 5 µl 10U/µl NEB T4 PNK, 4 µl 3U/µl NEB T4 DNA Polymerase and 1 µl 5U/µl Klenow, incubated for 30 minutes at room temperature and supernatant was discarded after separation on magnet. Beads were washed twice with 1X Tween Wash Buffer with 2 minute incubation at 55°C on a thermomixer with mixing, and afterwards resuspended in 100µl 1X NEB Buffer 2. For dA-tailing, beads were resuspended in 90 µl 1X NEB Buffer 2, 5 µl 10mM dATP and 5 µl 5U/µl Klenow (exo-, NEB M0212) and incubated for 30 minutes at 30°C. Beads were then washed twice as before. Beads

were subsequently washed in 1X NEB Quick Ligation Reaction Buffer (diluted from 2X, NEB B2200S). To ligate adapters, beads were suspended in 50  $\mu$ l 1X NEB DNA Quick Ligase Buffer and 3  $\mu$ l Illumina Indexed adapters and 2  $\mu$ l of NEB DNA Quick ligase (M2200) were added mixed and incubated for 15 minutes at room temperature. Beads were washed twice with 1X Tween Wash Buffer with 2 minutes at 55°C on a thermomixer as before. Beads washed with 1X Tris Buffer once and resuspended in 50  $\mu$ l of 1x Tris Buffer. KAPA qPCR assay was performed to estimate concentration and cycle number for final PCR. Final PCR was directly amplified off the T1 beads according to the qPCR results and DNA was cleaned with 1X AMPure Cleanup and eluted in 1X Tris Buffer and sequenced paired-end.

### **Bisulfite Sequencing**

Bisulfite sequencing was performed as described in (Urich et al., 2015). DNA from C57Bl/6J and SPRET/EiJ BMDMs was isolated using the PureLink Genomic DNA Kit (Thermo Fisher Scientific) as described by the manufacturer. 2  $\mu$ g of gDNA supplemented with 0.5% non-methylated  $\lambda$ -DNA (Promega) was used as input. DNA was fragmented to a 200 bp peak size using the Covaris microTube sonicator, size selected, end-repaired, adenylated and methylated adapters ligated exactly as detailed in Urich et al. Cytosine to uracil conversion was performed with the EZ DNA methylation-Gold kit (Zymo Research) with 450ng of adapter-ligated gDNA input as described within. Following bisulfite-treated DNA purification, reactions were amplified with four cycles, purified using one volume of AMPure XP bead solution and sequenced on an Illumina NextSeq 500 for 25 million and 31 million reads for C57Bl/6J and SPRET/EiJ, respectively. Conversion rates were 99.69%.

### **Hi-C sequencing**

In situ Hi-C was performed using the Arima-HiC kit (Arima Genomics Inc) as described by the manufacturer.

### **Quantification and Statistical Analysis**

Statistical details are indicated throughout the main text, in the Figure legends and within the supplemental Tables

### **Data mapping and shifting**

Custom genomes were generated for BALB/cJ, NOD/ShiLtJ, PWK/PhJ, and SPRET/EiJ from invariant positions of the mm10 sequence with alleles replaced by those reported in the VCF files (version v3) from the Mouse Genomes Project (Keane et al., 2011). For C57BL/6J the mm10 reference genome from the UCSC genome browser was used. CHIP-seq, ATAC-seq, GRO-seq and 5'GRO-seq data was mapped to custom genomes using bowtie2 (Langmead and Salzberg, 2012) with default parameters. RNA-seq data was mapped to custom genomes using STAR (Dobin et al., 2013) with default parameters. DNA methylation data was mapped to custom genomes using Bismark (Krueger and Andrews, 2011) (bismark -n 1 -l 40), Hi-C data was mapped to custom genomes using bowtie2 (Langmead and Salzberg, 2012) with default parameters and PLAC-seq data was mapped to custom genomes using BWA (version 0.7.15-r1140) (Li and Durbin, 2009) with default parameters. For visualization and data analysis, the strain genomes were shifted to the positions of the reference genome (mm10) using MARGE (Link et al., 2018). Deletions in the strains compared to the reference were not assigned with any reads. Mapped reads on insertions in the strain were shifted to the last overlapping position between strain and reference and summed up at this position. To overlap peaks with structural variances (SV) the SV file for all strain comparisons to C57 was downloaded from the mouse genome project webpage.

### **Generation of consensus motif file**

Position Probability Matrices (PPMs) of all pairs of DNA sequence motifs were generated and all pairwise combinations were aligned with Smith-Waterman algorithm (Smith and Waterman, 1981) without gaps. For each position in the alignment a Pearson correlation was calculated. Sets of motifs that had PPMs with a Pearson correlation of 0.9 or greater were merged by iteratively aligning each PPM within the set and averaging the nucleotide frequencies at each position, similar to the STAMP approach (Mahony and Benos, 2007). The threshold for motif

finding was set to have a false-positive rate lesser than 0.001 using the Biopython module motifs (Cock et al., 2009) with the function `distribution.threshold_fpr`.

### **IDR analysis**

Transcription factor ChIP-seq experiments were performed in two replicates with corresponding input experiments. HOMER (Heinz et al., 2010) tag directories were created for both replicates and both inputs and peaks were called with HOMER for each tag directory with relaxed peak finding parameters (-L 0 -C 0 -fdr 0.9) and the corresponding input directory. For ATAC-seq, no inputs were used, but the size was set to 200bp (-L 0 -C 0 -fdr 0.9 -minDist 200 -size 200). IDR (Li, 2011) was installed using Anaconda 3. To test for reproducibility between replicates, tag directories for input and ChIP-seq were pooled and pseudo replicates were generated. Peaks were called and IDR was run on both replicates and considered good if the replicate with more peaks had less than twice the number of peaks of the other replicate. To assess reproducibility within one experiment, each replicate was randomly split into two replicates. Peaks were called and IDR was run on each pseudo replicate. Experiments were considered self-consistent if the pseudo replicate with more peaks had less than twice the number of peaks of the other pseudo replicate. After passing these two quality control steps, IDR was performed on replicates and all optimal peaks defined by IDR were used for downstream analysis. For downstream analysis, the pooled input and experiment tag directories were used.

### **Histone modification ChIP-seq quality control**

IDR is not applicable for histone modification data. Therefore, for H3K4me2 and H3K27ac ChIP-seq, peaks were called on each replicate independently with HOMER's `findPeaks -style histone`, 8-fold enrichment over the input sample and normalization to 10 million mapped reads per experiment. These peaks then were merged using HOMER's `mergePeaks -size given` and subsequently similar peaks between both replicates were identified using HOMER's

getDifferentialPeaks with the parameter `–same`. Peaks that were significantly similar ( $p$ -value < 0.001) were kept.

### **ATAC-seq and ChIP-seq analysis**

To quantify the number of differentially bound transcription factor binding site and open chromatin, the optimal peak files from the IDR analysis were merged between two strains with HOMER's mergePeaks. Subsequently the merged file was annotated with HOMER's annotatePeaks.pl with parameters `mm10 –noann –nogene` and the pooled tag directories were used. Peaks bound more than 2-fold different between strains were colored light blue, peaks bound more than 4-fold different were colored dark blue. All tag counts reported throughout the manuscript are normalized to 10 million reads per sample.

### **De novo Motif analysis**

To identify motifs enriched in peak regions over random background, HOMER's de novo motif analysis (findMotifsGenome.pl) was modified to account for differences in the strain genomes. The hand-curated motif file was used to compare enriched consensus sequences with known motifs.

### **Super enhancers**

Super enhancers were identified using H3K27ac ChIP-seq data in HOMER using the findPeaks `–style super` and `–i<input sample>` parameters.

### **Network analysis**

To calculate the effect size of mutations within a motif on binding, the r.squaredGLMM function of the MuMIn package (Barton, 2017) in R was used. The marginal R value ( $\sqrt{R^2}$ ) was multiplied by 100 and reported as effect size. The network was visualized using Cytoscape 3.5.1 (Shannon et al., 2003) scaling the edges by effect size and nodes by percentage of peaks containing the motif of interest at least once.

## **RNA-seq analysis**

To compare strain-specific gene expression, first HOMER's analyzeRepeats with the option rna and the parameters –condenseGenes, -noadj, and –count exons was used on two replicates per strain. Differential gene expression was assessed with DESeq2 using HOMER's getDiffExpression.pl with the parameters –fdr 0.01 and –log2fold 1 (for 2-fold differently expressed genes) or –log2fold 2 (for 4-fold differently expressed genes). All genes shorter than 250bp were removed and for the remaining genes the TPM (transcript per kilobase million) values were plotted and colored according to fold change (2-fold different: light blue, 4-fold different: dark blue). For ratio ratio plots the TPM values of both replicates per treatment and strain were averaged and the induction of gene expression was calculated  $\text{avg}(\text{TPM}_{\text{KLA}+1})/\text{avg}(\text{TPM}_{\text{notx} + 1})$  on a log 2 scale. To assess the difference in interferon response, 46 genes associated with interferon response were manually selected and the average TPM values for both strains and both treatments were shown and used to calculate the fold difference in KLA response.

## **WGCNA analysis**

Raw gene counts defined by HOMER's analyzeRepeats was imported in R, processed with EdgeR version 3.16.5 (Robinson et al., 2010). Genes smaller than 250 bp, with less than 1 count per million (CPM) in at least 2 samples were discarded and WGCNA analysis (Version 1.61) (Langfelder and Horvath, 2008) was performed on the remaining genes. The coexpression network was created using a softpower value of 20. Tree cutting was performed with PAM stage, minimum module size of 250 genes and a cut height of 0.99 (Langfelder et al., 2008). The modules were ordered according to number of genes, and were assigned colors from the 'Spectral' palette of RcolorBrewer (1.1-2) (Neuwirt, 2014). Module trait correlations were calculated using 'treatment', 'strain combinations' and 'treatment and strain interaction' as surrogate trait variables. *P*-values based on the correlation scores were FDR multiple testing corrected with the Stats-package (Version 4.3.3.2), and only modules-trait correlated  $\text{FDR} < 1\text{E-}$

3 were considered to be significant. Modules were annotated using Metascape (Tripathi et al., 2015).

### **GRO-seq analysis**

To compare strain-specific nascent transcription the level of nascent transcripts at the gene bodies was assessed with HOMER's analyzeRepeats with the option rna and the parameters –condenseGenes –noadj and –count genes on two replicates per strain. Differential nascent transcription was assessed with DESeq2 using HOMER's getDiffExpression.pl with the parameters –fdr 0.01 and –log2fold 1 (for 2-fold different nascent transcription) or –log2fold 2 (for 4-fold different nascent transcription). All genes shorter than 250bp were removed and for the remaining gene bodies the TPM values for the pooled tag directories were used for visualization. For ratio-ratio plots analyzeRepeats was used with option rna and the parameters –condenseGenes, -tpm and –count genes using the pooled tag directories per strain and treatment.

### **5'GRO-seq analysis**

Tag directories for two replicates were pooled into one replicate to achieve more sequencing depth. 5'GRO-seq signal was assessed with HOMER's findPeaks –style tss using the pooled GRO-seq signal tag directories as input. 5'GRO-seq peaks were merged between strains using mergePeaks and signal was quantified with annotatePeaks.pl with parameters –fragLength 1 and –strand +. To quantify the distance between the 5'GRO-seq signals and the annotated transcription start sites (TSS), the distances annotatePeaks.pl reports were used.

To determine the percentage of genetic variation within TSS affection gene expression, all TSS were merged between all possible strains combinations with HOMER's mergePeaks –d given -strand. These sites then were centered and extended by the respective distances. For TSS on the negative strand the extension was reversed. The extended regions were annotated with mutations from the strains. Furthermore, the expression of each gene was measured by RNA-seq read counts using TPM. Genes on chromosome X and chromosome Y were excluded. All genes with less than 4 TPM were filtered out. For the remaining genes, the log2 fold change

was calculated. Genes were split into similarly expressed between strains (log<sub>2</sub> fold change between -0.5 and 0.5) and differently expressed (log<sub>2</sub> fold change greater than 1 or less than -1). The genes were associated with the closest 5' GRO-seq signal and promoter regions with natural genetic variation were defined as mutated, whereas regions without any genetic variation were defined as equal sequence.

### **CRD analysis**

The optimal peak files from all five strains were merged into one large file with HOMER's mergePeaks and annotated with the tag counts from the pooled IDR tag directories. The Pearson correlation coefficient (PCC) between all pairs of peaks was calculated with the perl module Statistics::Basic using the correlation sub function. To visualize the data, a sub part of the matrix was selected and visualized in R with heatmap.2 from the gplots package.

To define CRDs 4 or more consecutive peaks with a PCC of 0.8 or greater from each peak to the start peak was selected to plot length and average numbers of peaks. To annotate the signal at CRDs PU.1 or 5' GRO-seq signal was annotated at each original peak with annotatePeaks. For 5' GRO-seq the parameters `-fragLength 1` and `-strand +` was used. The signal of all peaks within one CRD was summed up to the final signal at the CRDs. In order to differentiation between highly similar and highly different CRDs between strains, the minimal and maximal binding strength was defined by locus and the difference between minimum and maximum was calculated as  $(\max - \min) / \max$ . Loci with a score greater than 0.6 were labeled as different, whereas loci with a score smaller than 0.5 were labeled as similar.

### **Hi-C analysis**

Hi-C fastq files were mapped separately and HOMER tag directories were created with `makeTagDirectory <read1>,<read2> -tbp 1`. Hi-C interaction matrices were visualized as observed interactions versus expected with Juicebox (Durand et al., 2016) and in the WashU (Zhou and Wang, 2012) genome browser as pairwise interactions. PC1 values were calculated using HOMER's `runHiCpca.pl` with `-res 50000 -superRes 100000`. TADs were called with HOMER's `findTADsAndCPs.pl` find with parameters `-res 3000 -superRes 15000` for C57 and

SPRET independently. To remove false positive regions, filterTADsAndCPs.pl was used. To compare TADs between strains the TADs were merged using merge2Dbbed.pl and then the inclusion ratio was quantified with findTADsAndCPs.pl score with parameters TBA.

The inclusion ratio was calculated by finding the average interaction counts as a function of distance for interactions within the TAD (intra-TAD) and for interactions between the TAD and the regions of upstream and downstream of the TAD of the same size (inter-TAD). To visualize TADs in the matrices for the CRDs TADbit-analysis (Serra et al., 2017) was used to call TADs on raw count interactions matrices in a 100kb window generated by HOMER.

### **PLAC-seq analysis**

PLAC-seq data was preprocessed with PrepPlac.sh script from the FithiChIP package version 1 (<https://github.com/ay-lab/FitHiChIP>). After mapping, two separate alignments files were created for short (< 1 kb) and long (> 10 kb) range interactions. Peak calling was performed on the short distance alignment file using MACS2 (Zhang et al., 2008) (version 2.1.1.20160309). BAM and bed peak files were used as input for FitHiChIP to calculate statistically significant interactions between bin size of 5000 bp. Significant interactions were calculated for each of the strains individually. A consensus PLAC-seq interaction set was generated considering only interactions that were identified in minimally two strains ( $Q < 0.01$ ). Overlap between CRDs and the PLAC-seq consensus set were calculated and the fraction of intra-CRD, inter-CRD, CRD to none CRD region, and interactions without CRDs were calculated. A 5-times bigger size-matched and peak-matched background set of consecutive ATAC-seq peaks with a PCC smaller than 0.6 was generated. Consequentially, the number of consensus PLAC-seq interactions with the CRDs and the background was counted and compared using the Kruskal Wallis between group test.

### **DNA-Methylation analysis**

For analysis only CpG data was considered. To account for additional CpG residues generated in SPRET due to mutations all CpGs present in both strains were considered in the downstream analysis and counted as 0 when not present. HOMER tag directories were created with

parameters –format bismark –genome mm10 –checkGC –minCounts 0. The optimal number of minCounts was assessed per experiment and tag directories were re-generated with parameters –format bismark –genome mm10 –checkGC –minCounts 5 for C57 and –minCounts 7 for SPRET. To assess the differences between both strains the transcription start sites of all genes were annotated with the percentage of methylated CpGs using HOMER’s annotatePeaks with option tss and parameters –mC. Subsequently the percentage of methylated CpGs was plotted for all gene promoters, for gene promoters of genes that are 2-fold differently expressed, as well as genes that are 4-fold differently expressed.

### **MARGE analysis**

Mutation Analysis for Regulatory Genomic Elements (MARGE) (Link et al., 2018) was used to generate custom genomes and shift the mapped data back to reference coordinates.

To model the impact of a motif on the binding of the measured TFs, a linear mixed model (LMM) was used. The binding of the TF is modeled as the fixed effect motif existence with random effects locus and genotype with the lme4 package (Douglas, 2015) in R (R Development Core Team, 2016). To calculate significance for each motif, the drop1 command was used. It compares a model including motif score with a model without motif score and reports the Akaike information criterion (AIC) (Akaike, 1973) for the difference.

### **F1 analysis**

F1 data was mapped to both parental genomes. Only reads without any mismatch were considered for downstream analysis. Tag directories were generated for perfectly aligned reads per parental genome, as well as for all reads that overlap loci with differences in the parental alleles. All loci without differences were discarded. To assign allele-specific reads, the ratio of reads overlapping mutations was calculated and subsequently all perfectly aligned reads for this locus were multiplied by this ratio \* 10 and assigned to the parental genomes. Loci annotated with 0 reads in one of the F1 alleles were filtered out. For GRO-seq analysis, gene bodies with less than 4 tag counts in either one of the parental or F1 alleles were filtered out.

For CHIP-seq analysis, loci with less than 16 reads were discarded. To determine *cis*-regulation, the difference of fold change between parental alleles and F1 alleles were calculated.

### **Data visualization**

All CHIP-seq, RNA-seq, GRO-seq, 5'GRO-seq and Bisulfite data was visualized in the UCSC genome browser (Kent et al., 2002). To show interactions for Hi-C and PLAC-seq, data was uploaded to the WashU browser (Zhou and Wang, 2012).

### **Data and Software Availability**

All raw data and processed data files were deposited to GEO under accession number GSE109965.

## KEY RESOURCES TABLE

| REAGENT or RESOURCE                                  | SOURCE              | IDENTIFIER         |
|--|---------------------|--------------------|
| <b>Antibodies</b>                                    |                     |                    |
| PU.1   | Santa Cruz          | Cat#sc-352X        |
| CEBPb  | Santa Cruz          | Cat#sc-150         |
| H3K4me2  | Millipore           | Cat#07-030         |
| H3K4me3  | Millipore           | Cat#04-745         |
| H3K27ac  | Active Motif        | Cat#39135          |
| cJun   | Santa Cruz          | Cat#sc-1694        |
| p65  | Santa Cruz          | Cat#sc-372X        |
| Usf2   | Santa Cruz          | Cat#sc-862X        |
| Runx1  | Santa Cruz          | Cat#sc365644       |
| <b>Chemicals, Peptides, and Recombinant Proteins</b> |                     |                    |
| DMEM high glucose                                    | Corning             | Cat#10-013-CV      |
| FBS  | Omega Biosciences   | Cat#FB-12          |
| 100X Penicillin/Streptomycin+L-glutamine             | Gibco               | Cat#10378-016      |
| Amphotericin B                                       | Hyclone             | Cat#SV30078.01     |
| RBC lysis buffer                                     | eBioscience         | Cat#00-4333-57     |
| mouse M-CSF  | Shenandoah Biotech  | Cat#200-08         |
| Kdo2 Lipid A (KLA)                                   | Avanti Polar Lipids | Cat#699500P        |
| Dynabeads Protein G                                  | Thermo Fischer      | Cat#10004D         |
| Speedbeads   | GE Healthcare       | Cat#65152105050250 |
| BrdU Antibody (IIB5) AC beads                        | Santa Cruz          | Cat#sc-32323 AC    |
| Millipore Ultrafree MC column                        | Millipore           | Cat#UFC30HVN       |
| Trizol LS  | Thermo Fischer      | Cat#10296010       |
| Dynabeads My One T1 Streptavidin beads               | Thermo Fischer      | Cat#65601          |
| <b>Critical Commercial Assays</b>                    |                     |                    |
| ChIP DNA Clean & Concentrator                        | Zymo Research       | Cat#D5205          |
| Quick RNA MiniPrep kit                               | Zymo Research       | Cat#R1055          |
| PureLink Genomic DNA Kit                             | Thermo Fischer      | Cat#K182001        |
| EZ DNA methylation-Gold kit                          | Zymo Research       | Cat#D5005          |
| Nextera DNA Library Prep Kit                         | Illumina            | Cat#FC-121-1030    |
| NEXTflex® DNA Barcodes                               | Bioo Scientific     | Cat#NOVA-514104    |
| Arima HiC kit  | Arima Genomics Inc  | NA                 |
| AMPure Cleanup                                       | Beckman Coulter     | CAT#A63880         |
| <b>Deposited Data</b>                                |                     |                    |
| Raw and analyzed data                                | This paper          | GEO:               |
| <b>Experimental Models: Organisms/Strains</b>        |                     |                    |
| BALB/cJ  | Jackson Laboratory  | Stock No: 000651   |
| C57Bl/6J   | Jackson Laboratory  | Stock No: 000664   |

|  |                                     |   |
|--|-------------------------------------|---|
| NOD/ShiLtJ                             | Jackson Laboratory                  | Stock No: 001976  |
| PWK/PhJ                                | Jackson Laboratory                  | Stock No: 003715  |
| SPRET/EiJ                              | Jackson Laboratory                  | Stock No: 001146  |
| <b>Software and Algorithms</b>         |                                     |   |
| Bowtie2                                | Langmead and Salzberg, 2012         | <a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>   |
| UCSC genome browser                    | Kent et al., 2002                   | <a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a>   |
| STAR                                   | Dobin et al., 2013                  | <a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>   |
| Bismark                                | Krueger and Andrews, 2011           | <a href="https://www.bioinformatics.babraham.ac.uk/projects/bismark/">https://www.bioinformatics.babraham.ac.uk/projects/bismark/</a>                                 |
| MARGE                                  | In preparation                      | <a href="https://github.com/vlink/marge">https://github.com/vlink/marge</a>   |
| Irreproducibility Discovery Rate (IDR) | Li et al., 2011                     | <a href="https://www.encodeproject.org/software/idr/">https://www.encodeproject.org/software/idr/</a>   |
| HOMER                                  | Heinz et al., 2010                  | <a href="http://homer.ucsd.edu/homer/">http://homer.ucsd.edu/homer/</a>   |
| Cytoscape 3.5.1                        | Shannon et al., 2013                | <a href="http://www.cytoscape.org/">http://www.cytoscape.org/</a>   |
| TADbit                                 | Serra et al. 2017                   | <a href="https://github.com/3DGenomes/TADbit">https://github.com/3DGenomes/TADbit</a>   |
| Juicebox                               | Durand et al., 2016                 | <a href="http://www.aidenlab.org/software.html">http://www.aidenlab.org/software.html</a>   |
| WashU Browser                          | Zhou and Wang, 2013                 | <a href="http://epigenomegateway.wustl.edu/browser/">http://epigenomegateway.wustl.edu/browser/</a>   |
| R package: MuMIn                       | Barton 2017                         | <a href="https://cran.r-project.org/web/packages/MuMIn/index.html">https://cran.r-project.org/web/packages/MuMIn/index.html</a>                                       |
| R package: DESeq2                      | Love et al., 2014                   | <a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>                         |
| R package: EdgeR                       | Robinson, McCarthy, and Smyth 2010  | <a href="http://bioconductor.org/packages/release/bioc/html/edgeR.html">http://bioconductor.org/packages/release/bioc/html/edgeR.html</a>                             |
| R-package: WGCNA                       | Langfelder and Horvath 2008         | <a href="https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/">https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/</a> |
| R-package: Dynamic Tree cutting        | Langfelder, Zhang, and Horvath 2008 | <a href="https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/BranchCutting/">https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/BranchCutting/</a>     |
| R-package: RcolorBrewer                | Neuwirt 2014                        | <a href="https://cran.r-project.org/web/packages/RColorBrewer/index.html">https://cran.r-project.org/web/packages/RColorBrewer/index.html</a>                         |

|                                |                      |   |
|--------------------------------|----------------------|---|
| R-package: gplots              |                      | <a href="https://cran.r-project.org/web/packages/gplots/index.html">https://cran.r-project.org/web/packages/gplots/index.html</a>   |
| Metascape                      | Tripathi et al. 2015 | <a href="http://metascape.org">http://metascape.org</a>   |
| Perl module: Statistics::Basic | Miller 2014          | <a href="http://search.cpan.org/~jettero/Statistics-Basic-1.6611/lib/Statistics/Basic.pod">http://search.cpan.org/~jettero/Statistics-Basic-1.6611/lib/Statistics/Basic.pod</a> |

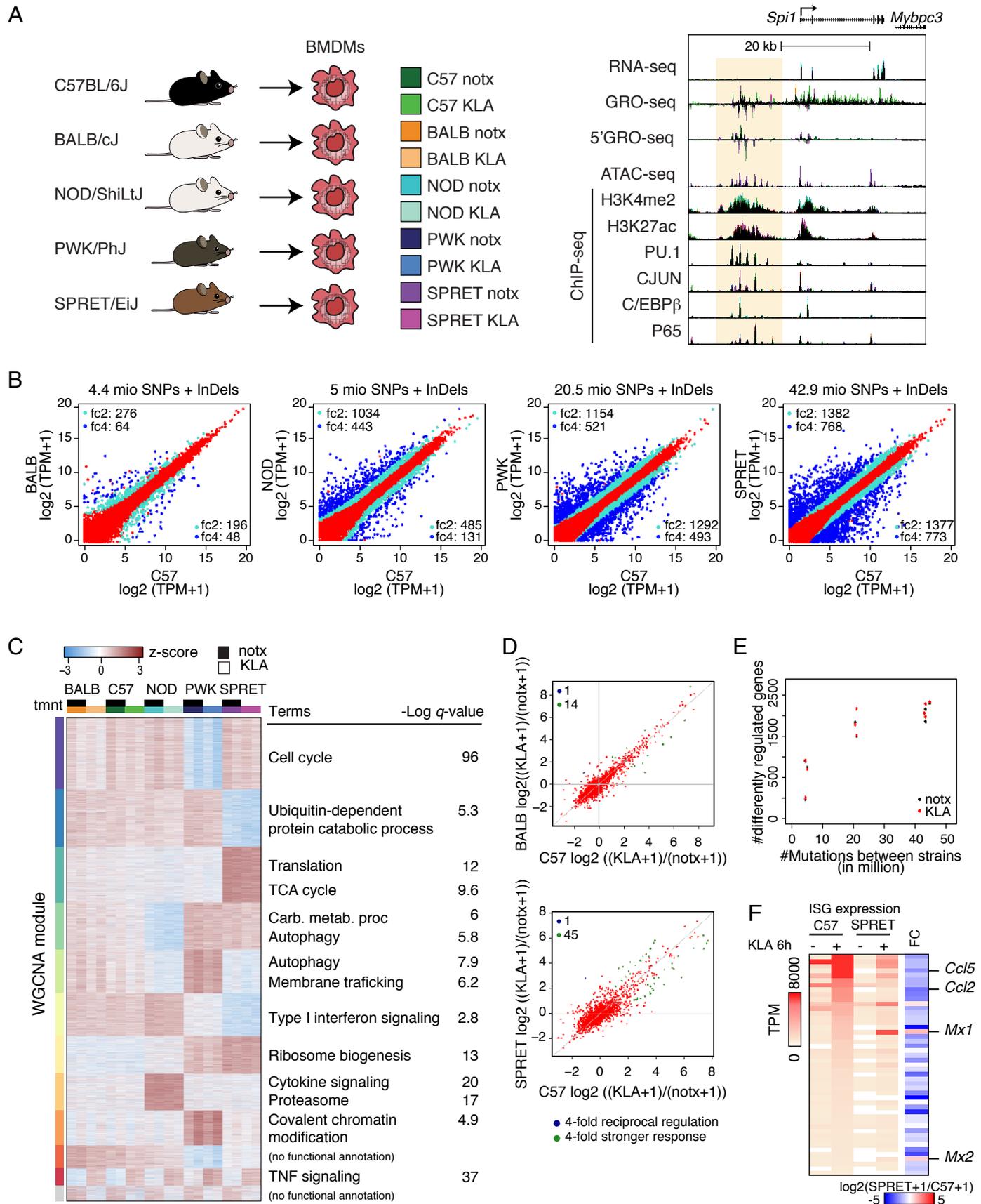


Figure 1

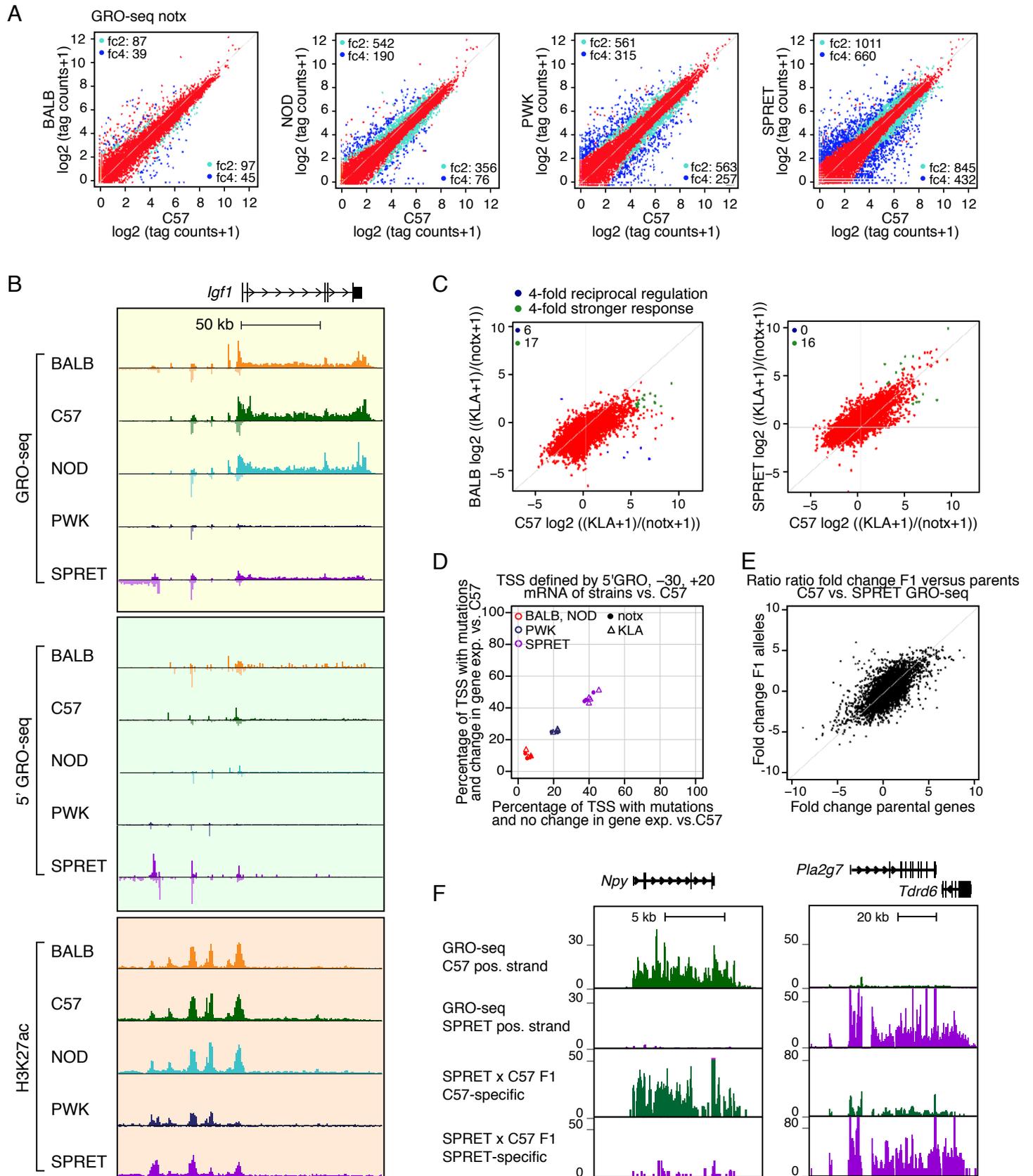


Figure 2

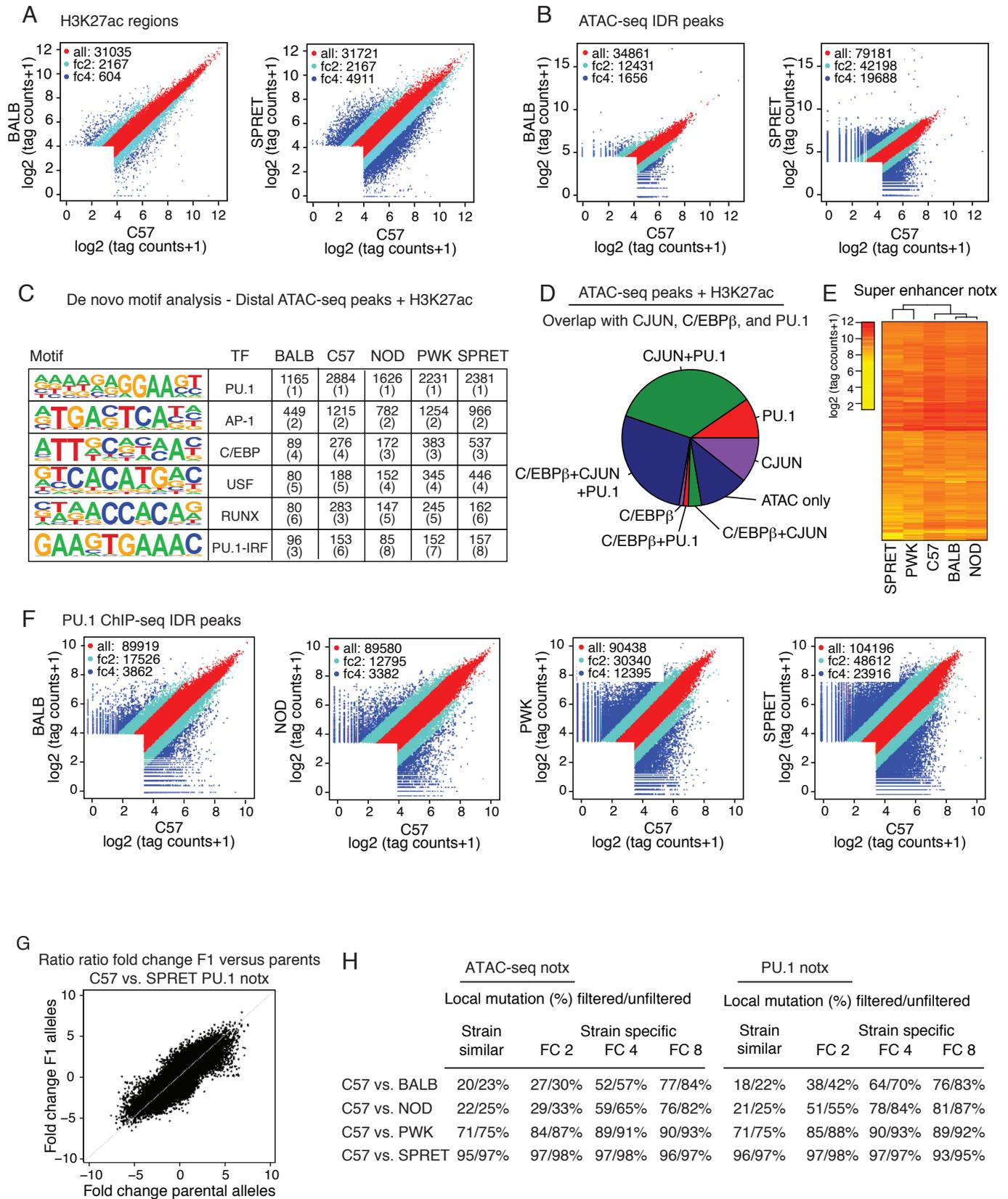


Figure 3

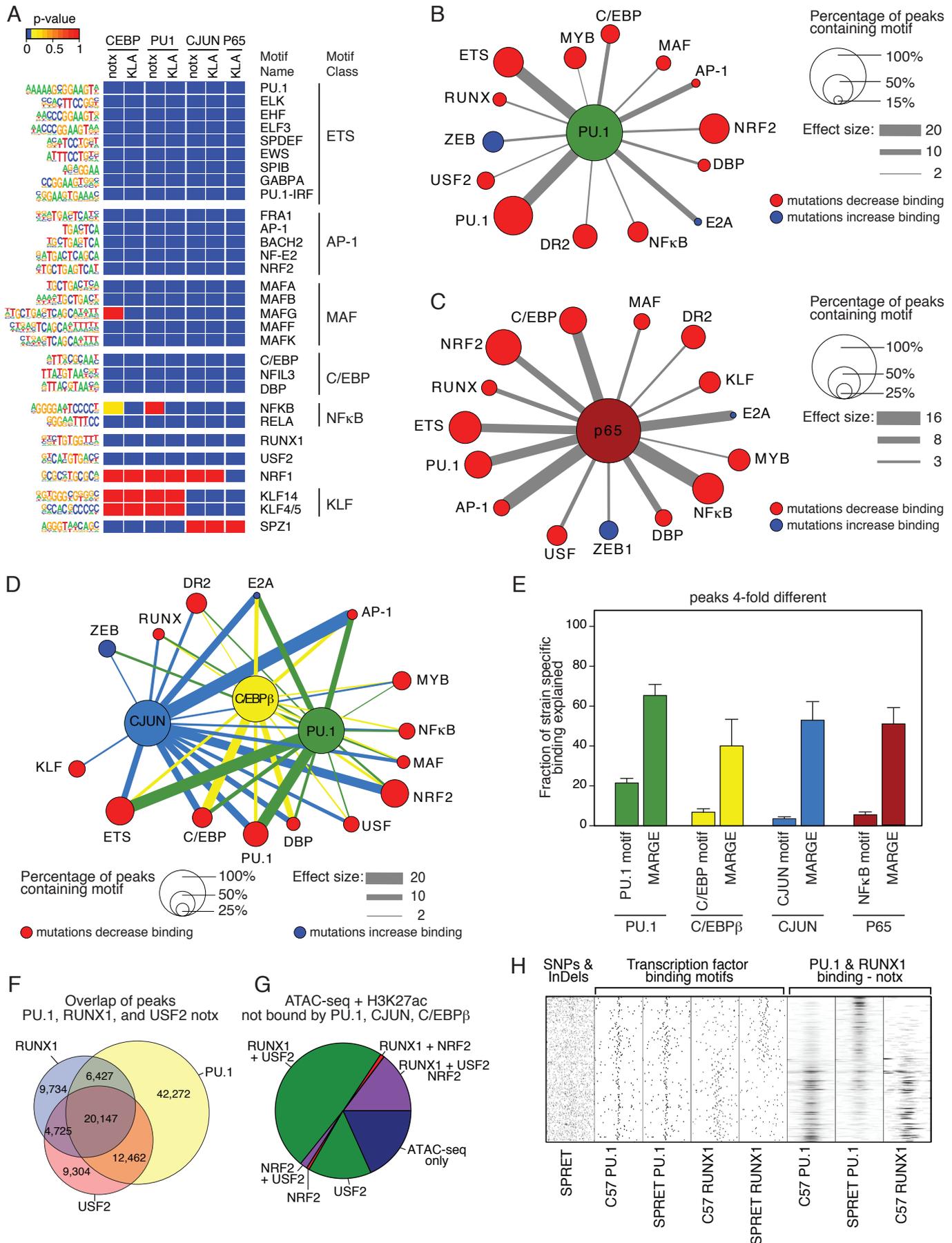


Figure 4

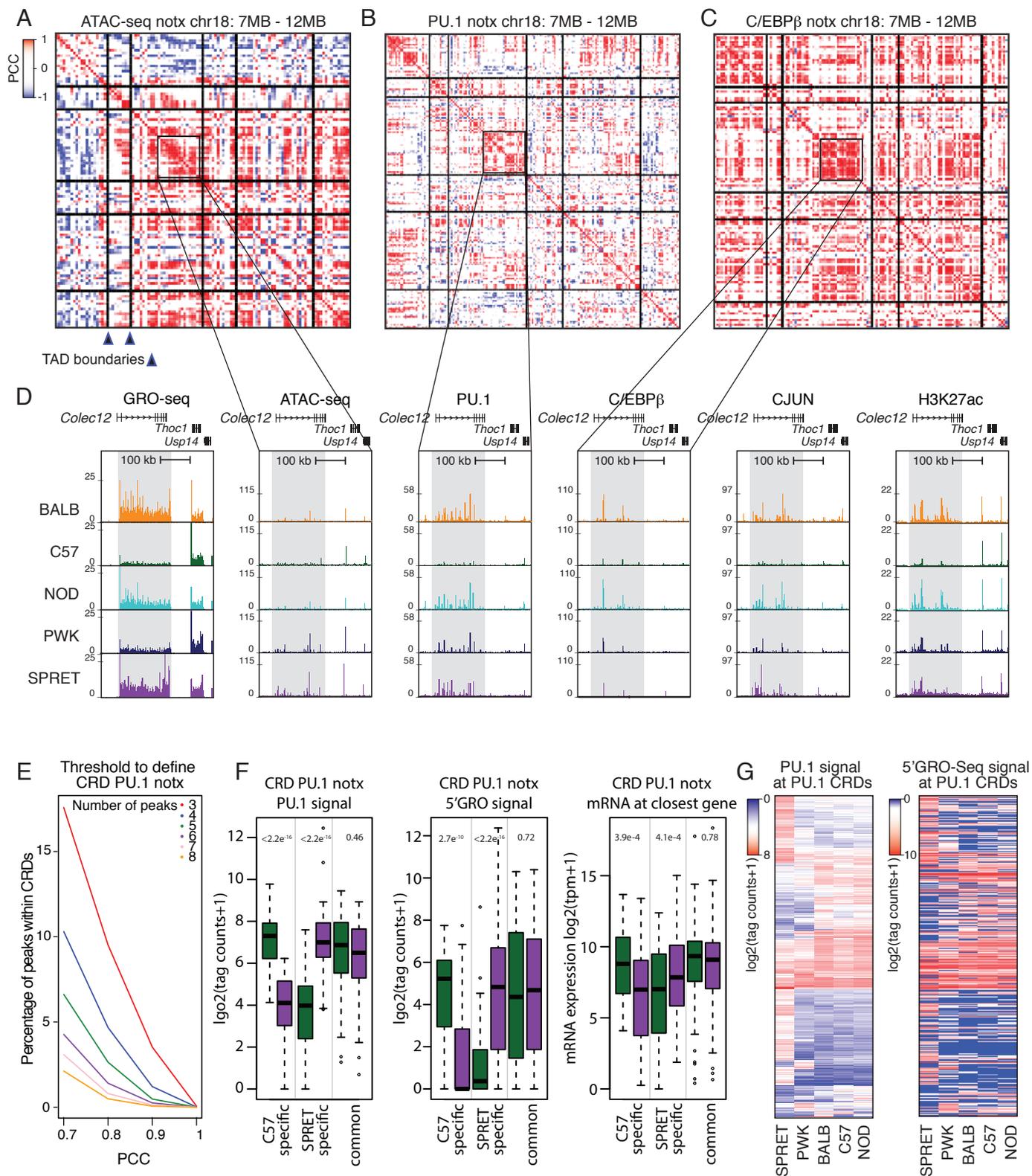


Figure 5

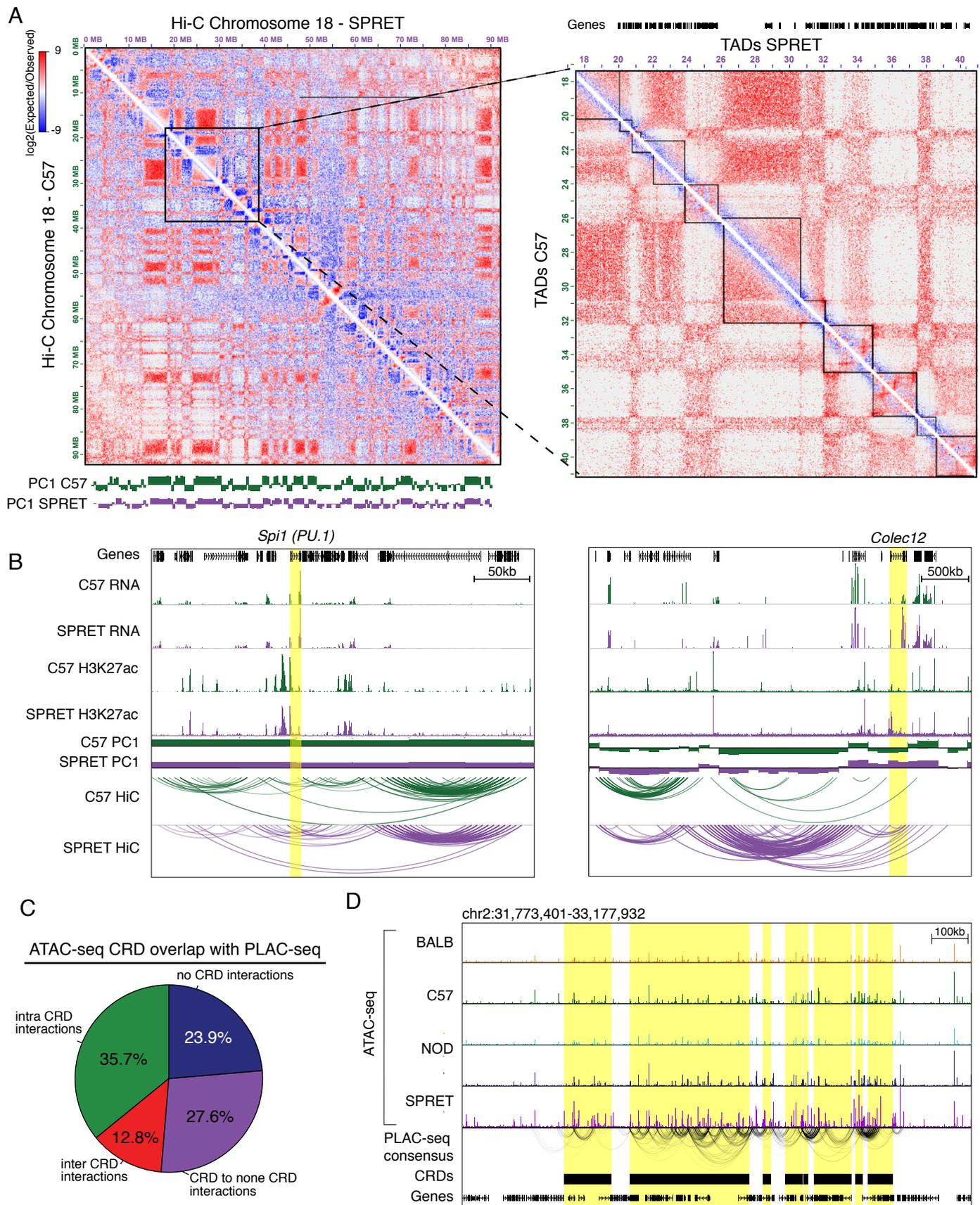
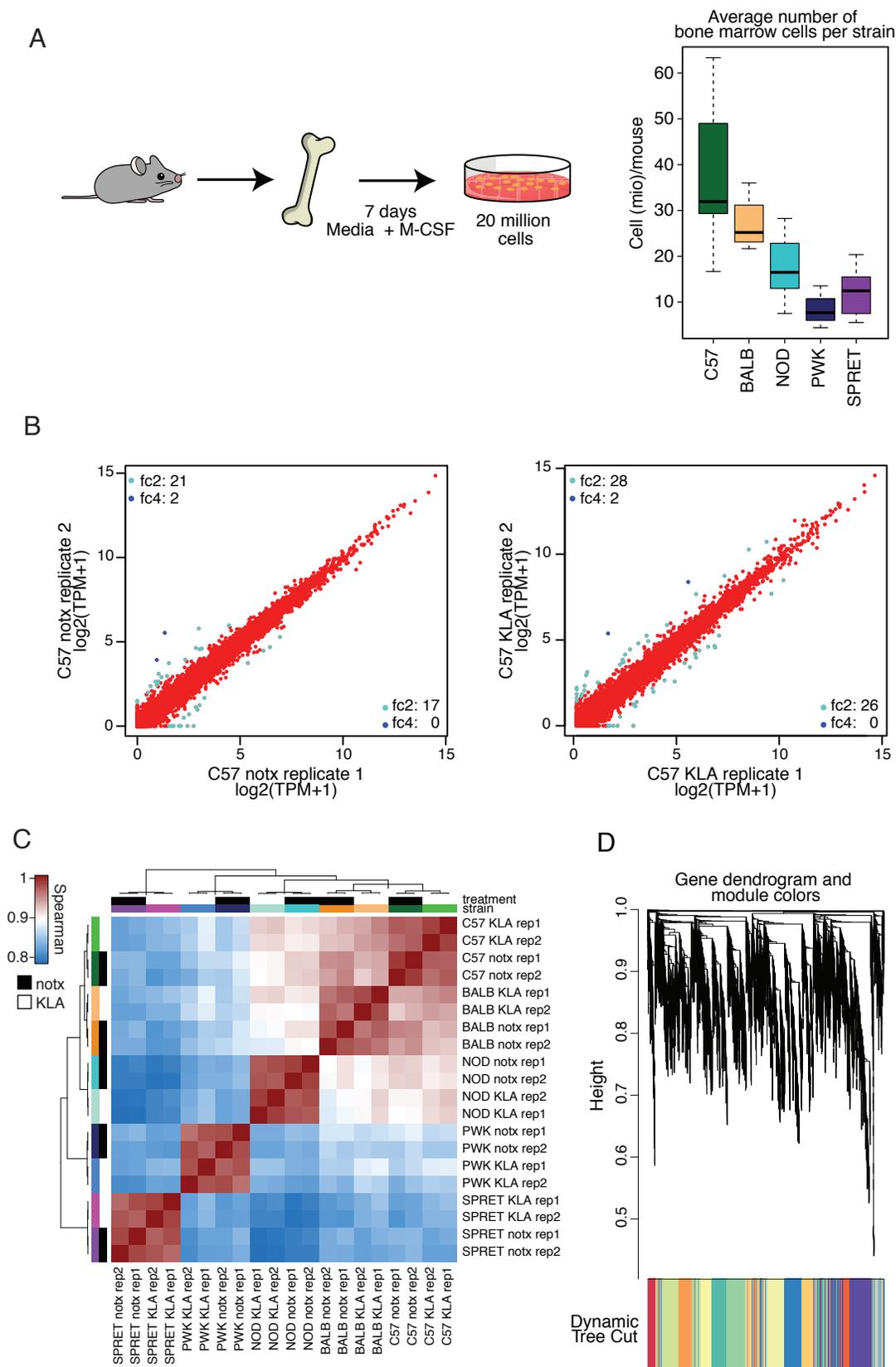


Figure 6



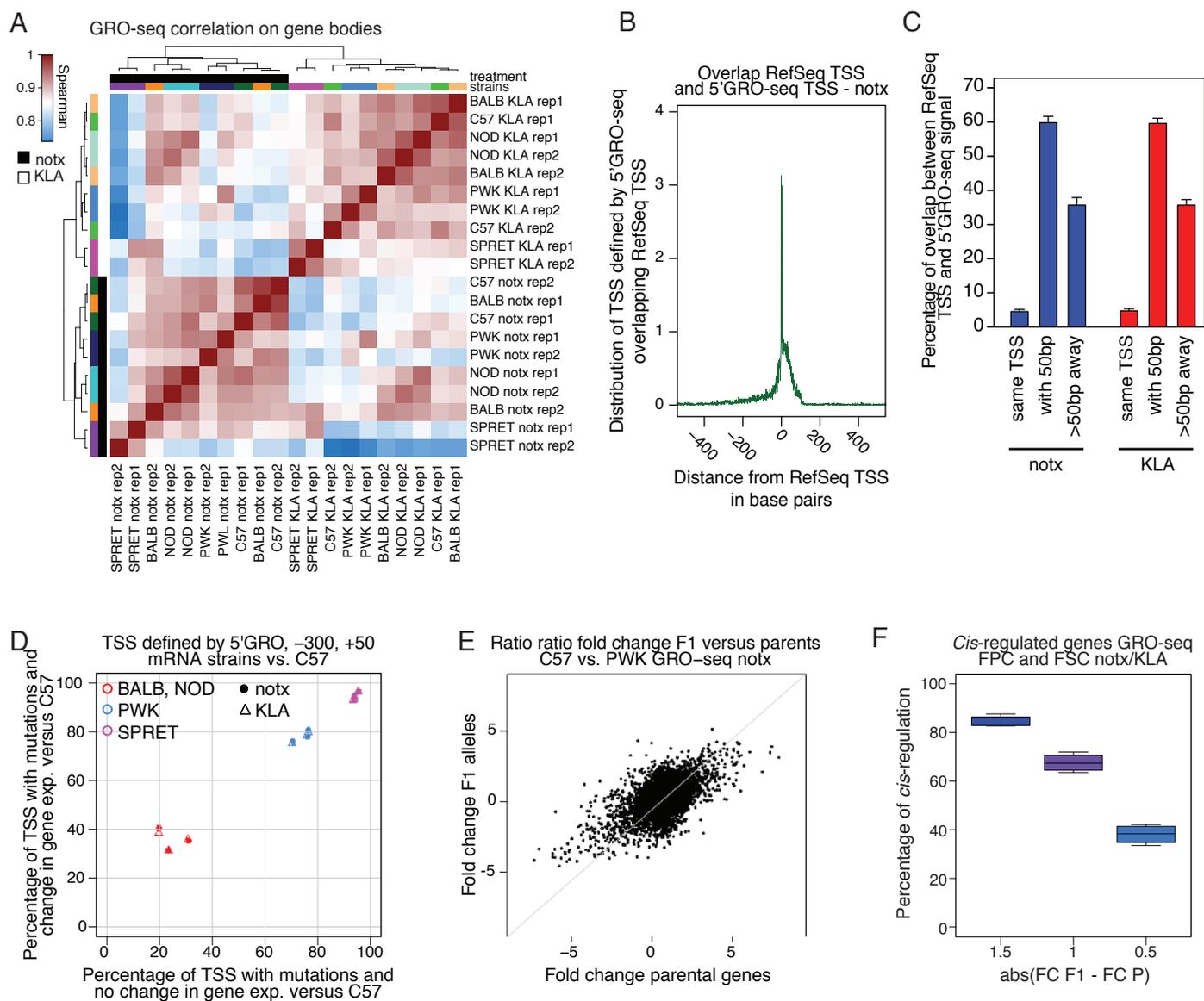
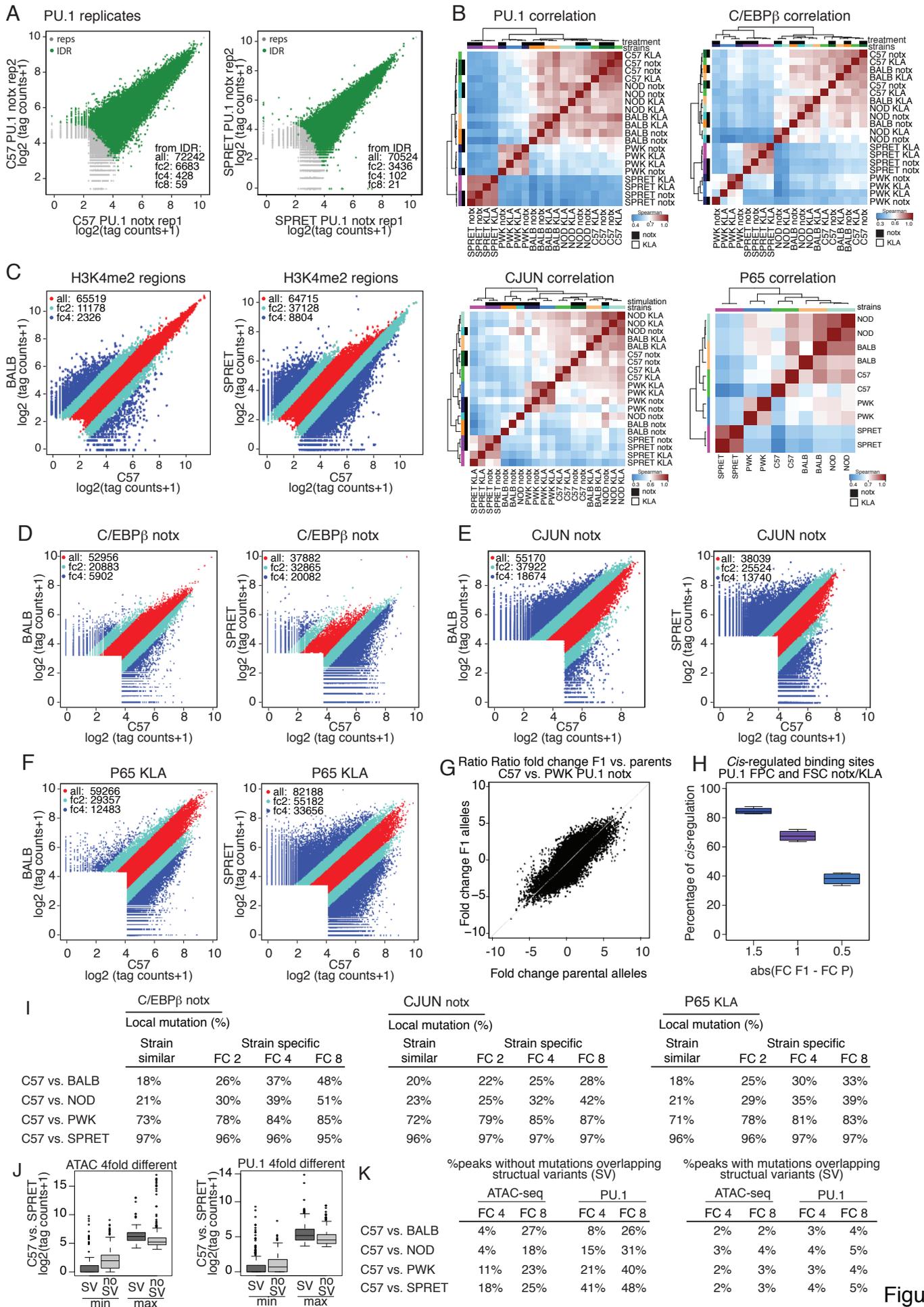
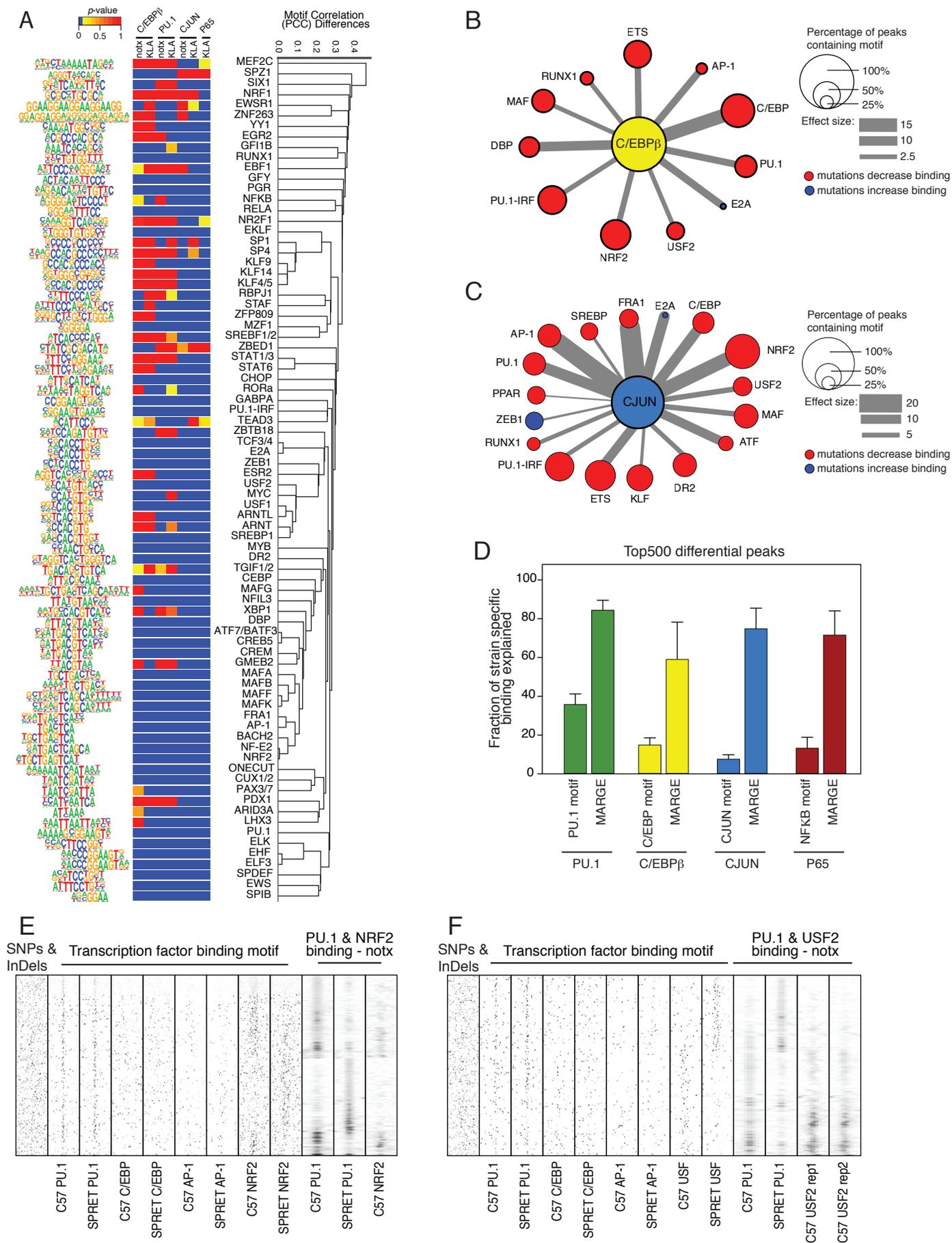
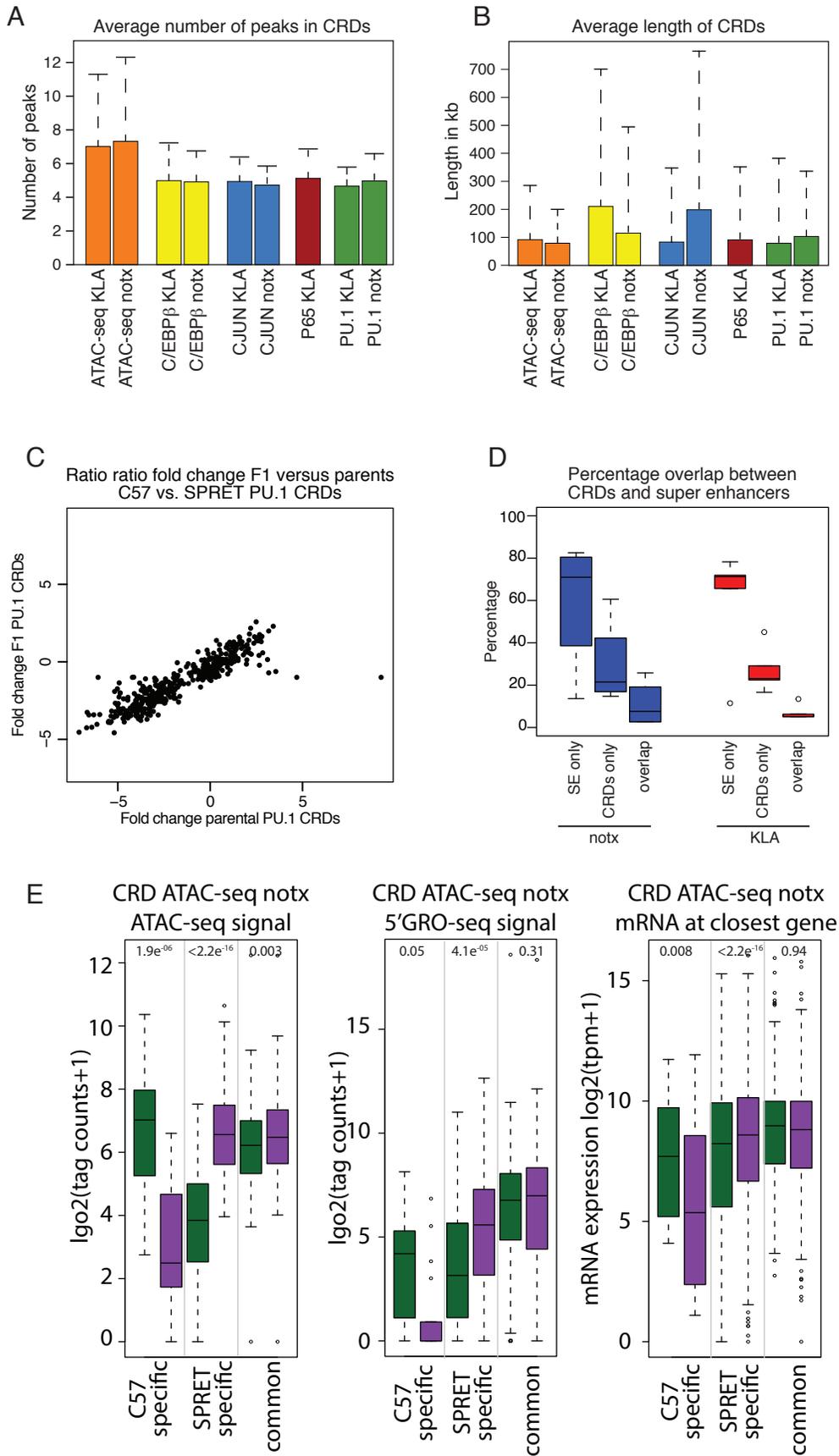


Figure S2











# Chapter 5

## Additional contributions

### **SREBP1 Contributes to Resolution of Pro-inflammatory TLR4 Signaling by Reprogramming Fatty Acid Metabolism**

Yumiko Oishi, Nathanael J. Spann, **Verena M. Link**, Evan D. Muse, Tobias Strid, Chantle Edillor, Matthew J. Kolar, Takashi Matsuzaka, Sumio Hayakawa, Jenhan Tao, Minna U. Kaikkonen, Aaron F. Carlin, Michael T. Lam, Ichiro Manabe, Hitoshi Shimano, Alan Saghatelian, and Christopher K. Glass

*Cell Metabolism* (2017) 25:412–427

### **Sympathetic neuron-associated macrophages contribute to obesity by importing and metabolizing norepinephrine**

Roksana M. Pirzgalska, Elsa Seixas, Jason S. Seidman, **Verena M. Link**, Noelia Martinez Snchez, Ins Mah, Raquel Mendes, Vitka Gres, Nadiya Kubasova, Imogen Morris, Bernardo A. Ars, Chelsea M. Larabee, Miguel Vasques, Francisco Tortosa, Ana L. Sousa, Sathya-vathy Anandan, Erin Tranfield, Maureen K. Hahn, Matteo Iannaccone, Nathanael J. Spann, Christopher K. Glass and Ana I. Domingos

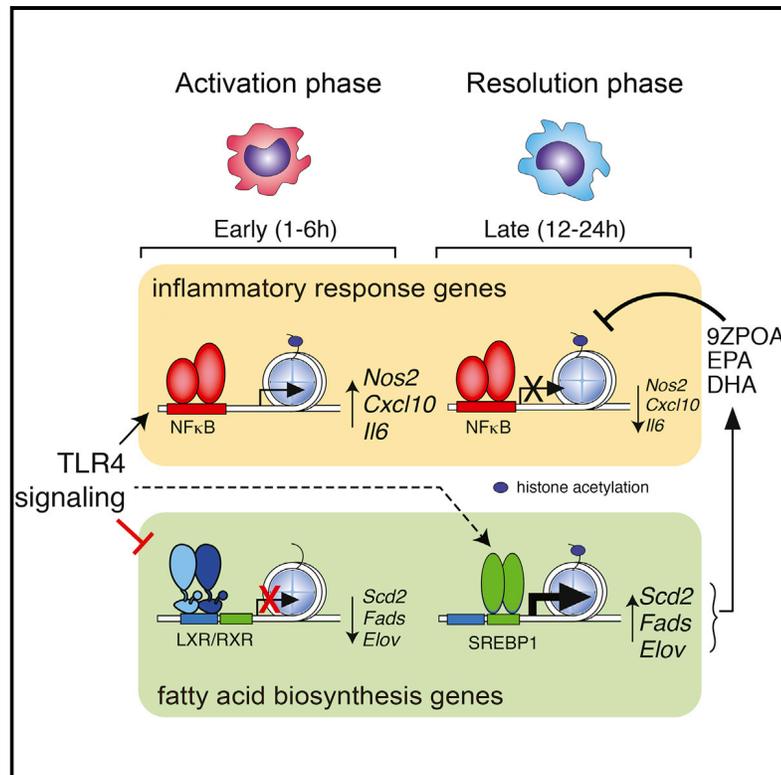
*Nature Medicine* (2017) 23:1309–1318



# Cell Metabolism

## SREBP1 Contributes to Resolution of Pro-inflammatory TLR4 Signaling by Reprogramming Fatty Acid Metabolism

### Graphical Abstract



### Authors

Yumiko Oishi, Nathanael J. Spann, Verena M. Link, ..., Hitoshi Shimano, Alan Saghatelian, Christopher K. Glass

### Correspondence

yuoishi-circ@umin.ac.jp (Y.O.), ckg@ucsd.edu (C.K.G.)

### In Brief

Oishi and Spann et al. identify a biphasic macrophage gene expression program underlying anti-inflammatory fatty acid production following TLR4 activation. The late anti-inflammatory program is dependent on SREBP1. This is surprising given the known roles of SREBP1 in promoting IL1 $\beta$  production during the induction of inflammation and highlights its dual roles.

### Highlights

- TLR ligands induce reciprocal inflammatory and lipid biosynthetic gene expression
- SREBP1 levels and activity are increased 12–24 hr following TLR4 activation
- SREBP1 drives the production of anti-inflammatory fatty acids in macrophages
- Loss of SREBP1 results in impaired resolution of inflammatory responses

### Accession Numbers

GSE79423



# SREBP1 Contributes to Resolution of Pro-inflammatory TLR4 Signaling by Reprogramming Fatty Acid Metabolism

Yumiko Oishi,<sup>1,2,9,\*</sup> Nathanael J. Spann,<sup>1,9</sup> Verena M. Link,<sup>1,8</sup> Evan D. Muse,<sup>1,3</sup> Tobias Strid,<sup>1</sup> Chantle Edillor,<sup>1</sup> Matthew J. Kolar,<sup>4</sup> Takashi Matsuzaka,<sup>5</sup> Sumio Hayakawa,<sup>2</sup> Jenhan Tao,<sup>1</sup> Minna U. Kaikkonen,<sup>1,6</sup> Aaron F. Carlin,<sup>1</sup> Michael T. Lam,<sup>1</sup> Ichiro Manabe,<sup>7</sup> Hitoshi Shimano,<sup>5</sup> Alan Saghatelian,<sup>4</sup> and Christopher K. Glass<sup>1,10,\*</sup>

<sup>1</sup>Department of Cellular and Molecular Medicine, School of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Department of Cellular and Molecular Medicine, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

<sup>3</sup>Scripps Translational Science Institute, La Jolla, CA 92037, USA

<sup>4</sup>Salk Institute for Biological Studies, La Jolla, CA 92037, USA

<sup>5</sup>Department of Internal Medicine (Endocrinology and Metabolism), Faculty of Medicine, Graduate School of Comprehensive Human Sciences, International Institute for Integrative Sleep Medicine (WPI-IIS), and Center for Tsukuba Advanced Research Alliance, University of Tsukuba, Ibaraki Prefecture 305-8571, Japan

<sup>6</sup>Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, P.O. Box 1627, 70211 Kuopio, Finland

<sup>7</sup>Department of Aging Research, Graduate School of Medicine, Chiba University, Chiba 260-8670, Japan

<sup>8</sup>Department II, Faculty of Biology, Ludwig-Maximilians Universität München, Planegg-Martinsried 82152, Germany

<sup>9</sup>Co-first author

<sup>10</sup>Lead Contact

\*Correspondence: [yuoishi-circ@umin.ac.jp](mailto:yuoishi-circ@umin.ac.jp) (Y.O.), [ckg@ucsd.edu](mailto:ckg@ucsd.edu) (C.K.G.)

<http://dx.doi.org/10.1016/j.cmet.2016.11.009>

## SUMMARY

Macrophages play pivotal roles in both the induction and resolution phases of inflammatory processes. Macrophages have been shown to synthesize anti-inflammatory fatty acids in an LXR-dependent manner, but whether the production of these species contributes to the resolution phase of inflammatory responses has not been established. Here, we identify a biphasic program of gene expression that drives production of anti-inflammatory fatty acids 12–24 hr following TLR4 activation and contributes to downregulation of mRNAs encoding pro-inflammatory mediators. Unexpectedly, rather than requiring LXRs, this late program of anti-inflammatory fatty acid biosynthesis is dependent on SREBP1 and results in the uncoupling of NF $\kappa$ B binding from gene activation. In contrast to previously identified roles of SREBP1 in promoting production of IL1 $\beta$  during the induction phase of inflammation, these studies provide evidence that SREBP1 also contributes to the resolution phase of TLR4-induced gene activation by reprogramming macrophage lipid metabolism.

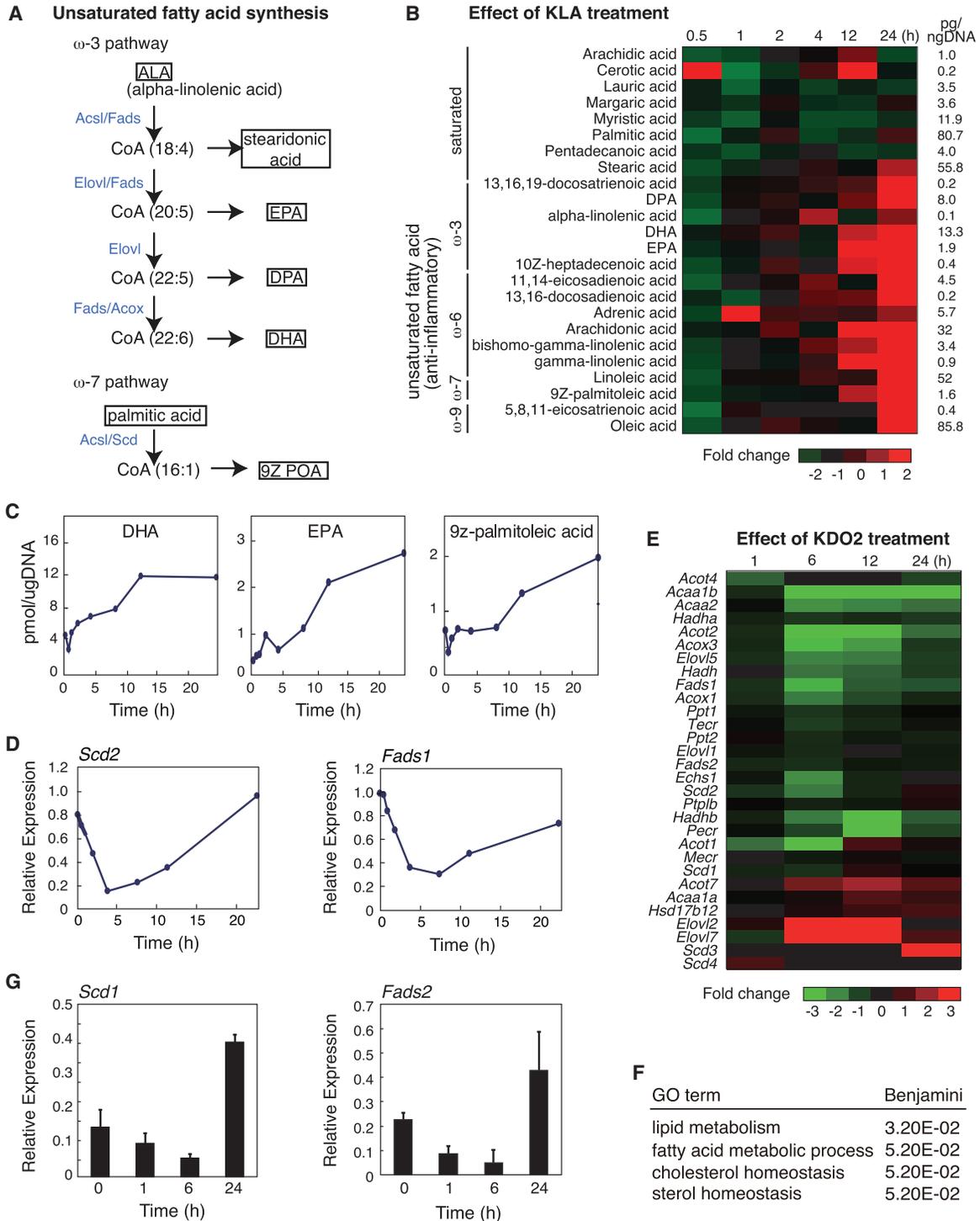
## INTRODUCTION

Failure to resolve endogenous or extrinsic inflammatory stimuli can lead to a chronic state of low-grade inflammation that results in cellular dysfunction and tissue damage (Tabas and Glass, 2013). Recent studies have shown that the immune and metabolic systems are highly integrated with one another (Cildir

et al., 2013). For instance, increased infiltration of pro-inflammatory macrophages in adipose tissue, liver, and skeletal muscle and their release of cytokines that impair local insulin signaling contribute to insulin resistance (Lumeng et al., 2008; Osborn and Olefsky, 2012; Tencerova et al., 2015; Varma et al., 2009; Wynn et al., 2013; Xu et al., 2003). In addition, immune cell function itself is coordinately regulated with cellular metabolism (Spann and Glass, 2013). For example, upon inflammatory activation, macrophages rapidly induce glycolysis through HIF-1 $\alpha$  and NF $\kappa$ B, enabling them to trigger microbicidal activity even in a hypoxic inflammatory tissue environment (Huang et al., 2014a; Rodríguez-Prados et al., 2010; Tannahill et al., 2013). In contrast, macrophages display a shift to oxidative metabolism of glucose and fatty acids and acquire an anti-inflammatory phenotype in the context of tissue repair and remodeling (Mantovani et al., 2013; Rodríguez-Prados et al., 2010).

Macrophage activation in response to ligation of TLR4 provides a paradigm for investigation of molecular mechanisms that positively and negatively regulate inflammatory responses (Iyer et al., 2010; Medzhitov and Horng, 2009). TLR4 signaling induces immediate/early gene expression through activation of latent transcription factors that include members of the NF $\kappa$ B, IRF, and AP-1 families (Glass and Natoli, 2016; Medzhitov and Horng, 2009). These factors in turn induce secondary response genes via the production of type I interferons, TNF $\alpha$ , and other signaling molecules. Collectively, the immediate/early and secondary responses drive expression of inflammatory response genes that support innate immunity and set the stage for adaptive immunity. TLR4 signaling also results in downregulation of a broad program of gene expression, although molecular mechanisms are less well characterized.

Recent lipidomic analysis in macrophages revealed an immediate reduction of fatty acid synthesis in response to TLR4 activation, followed by an increase in eicosanoid synthesis that was



**Figure 1. Activation of TLR4 Reprograms Macrophage Fatty Acid Metabolism**

(A) Pathway maps illustrating omega-3 and omega-7 pathways. The enzymes catalyzing each step are highlighted in blue.

(B) Lipidomic analysis of saturated and unsaturated fatty acids (omega-3, omega-6, omega-7, and omega-9) in thioglycollate-elicited macrophages treated with KLA for 0, 1, 6, 12, and 24 hr.

(C) Cellular content of omega-3 (DHA and EPA) and omega-7 (9Z-POA) fatty acids in thioglycollate-elicited macrophages treated with KLA for 0, 1, 6, 12, and 24 hr.

(D) Relative mRNA expression levels for *Scd2* and *Fads1* determined by microarray analysis of RNA from thioglycollate-elicited macrophages treated with KLA for 0, 1, 6, and 24 hr.

(E) Heatmap of mRNA expression levels determined by RNA-seq analysis of BMDMs with KLA for 0, 1, 6, and 24 hr (FDR <0.01 and RPKM >0.5).

(legend continued on next page)

linked to the arachidonic acid pathway and delayed responses characterized by sphingolipid and sterol biosynthesis (Dennis et al., 2010). Lipid uptake is activated by chronic (~24 hr) LPS treatment, leading triglycerides to accumulate in lipid droplets within macrophages (Feingold et al., 2012; Huang et al., 2014b). These changes in lipid metabolism may be linked to changes in macrophage activity over the time course of the response to LPS.

Macrophages can also synthesize anti-inflammatory fatty acids under the control of liver X receptors (LXRs)  $\alpha$  and  $\beta$  (Li et al., 2013; Spann et al., 2012). The LXR pathway is derepressed following genetic deletion of the nuclear receptor co-repressor NCoR, leading to increased production of 9Z palmitoleic acid and polyunsaturated omega-3 and omega-9 fatty acids (Li et al., 2013). These fatty acids exert anti-inflammatory functions in macrophages in part by binding to G protein coupled receptors (Oh et al., 2010). NCoR deletion in macrophages conferred protection of mice from high fat diet-induced inflammation and insulin resistance (Li et al., 2013). Therefore, it is possible that anti-inflammatory fatty acids produced by the macrophage act in an autocrine/paracrine manner to regulate its function autonomously, as well as the functions of surrounding parenchymal cells.

Cholesterol and fatty acid homeostasis are regulated at the level of transcription by LXRs and SREBPs 1 and 2 (Goldstein et al., 2006; Hong and Tontonoz, 2014; Horton et al., 2002). Their roles in cholesterol homeostasis are largely antagonistic. SREBPs (primarily SREBP2) drive transcriptional programs that increase cellular cholesterol synthesis and import (Horton et al., 2002), while LXRs induce expression of genes that mediate cholesterol efflux and inhibit import (Hong and Tontonoz, 2014). In contrast, LXRs and SREBPs (primarily SREBP1) function in a coordinate manner to positively regulate fatty acid biosynthesis. LXRs directly activate the expression of SREBP1c, and both LXRs and SREBP1 bind to and activate numerous genes involved in fatty acid biosynthesis (Repa et al., 2000a; Schultz et al., 2000). Further, at co-bound genomic loci, SREBP functions in a permissive manner, allowing signal-specific tailoring of LXR-mediated activation of lipid metabolic gene expression profiles (Spann et al., 2012), resulting in context-dependent synthesis and output of select lipid species.

LXRs and SREBPs also play important roles in regulating macrophage activation. LXRs primarily function to inhibit inflammatory responses by antagonizing pro-inflammatory transcription factors, such as NF $\kappa$ B (Ghisletti et al., 2009; Hong and Tontonoz, 2014), and by activating genes with anti-inflammatory activities, such as *Mer* and *Abca1* (A-Gonzalez et al., 2009; Ito et al., 2015). In contrast, SREBP1 has been found to promote the acute inflammatory response by regulating genes involved in the production of active IL1 $\beta$  (Im et al., 2011; Reboli et al., 2014). Further, the LXR pathway is subject to negative regulation by TLR4 (Castrillo et al., 2003). This suggests that macrophage fatty acid synthesis is influenced by TLR signaling via temporal modulation of LXR activities.

To address the question of whether TLR4 signaling regulates the production of anti-inflammatory fatty acids, we analyzed lipidomic data generated by the LIPID MAPS consortium evaluating the temporal response of primary mouse macrophages to the specific TLR4 agonist Kdo2 LIPID A (KLA) (<http://www.lipidmaps.org/>) (Dennis et al., 2010). This analyses revealed that the intracellular content of anti-inflammatory mono- and poly- ( $\omega$ -3,  $\omega$ -7, and  $\omega$ -9) unsaturated fatty acids was rapidly decreased at early time points of TLR4-mediated inflammation; while the resolution phase was characterized by increased intracellular unsaturated fatty acid levels. This temporal pattern of changes in specific lipid species was correlated with changes in mRNAs encoding corresponding biosynthetic enzymes. Unexpectedly, we found that the late upregulation of unsaturated fatty acid synthesis was independent of LXR, but was instead driven by SREBP1. Anti-inflammatory fatty acid synthesis was compromised in *Srebf1*<sup>-/-</sup> macrophages at late time points compared to wild-type (WT) macrophages, concomitant with a hyper-inflammatory state due to impaired resolution of NF $\kappa$ B associated activity and gene expression. Supplementation with exogenous mono- and polyunsaturated fatty acids rescues the late hyper-inflammatory response in both *Srebf1*<sup>-/-</sup> macrophages and *Srebf1*<sup>-/-</sup> mice. Collectively, these findings provide evidence that SREBP1 contributes to resolution of pro-inflammatory TLR4 signaling by reprogramming fatty acid metabolism.

## RESULTS

### TLR4 Signaling Reprograms Macrophage Fatty Acid Metabolism

To investigate changes in macrophage fatty acid levels throughout the course of an inflammatory response, we utilized lipidomic data generated by the LIPID MAPS Consortium (<http://www.lipidmaps.org/>) (Dennis et al., 2010). Metabolic pathways responsible for generation of long chain omega-3 fatty acids and 9Z palmitoleic acid (9Z-POA) are shown in Figure 1A. Activation of TLR4 by KLA, a chemically defined substructure of bacterial lipopolysaccharide (LPS) that is specifically recognized by Toll-like receptor 4 (Raetz et al., 2006), rapidly and transiently decreased the cellular content of most fatty acids analyzed (Figure 1B). Unexpectedly, in addition to the known upregulation of omega-6 fatty acids, such as arachidonic acid, the cellular content of anti-inflammatory omega-3, omega-7, and omega-9 fatty acids was also significantly increased during the late inflammatory response (12–24 hr after KLA treatment) (Figures 1B and 1C). Based on estimates of cell volume, maximum intracellular concentrations of DHA are on the order of 10  $\mu$ M and EPA and 9Z-POA are on the order of 2  $\mu$ M.

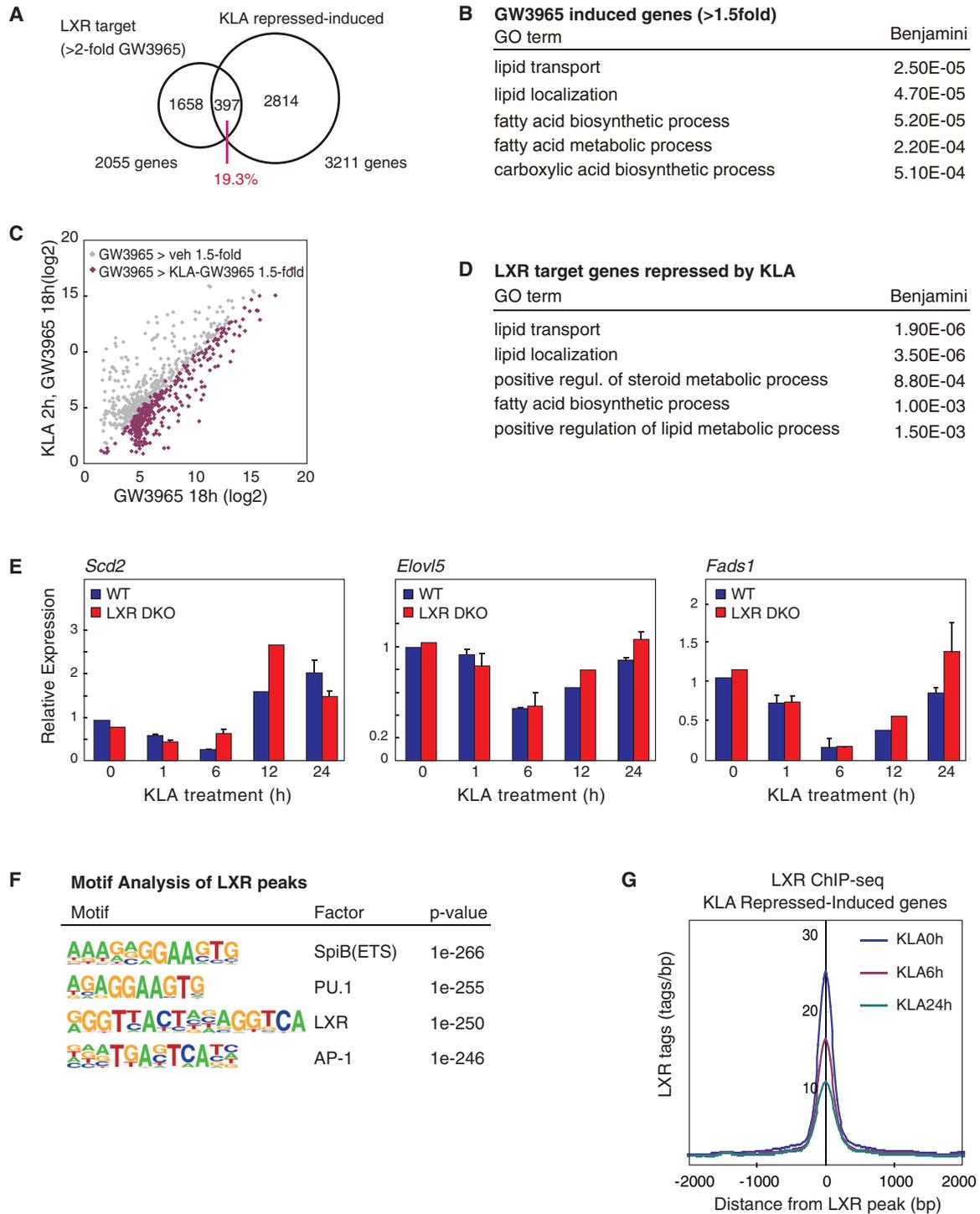
Analysis of microarray data from the same KLA-treated macrophages, generated by the LIPID MAPS Consortium (<http://www.lipidmaps.org/>) (Dennis et al., 2010), demonstrated biphasic expression of genes encoding enzymes involved in mono-unsaturated and omega-3 polyunsaturated fatty acid biosynthesis, exemplified by *Scd2*, *Fads1*, *Acox3*, and *Elovl5*

(F) Functional annotations associated with genes exhibiting KLA repressed-induced temporal expression patterns.

(G) Relative mRNA expression of *Scd2* and *Fads2* in human monocyte-derived macrophages treated with KLA for 0, 1, 6, and 24 hr.

Values are expressed as mean  $\pm$  SEM. \*p < 0.05 and \*\*p < 0.01.

See also Figure S1.



**Figure 2. Genes Required for Anti-inflammatory Fatty Acid Synthesis Demonstrate Biphasic Temporal Expression Patterns following TLR4 Ligation**

(A) Venn diagram of overlap between LXR target genes (GW3965 induced genes >2-fold versus untreated) and KLA repressed-induced genes.  
 (B) Functional annotations associated with LXR target genes induced by GW3965 treatment.  
 (C) Scatterplot depicting the relationship between fold change of LXR target genes (GW3965 >1.5-fold versus untreated) comparing RNA-seq data from thioglycollate-elicited macrophages treated with GW3965 (18 hr), with or without KLA pretreatment (100 ng/mL for 2 hr).  
 (D) Functional annotations associated with LXR target genes repressed by KLA treatment.  
 (E) *Scd2*, *Elovl5*, and *Fads1* mRNA expression in *LXRα/β<sup>-/-</sup>* and WT thioglycollate-elicited macrophages treated with KLA for 0, 1, 6, 12, and 24 hr.

(legend continued on next page)

(Figures 1D and S1, available online). To independently confirm these findings, we performed RNA sequencing (RNA-seq) throughout a time course of KLA treatment. These experiments also revealed a common biphasic expression pattern for many of the genes involved in synthesis of unsaturated fatty acids, exemplified by *Scd1/2*, *Elovl5*, *Fads1*, *Acsl3*, and *Acox3* (Figure 1E). This temporal pattern is characterized by an initial transient reduction within 1–6 hr of TLR4 activation and subsequent activation in the late phase of the TLR4 response (Figures 1D and 1E). The rapid decrease in lipid species observed at 30 min (Figure 1B) precedes the decrease in mRNA levels of biosynthetic genes (Figures 1D and 1E), indicating that the initial phase of reduced fatty acid levels is determined by post-transcriptional mechanisms. However, the increase in levels of mono- and polyunsaturated fatty acids between 12 and 24 hr is correlated with increases in mRNA levels of corresponding biosynthetic genes. Whole transcriptome analysis revealed that among the 22,455 measurable transcripts, 2,993 genes with RefSeq annotations were reduced >2.0-fold at 6 hr and subsequently increased 2.0-fold at 24 hr after KLA treatment. We define this subset of KLA-regulated genes as KLA repressed-induced genes. Consistent with temporal changes in macrophage fatty acid content, the entire set of KLA repressed-induced genes was significantly enriched for functional annotations linked to lipid metabolism (Figure 1F). Experiments in KLA-treated human monocyte-derived macrophages revealed a similar biphasic expression pattern for genes involved in the synthesis of unsaturated fatty acids, exemplified by *Scd1* and *Fads2* (Figure 1G). The temporal pattern characterized by induction in the late phase of the TLR4 response suggests that the observed temporal dynamics of specific fatty acid metabolic reprogramming is conserved in humans. Collectively, these findings indicate that TLR4 signaling induces a biphasic reprogramming of fatty acid metabolism in macrophages through transcriptional and post-transcriptional mechanisms.

### Biphasic Expression of Fatty Acid Biosynthetic Genes Is Independent of LXRs

Many enzymes involved in unsaturated fatty acid synthesis are products of LXR-regulated target genes (Calkin and Tontonoz, 2012; Hong and Tontonoz, 2014). Because TLR4 activation can repress LXR induction of gene expression (Castrillo et al., 2003), it is possible that altered LXR activity could account for the biphasic pattern of expression observed for genes involved in mono- and polyunsaturated fatty acid biosynthesis. To address this possibility, we performed RNA-seq analysis of RNA recovered from macrophages treated with either vehicle or the synthetic LXR agonist GW3965. Approximately one fifth (19.3%) of LXR target genes (GW3965 > vehicle 2-fold) are represented as KLA repressed-induced genes (Figure 2A). In addition, gene ontology analysis revealed that both LXR-induced and KLA repressed-induced genes are enriched for similar functional annotations, including lipid metabolism and fatty acid metabolic process (Figure 2B). To further examine the extent

to which TLR4-mediated inflammation repressed LXR-dependent gene expression, macrophages were pretreated with vehicle or KLA, followed by treatment with either vehicle or the LXR agonist GW3965. RNA-seq revealed that LXR target gene activation was markedly attenuated by KLA pretreatment (~42% GW3965-induced genes; Figure 2C), consistent with previous findings (Joseph et al., 2003). These TLR4-compromised, LXR target genes were significantly enriched for functional annotations linked to lipid transport, lipid localization, and fatty acid biosynthetic process (Figure 2D). These data suggest that the macrophage LXR-regulatory program involved in synthesis of unsaturated fatty acid is repressed in the early phase of TLR4 activation, which could be important for allowing appropriate induction of the inflammatory response.

To assess whether the repression of LXR activity is required for the early reduction of unsaturated fatty acid related gene expression and production following KLA treatment, we took advantage of LXR-deficient macrophages. The temporal dynamics of TLR4 activation was assessed by expression profiling of KLA treated macrophages prepared from WT and *LXR $\alpha/\beta$ <sup>-/-</sup>* mice (Repa et al., 2000b). Unexpectedly, qPCR analysis of the temporal mRNA expression patterns of genes involved in unsaturated fatty acid synthesis, exemplified by *Scd2*, *Elovl5*, and *Fads1*, revealed similar patterns in *LXR $\alpha/\beta$ <sup>-/-</sup>* and WT macrophages (Figure 2E). We further evaluated effects of TLR4 activation on the genome-wide location of endogenous LXRs by chromatin immunoprecipitation (ChIP)-seq. These studies revealed co-localization of LXRs with macrophage lineage-determining factors PU.1 and AP-1 based on motif co-enrichment (Figure 2F), consistent with previous studies using tagged LXRs in RAW264.7 macrophages (Heinz et al., 2010). Further, these studies revealed an unexpected finding that LXR binding at KLA repressed-induced loci significantly decreases in the late phase of inflammation (Figures 2G and S2). Thus, temporal changes in LXR binding are disassociated from late phase induction of KLA repressed-induced genes.

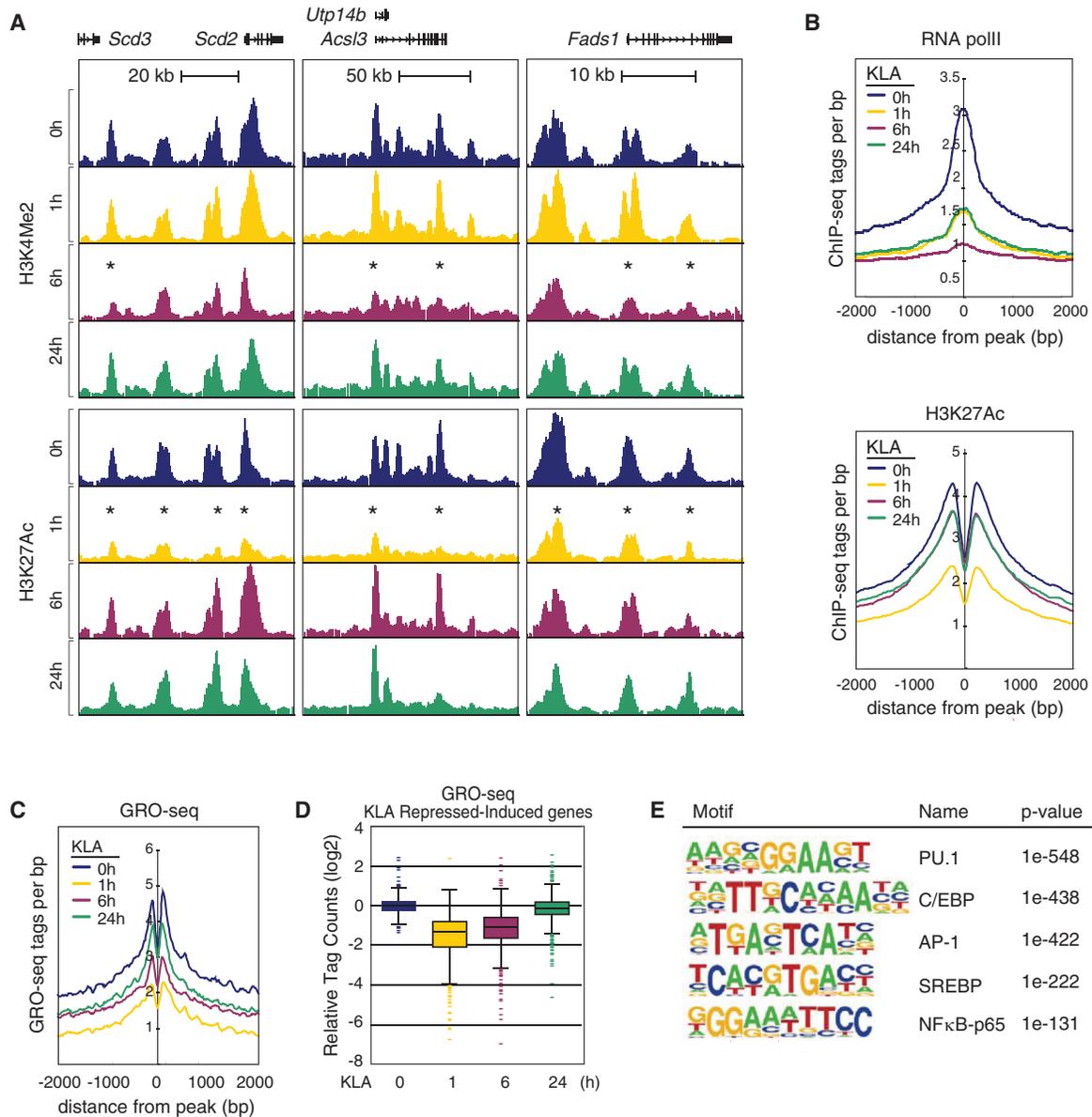
### TLR4 Signaling Reprograms Enhancer Activities Near Repressed-Induced Genes

The unexpected finding that LXR is dispensable for late activation of genes directing fatty acid metabolism prompted us to analyze the local enhancer landscapes of these genes for candidate regulators associated with the temporal profile of repressed-induced genes. To identify enhancers exhibiting temporal activities associated with KLA repressed-induced genes, we performed ChIP-seq to analyze the dimethylation status of lysine 4 of histone H3 (H3K4me2), acetylation status of lysine 27 of histone H3 (H3K27ac), and RNA polymerase II (RNA polII) in naive and KLA-stimulated macrophages. Whereas H3K27ac and RNA polII correlate positively with active transcriptional activity (Creyghton et al., 2010; Kaikkonen et al., 2013), deposition of H3K4me2 has been demonstrated as an indicator of both previous and current local transcription (He et al., 2010; Kaikkonen et al., 2013; Ostuni et al., 2013).

(F) De novo motif analysis of LXR peaks in WT thioglycollate-elicited macrophages.

(G) Normalized distribution LXR ChIP-seq tag density, at enhancers vicinal to KLA repressed-induced genes, in thioglycollate-elicited macrophages treated with KLA for 0, 1, 6, and 24 hr.

See also Figure S2.



**Figure 3. Temporal Dynamics of *cis*-Regulatory Elements Associated with KLA Repressed-Induced Genes**

(A) UCSC genome browser images illustrating normalized tag counts for H3K4Me2 and H3K27Ac at the LXR target genes in BMDMs treated with KLA for 0, 1, 6, and 24 hr.

(B) Distribution of RNA polII and H3K27Ac tag densities in vicinity of KLA repressed-induced enhancers in BMDMs treated with KLA for 0, 1, 6, and 24 hr.

(C) Distribution of GRO-seq tags at KLA repressed-induced enhancers in thioglycollate-elicited macrophages treated with KLA for 0, 1, 6, and 24 hr.

(D) Relative distribution of GRO-seq tags at gene bodies of KLA repressed-induced genes in thioglycollate-elicited macrophages treated with KLA for 0, 1, 6, and 24 hr.

(E) Sequence motifs enriched at enhancers associated with KLA repressed-induced genes.

Consistent with their reduced mRNA levels (Figure 1D), genomic loci of representative KLA repressed-induced genes, exemplified by *Scd2/3*, *Acs13*, and *Fads1*, are associated with decreased H3K27ac and H3K4me2 during the early phase inflammatory response at 1 and 6 hr post-KLA, respectively (Figure 3A, asterisks). Prior studies revealed that the lineage determining transcription factor (LDTF) PU.1 is necessary for establishing macrophage-specific cistromes for signal responsive transcription factors (Heinz et al., 2010). Centering our analysis

on PU.1-bound regions, we analyzed the temporal pattern of relevant features at enhancers associated with repressed-induced genes. Chromatin features of active transcription, defined by RNA polII and H3K27ac, were decreased during the early inflammation phase (at 1 hr post-KLA), then subsequently increased at 6 and 24 hr post-KLA stimulation (Figure 3B), preceding increased levels of nearby mRNA.

As a more direct analysis of active transcription, we analyzed global run-on (GRO)-seq data (Kaikkonen et al., 2013) to

measure nascent transcript levels at KLA repressed-induced loci. Consistent with enhancer ChIP-seq data, GRO-seq revealed that nascent RNA transcription at KLA repressed-induced enhancers follows a similar temporal profile, exhibited by early transcriptional repression and late response induction (Figure 3C). Further, GRO-seq analysis revealed a conserved temporal pattern of transcription at associated KLA repressed-induced gene bodies (Figure 3D). Collectively, these results suggest that KLA repressed-induced enhancer activity and gene transcription is transiently inhibited following inflammatory activation, ensuring decreased unsaturated fatty acid synthesis in macrophages. This is followed by subsequent late phase induction of relevant macrophage transcription and gene expression, culminating in increased unsaturated fatty acid synthesis and output by macrophages.

To define transcription factors potentially determining this temporal regulation of KLA repressed-induced genes, we performed motif analysis on enhancers exhibiting the repressed/induced pattern of chromatin features. As expected, de novo motif analysis identified motifs for the macrophage LDTFs, PU.1, C/EBP, and AP-1, as the most highly enriched sequences. Unexpectedly, an SREBP response element (SRE) was also highly enriched in repressed-induced associated enhancers (Figure 3E). Given its role in regulation of fatty acid metabolism in various cell types, these findings suggested that SREBP1 might be a determinant of late inflammatory phase regulatory dynamics leading to induction of genes necessary for unsaturated fatty acid biosynthesis.

### SREBP1 Activity Is Induced during the Resolution Phase of the Inflammatory Response

We previously demonstrated that LXR and SREBP1 not only co-localize to representative genes involved in maintaining cholesterol and fatty acid homeostasis, but their coordinate regulatory actions can control context-specific expression profiles (Spann et al., 2012). To investigate the potential relationships of LXR and SREBP1 in controlling macrophage lipid metabolism following TLR4 activation, we performed ChIP-seq of SREBP1 and LXR in mouse primary macrophages stimulated with ligands for LXR and TLR4 for 24 hr. As expected, SREBP1 recruitment was observed in the enhancers of lipid synthesis-related genes, as exemplified by *Scd2*, *Acsl3*, and *Srebf1* following GW3965, but not desmosterol treatment, which is a potent suppressor of SREBP processing (Figure 4A). Further, LXR and SREBP1 cis-tromes exhibited significant overlap when comparing genome-wide binding profiles (Figures 4A and S3A). The genes associated with LXR-SREBP1 co-bound sites were enriched for functional annotations for fatty acid metabolism, fatty acid biosynthesis, and elongation (Figure S3B). Further, temporal patterns for direct measurement of enhancer activity levels, demonstrated by H4K5ac and GRO-seq, revealed these LXR-SREBP1 co-bound regions exhibited a coordinate KLA repressed-induced profile (Figure S3C).

Remarkably, KLA treatment also dramatically increased the binding of SREBP1 at enhancer-like regions associated with genes required for mono- and polyunsaturated fatty acid biosynthesis (Figure 4A), consistent with the enrichment of the SREBP recognition element in repressed-induced enhancers (Figure 3E). Furthermore, rigorous peak analysis using HOMER defined

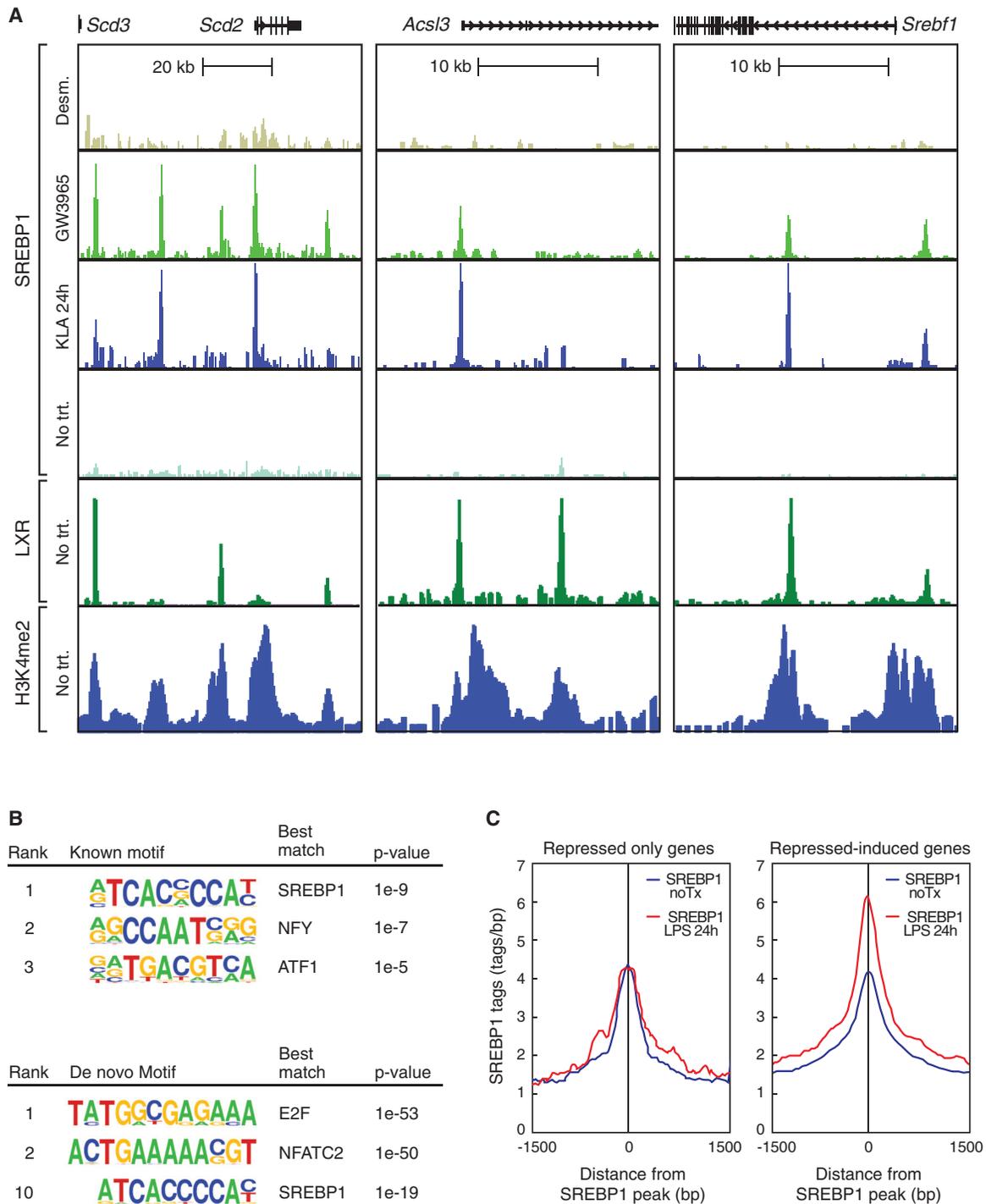
peaks along with irreproducible discovery rate (IDR) analysis identified the top known motif in the KLA-induced SREBP1 cis-trome as matching the consensus sterol response element (Figure 4B, top). The sterol response element was independently identified by de novo motif analysis of IDR-defined SREBP1 binding sites (Figure 4B, bottom). Multiple independent experiments indicated that the late phase KLA induction of SREBP1 binding activity was associated with parallel increased nuclear levels of mature SREBP1 protein, as determined by western blotting (Figure S3D). Intriguingly, the late phase increase in SREBP1 recruitment is specific to KLA repressed-induced associated promoters and enhancers, as binding is not changed at regions of solely KLA-repressed genes (Figure 4C).

### SREBP1 Drives TLR-Responsive Late Activation of Repressed-Induced Genes

The observation that SREBP1 was recruited to the genes involved in unsaturated fatty acid synthesis in the late inflammatory response led us to examine the consequences of *Srebf1* deletion in the inflammatory response of macrophages on a genome-wide scale. We performed RNA-seq analysis of KLA treated bone marrow-derived macrophages (BMDMs) prepared from WT and *Srebf1*<sup>-/-</sup> mice (Shimano et al., 1997) (Figures 5A and 5B). We identified 2,995 significantly expressed transcripts with RefSeq annotations exhibiting the KLA repressed-induced phenotype; characterized by reduced levels of >2-fold at 6 hr and subsequently increased >2-fold at 24 hr after KLA treatment. The expression of 1,047 of these genes (~35%), in the KLA repressed-induced group, were significantly reduced at 24 hr post-KLA treatment in *Srebf1*<sup>-/-</sup> macrophages compared to levels in WT cells (Figures 5A and 5B). Both RNA-seq and qPCR analysis confirmed that *Srebf1*<sup>-/-</sup> macrophages demonstrated significant reduction in the expression of genes mediating mono- and polyunsaturated fatty acid biosynthesis, exemplified by *Scd1/2*, *Acsl3*, *Fads1/2*, and *Acot2*, during the resolution phase of inflammation at 24 hr post-KLA treatment (Figures 5C and S4A).

We independently confirmed the requirement of SREBP1 in the regulation of these genes by using small interfering RNAs (siRNAs) specifically targeting *Srebf1* or *Scap*. qPCR analysis indicated that the *Srebf1* knockdown was sufficient to inhibit the late phase inflammation induction of unsaturated fatty acid related gene expressions, exemplified by *Scd2* (Figure 5D). RNA-seq analysis further confirmed that siRNA-mediated *Srebf1* knockdown led to significantly compromised late phase induction of many genes controlling synthesis of mono- and polyunsaturated fatty acids, including *Acsl3* and *Fads1/2*, in macrophages at 24 hr post-KLA treatment (Figures S4B–S4D). Further, knockdown of *Scap*, which is required for SREBP processing and activation (Horton et al., 2002), resulted in a similar compromise in late phase induction of gene expression (Figure 5E).

To gain further insight into the mechanism by which the repressed-induced gene expression is compromised in *Srebf1*<sup>-/-</sup> macrophages, we performed ChIP-seq of H3K27ac and RNA polII to evaluate local enhancer activity at KLA repressed-induced loci. Both H3K27ac and RNA polII levels were markedly decreased at post-KLA 24 hr, in *Srebf1*<sup>-/-</sup> macrophages, at KLA repressed-induced gene bodies, as



**Figure 4. SREBP1 Is Recruited to Repressed-Induced Enhancers during the Resolution Phase of the Inflammatory Response**

(A) UCSC genome browser images illustrating normalized tag counts for SREBP1, LXR, and H3K4Me2, at indicated loci, in thioglycollate-elicited macrophages treated with vehicle, KLA, GW3965, or desmosterol for 24 hr.

(B) Known and de novo motifs identified in regions bound by SREBP1 in the late inflammatory response. For ChIP-seq peaks used in motif analysis, the peaks for each SREBP1 ChIP were identified using Homer, and we calculated the IDR to measure the consistency between replicate experiments for the strength of binding at each loci and then retained SREBP1 peaks with IDR <0.05.

(C) Distribution of SREBP1 tag densities, at enhancers associated with genes exhibiting either repressed-repressed or repressed-induced temporal expression patterns, in thioglycollate-elicited macrophages treated with KLA for 24 hr.

See also [Figure S3](#).

exemplified by *Scd2* (Figure 5F). Further, the normal temporal dynamics, characterized by the late phase increase in both H3K27ac and RNA polII levels, were globally compromised in *Srebf1*<sup>-/-</sup> macrophages when looking at profiles for all KLA repressed-induced associated enhancers (Figure 5G), consistent with both the deficient late phase expression recovery, as measured by RNA-seq, and increased late recruitment of SREBP1 to these enhancers. These results suggest a significant role for SREBP1 in the late phase induction of repressed-induced genes that control fatty acid biosynthesis.

To investigate whether signaling through other TLRs would exert similar effects on SREBP1 target genes, we assessed relevant gene expression levels throughout a time course of PAM3CSK4 (TLR2 agonist) and Poly(I:C) (TLR3 agonist) treatment, comparing temporal responses in *Srebf1*<sup>-/-</sup> and WT macrophages. Similar to the temporal dynamics of TLR4 activation, TLR2 and TLR3 responses of unsaturated fatty acid biosynthetic genes demonstrated a repressed-induced expression profile (Figures S4E and S4F) dependent on SREBP1. However, the induced phase was much more pronounced in the case of TLR2 activation. This result suggests a predominant role of the MyD88 pathway, which is used by both TLR4 and TLR2, but not TLR3, which instead signals primarily through TRIF.

### SREBP1 Is Necessary for Resolution of the TLR-Mediated Inflammation

Previous studies demonstrated that unsaturated fatty acids such as EPA, DHA, and 9Z-POA have potent anti-inflammatory effects in macrophages by antagonizing inflammatory signaling through GPCRs, nuclear receptors, and other mechanisms (Cao et al., 2008; Li et al., 2013; Oh et al., 2010). To investigate whether the late phase of expression of genes involved in mono- and polyunsaturated fatty acid biosynthesis contributes to the resolution phase of TLR4 signaling, we evaluated the temporal expression profiles of genes that are induced following KLA treatment in WT and *Srebf1*<sup>-/-</sup> macrophages. Indeed, RNA-seq analysis revealed that *Srebf1*<sup>-/-</sup> macrophages demonstrated delayed resolution and often exaggerated gene expression upon TLR4 activation, relative to their WT counterparts (Figures 6A and 6B). In WT macrophages, 964 significantly expressed transcripts (with RefSeq annotations) were detected that were increased >2.0-fold at 6 hr after KLA treatment and subsequently decreased >2.0-fold at 24 hr post-KLA treatment (defined herein as KLA induced-repressed genes). There were 247 of these induced-repressed genes that demonstrated significantly increased expression, at 24 hr post-KLA treatment, in *Srebf1*<sup>-/-</sup> macrophages compared to WT (Figures 6A and 6B). This set of KLA induced-repressed genes had significant enrichment of functional annotations for immune response, regulation of cytokine production, and inflammatory response (Figure 6C). qPCR analysis confirmed that inflammatory gene expressions, as exemplified by *Nos2*, *Cxcl1*, *Cxcl9*, and *Il1a*, are significantly increased at 24 hr post-KLA treatment in *Srebf1*<sup>-/-</sup> macrophage (Figure 6D). Further, siRNA-mediated knockdown experiments confirmed the requirement of SREBP1 for appropriate resolution of inflammatory gene expression, as exemplified by *Cxcl2*, *Nos2*, *Cxcl1*, *Il1a*, *Il12b*, and *Il6* (Figure S5A).

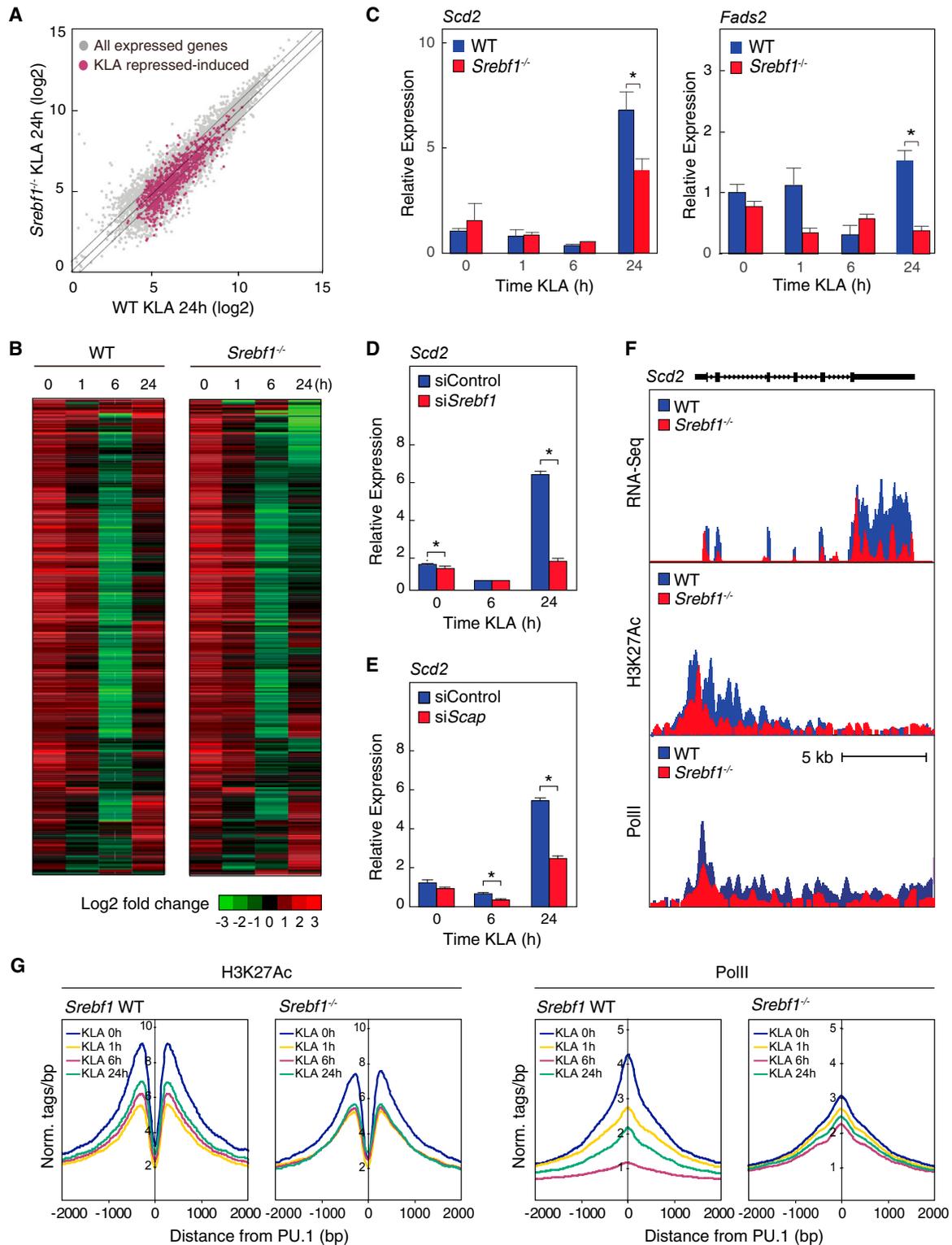
We further assessed relevant pro-inflammatory gene expression levels throughout a time course of PAM3CSK4 (TLR2

agonist) and Poly(I:C) (TLR3 agonist) treatment, comparing temporal responses in *Srebf1*<sup>-/-</sup> and WT macrophages. Similar to the temporal dynamics of TLR4 activation, TLR2 and TLR3 responses of pro-inflammatory genes demonstrated an induced-repressed expression profile (Figures S5E and S5F). Further, resolution of TLR2- and TLR3-mediated inflammatory gene expression was drastically compromised in *Srebf1*<sup>-/-</sup> macrophages, relative to their WT counterparts (Figures S5E and S5F). Interestingly, Pam3 induced genes showed delayed resolution, similar to TLR4 response, whereas Poly(I:C) induced genes were hyper-responsive throughout the time course. These results are consistent with a MyD88-dependent induction of SREBP1 mediating late resolution of TLR2 and TLR4 responses. The hyper-activation of *Ifnb1* and *Ifna4* in response to TLR3 agonist in *Srebf1*<sup>-/-</sup> macrophages may reflect a different mechanism.

To further define the SREBP-dependent temporal regulatory pattern, we analyzed ChIP-seq data for RNA polII in KLA-treated *Srebf1*<sup>-/-</sup> and WT macrophages. Normalized tag density plots at induced-repressed genes revealed increased levels of RNA polII in *Srebf1*<sup>-/-</sup> versus WT macrophages (Figure 6E). The average tag density levels, between *Srebf1*<sup>-/-</sup> and WT, demonstrated the most significant differentials at 24 hr post-KLA (Figure 6E). These distinct patterns are exemplified for *Cxcl2*, *Nos2*, *Cxcl1*, and *Il1a* in Figure 6F. Consistent with ChIP-seq and mRNA expression data, GRO-seq analysis revealed a similarly conserved temporal pattern of transcription at associated KLA induced-repressed gene bodies (Figure S5B). These results suggest that the temporal dynamics of induced-repressed inflammatory genes are regulated via local enhancer activities driven by KLA responsive transcription factor complexes.

Given the role of NFκB as a primary driver of TLR4-mediated responses, we further performed ChIP-seq of the p65 component of NFκB to determine whether the increased inflammatory gene expression, exhibited by *Srebf1*<sup>-/-</sup> macrophages, was due to increased p65 recruitment to the induced-repressed loci. Unexpectedly, the ChIP-seq analysis revealed a strikingly similar pattern of p65 binding in KLA treated *Srebf1*<sup>-/-</sup> and WT macrophages (Figures 6F, S5C, and S5D). Further, the similarity of p65 binding, comparing *Srebf1*<sup>-/-</sup> and WT profiles, remains consistent whether looking at all repressed-induced loci (correlation co-efficient = 0.949817) or the subset of induced-repressed loci demonstrating the most significant alterations upon loss of *Srebp1* (correlation co-efficient = 0.9438753) (Figures S5C and S5D). This finding is consistent with previous studies, suggesting that the repressive actions of unsaturated fatty acids on NFκB activity are independent of changes in factor binding (Li et al., 2013).

The gene expression pattern observed in *Srebf1*<sup>-/-</sup> macrophages predicts that the late phase of mono- and polyunsaturated fatty acid production would be compromised in these cells. We therefore performed lipidomic analysis of KLA-treated *Srebf1*<sup>-/-</sup> and WT macrophages to assess changes in fatty acid levels. Consistent with the altered gene expression patterns, *Srebf1*<sup>-/-</sup> macrophages demonstrated marked decreases in unsaturated fatty acid production, as exemplified by DHA, EPA, and 9Z-POA; with most dramatic differentials, between *Srebf1*<sup>-/-</sup> and WT macrophages, occurring at 24 hr post-KLA



**Figure 5. *Srebf1*<sup>-/-</sup> Macrophages Exhibit Reduced Fatty Acid Biosynthetic Gene Expression during the Resolution Phase of the TLR4 Response**

(A) Scatterplot depicting the relationship between fold change of KLA repressed-induced genes, comparing RNA-seq from WT versus *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for 24 hr. The gray dots show all expressed genes. The red dots represent all KLA repressed-induced genes.

(B) Hierarchical clustering and heatmap of the fold change in expression levels of KLA repressed-induced genes in WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for the indicated times (FDR <0.01 and RPKM >0.5).

(legend continued on next page)

treatment (Figures 7A and S6A). This late phase-specific decrease is consistent with the possibility that these anti-inflammatory fatty acids contribute to the resolution phase of the TLR4 response.

To further test the link between late phase resolution of inflammation and unsaturated fatty acid output, fatty acid rescue experiments were performed in which exogenous unsaturated fatty acids were added to KLA-treated *Srebf1*<sup>-/-</sup> and WT macrophages. In these experiments, mono- (9Z-POA) and polyunsaturated (EPA/DHA) fatty acids were supplemented, either alone or simultaneously, at 12 hr post-KLA treatment to mimic late phase accumulation. Cells were then harvested at 24 hr post-KLA treatment. Addition of exogenous unsaturated fatty acids led to significant reduction of inflammatory gene expression in macrophages (Figure 7B). Further, this unsaturated fatty acid-specific repressive effect was more pronounced in *Srebf1*<sup>-/-</sup> cells, relative to their WT counterparts (Figure 7B), consistent with intact production of these fatty acid species in WT macrophages. Similar results were observed in siRNA-mediated knockdown cells (Figure S6B).

To investigate the role of SREBP1-mediated unsaturated fatty acid output in modulating the inflammatory response in vivo, *Srebf1*<sup>-/-</sup> mice were challenged with a sublethal dose of LPS. Consistent with the increased late phase inflammatory gene expression patterns observed in the *Srebf1*<sup>-/-</sup> macrophages (Figures 6B–6F), circulating cytokine levels of IL-6 and IL-1 $\alpha$  remained significantly higher in *Srebf1*<sup>-/-</sup> mice at 24 hr post-LPS injection (Figure 7C); thus suggesting a compromised resolution of inflammation relative to their WT counterparts. In addition, supplementation of exogenous EPA, prior to the LPS challenge, protected *Srebf1*<sup>-/-</sup> mice from an exaggerated inflammatory response, restoring circulating cytokine levels to those seen in WT mice (Figure 7C).

## DISCUSSION

Emerging evidence suggests that the immune system and lipid metabolism are coordinately regulated at multiple levels within the body. Here, we demonstrate a reciprocal relationship between cellular levels of anti-inflammatory fatty acids and the temporal induction and resolution of pro-inflammatory gene expression following TLR4 activation (Figure 7D). Anti-inflammatory fatty acid levels rapidly fall following KLA treatment, in advance of downregulation of mRNAs encoding corresponding biosynthetic enzymes. Given the ability of these fatty acid species to suppress NF $\kappa$ B-dependent gene expression, their downregulation is likely to be necessary for a full TLR4 response. At 12–24 hr following TLR4 ligation, anti-inflammatory fatty acid levels rise, concurrent with increased expression of mRNAs encoding biosynthetic enzymes and

decreased expression of mRNAs encoding pro-inflammatory mediators.

The mechanisms responsible for downregulation of lipid biosynthetic genes remain to be established. Repression of basal LXR-dependent gene expression does not account for this effect because a similar pattern of gene expression was observed in LXR double knockout macrophages. Thus, while TLR signaling blunts the ability of LXR agonists to induce target gene expression, alternative mechanisms must account for the observed downregulation. During the initial phase of the TLR4 response, the p65 component of NF $\kappa$ B is recruited to many of the enhancer elements associated with the set of repressed-induced genes. This TLR-induced p65 binding is associated with loss of both co-activator recruitment and active chromatin features at these loci, correlating with their loss of transcriptional activity and expression. Our unpublished results using an NF $\kappa$ B inhibitor, suggest a requirement for NF $\kappa$ B activity in mediating the early phase repression of these genes (data not shown). However, a direct role of p65-containing NF $\kappa$ B complexes in downregulation of these genes remains to be established.

Unexpectedly, the late upregulation of mRNAs encoding enzymes required for synthesis of anti-inflammatory fatty acids was independent of LXRs and instead required SREBP1. Consistent with these findings, ChIP-seq experiments indicated a reduction of LXR binding to enhancers associated with repressed-induced genes, but a marked increase in the binding of SREBP1. This KLA-induced binding of SREBP1 to *cis* regulatory elements at late time points was associated with increases in the total nuclear content of processed SREBP1. Further, the late increase in SREBP1 binding was associated with increases in chromatin features associated with active enhancers. SREBP1 KO macrophages, or macrophages in which siRNAs were used to knock down SREBP1, displayed compromised late induction of repressed-induced genes and reduced production of anti-inflammatory fatty acids. While our data clearly provide evidence for a novel role of SREBP1 in transcriptionally tailoring specific macrophage lipid metabolic output, driving late phase synthesis of anti-inflammatory unsaturated fatty acids, the mechanisms controlling both the late phase induction of SREBP1 recruitment and the SREBP1-target activation specificity are not entirely clear. An understanding of these mechanisms could be important in identifying novel targets for development of SREBP1-centric interventions of various inflammatory disease states.

Recent studies provided evidence that SREBP-1a is required for the formation of the inflammasome and secretion of IL-1 $\beta$  in response to systemic inflammation (such as endotoxic shock) (Im et al., 2011). Consistent with these findings, we observed increased secretion of IL-1 $\beta$  protein following KLA treatment of

(C) Relative mRNA expression of *Scd2* and *Fads2* in WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for the indicated times.

(D) Relative mRNA expression of *Scd2* mRNA KLA-treated thioglycollate-elicited macrophages, transfected with siRNA control or targeting *Srebf1*.

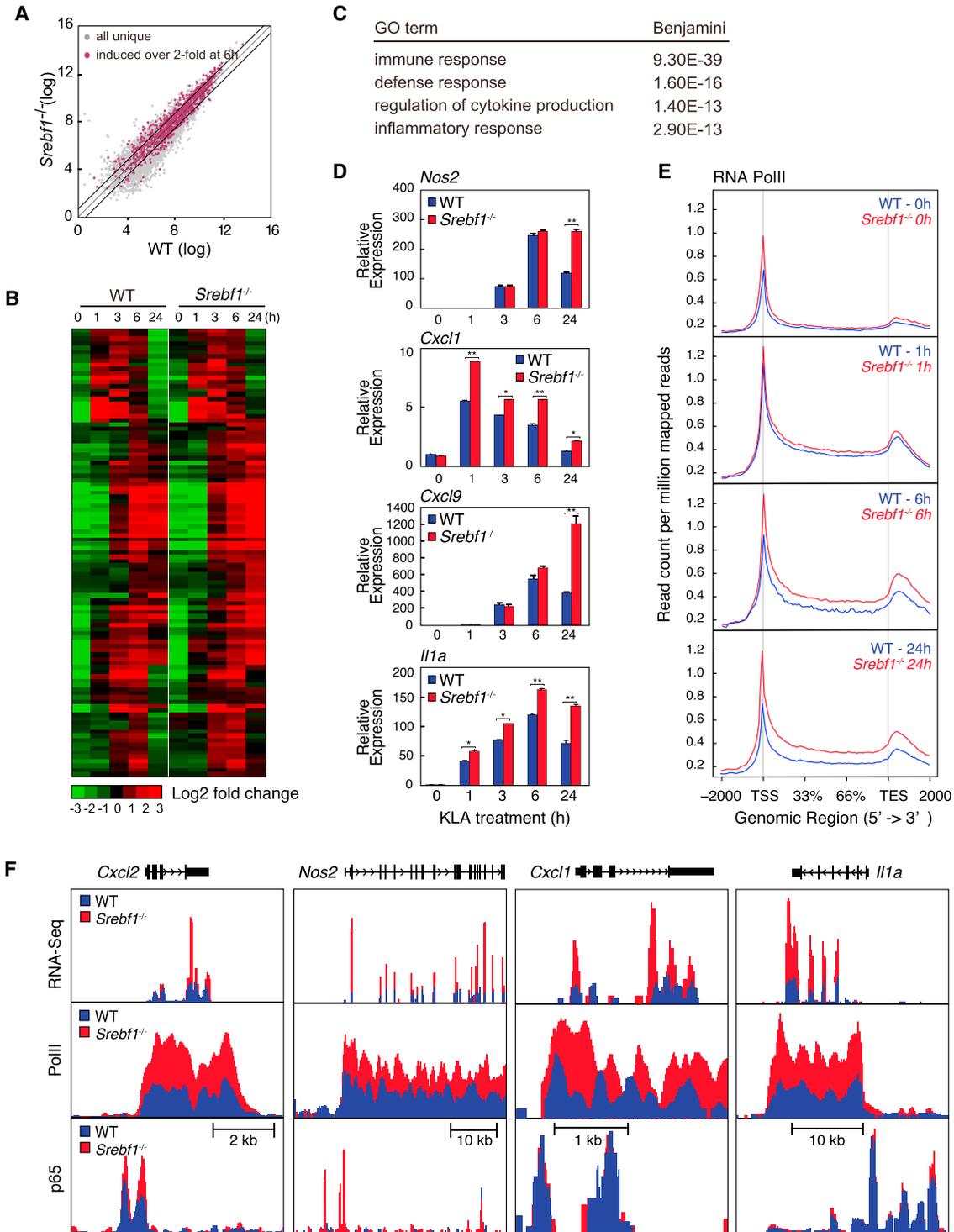
(E) Relative mRNA expression of *Scd2* mRNA KLA-treated thioglycollate-elicited macrophages, transfected with siRNA control or targeting *Scap*.

(F) Distribution of RNA-seq, H3K27ac, and RNA polII tag densities at the *Scd2* locus in WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for 24 hr.

(G) Distribution of H3K27Ac and RNA polII tag densities in the vicinity of enhancers associated with KLA repressed-induced genes in WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for the indicated times.

Values are expressed as mean  $\pm$  SEM. \* $p$  < 0.05 and \*\* $p$  < 0.01.

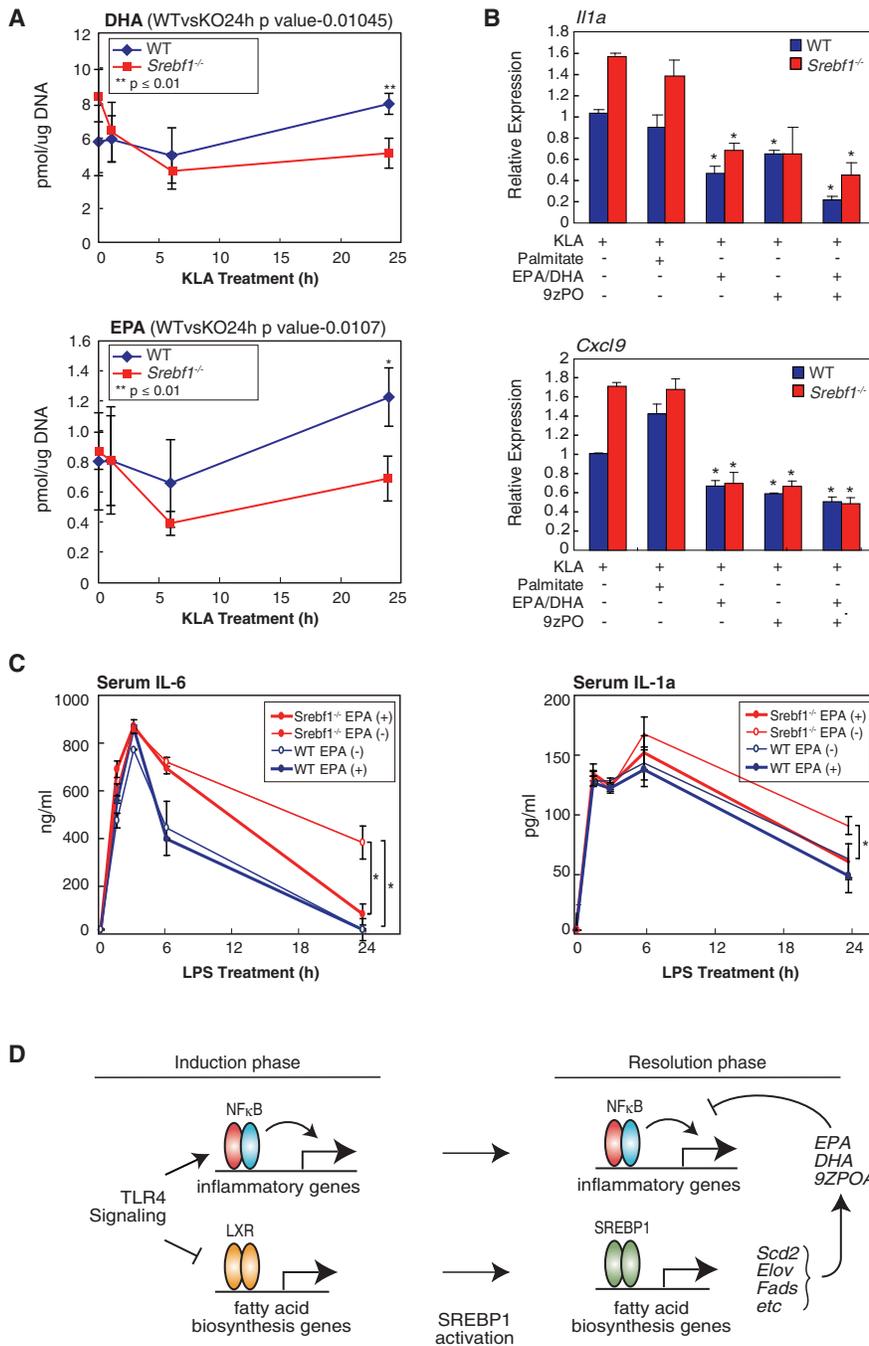
See also Figure S4.



**Figure 6. *Srebf1*<sup>-/-</sup> Macrophages Exhibit a Hyper-inflammatory Phenotype**

(A) Scatterplot depicting the relationship between fold change of KLA induced-repressed genes, comparing RNA-seq from WT versus *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for 24 hr. The gray dots show all uniquely expressed genes. The red dots represent all KLA induced-repressed genes.  
 (B) Hierarchical clustering and heatmap of the fold change in expression levels of KLA induced-repressed genes, comparing RNA-seq data from WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for 24 hr (FDR <0.01 and RPKM >0.5).  
 (C) Functional annotations associated with KLA induced-repressed genes.  
 (D) Relative mRNA expression of inflammatory genes in WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for the indicated times.  
 (E) Distribution of RNA polII tag densities at loci of KLA induced-repressed genes WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for indicated times.

(legend continued on next page)



**Figure 7. SREBP1 Is Necessary for Resolution of Inflammation by Driving Appropriate Macrophage Production of Anti-inflammatory Unsaturated Fatty Acids in Late Inflammatory Response**

(A) Lipidomics analysis of unsaturated fatty acid (EPA and DHA 9Z-PO) levels in KLA treated WT and *Srebf1*<sup>-/-</sup> BMDMs.

(B) Relative mRNA expression of inflammatory genes in WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for 24 hr, with or without supplementation with the indicated exogenous fatty acids (20 μM) at 12 hr post-KLA treatment.

(C) Serum levels of cytokines IL-6 and IL-1a, as quantified by ELISA, in WT and *Srebf1*<sup>-/-</sup> mice treated with 5 mg/kg LPS for 0, 1, 3, 6, and 24 hr, with or without EPA supplementation as indicated.

(D) Model for integrated actions of NFkB, LXRs, and SREBP1 during the induction and resolution phases of the TLR4 response.

Values are expressed as mean ± SEM. \*p < 0.05 and \*\*p < 0.01.

See also Figure S6.

TLR3; thus indicating that SREBP1 is genetically required for the normal resolution phase of varied TLR responses in macrophages. Our findings further suggest that SREBP1-driven synthesis of anti-inflammatory fatty acids contributes to this resolution phase. The late TLR4-mediated increase of these fatty acid species is compromised in *Srebf1*<sup>-/-</sup> macrophages, and supplementation of exogenous anti-inflammatory, both in cultured macrophages and in vivo, reverses hyper-induction of pro-inflammatory gene expression caused by loss of SREBP1. This is consistent with the presence of higher concentrations of these species in WT macrophages. Although our studies focused on 9Z-POA, DHA, and EPA, it is possible that additional anti-inflammatory metabolites of polyunsaturated fatty acids, such as resolvins protectins and fatty acid hydroxyl fatty acids, are also generated by the late SREBP1-dependent program of gene expression. Interestingly, loss of SREBP1

*Srebf1*<sup>-/-</sup> macrophages as compared to WT controls (data not shown). However, our studies also demonstrated that a subset of TLR4-responsive, pro-inflammatory genes was hyper-activated in *Srebf1*<sup>-/-</sup> macrophages at 12–24 hr following KLA treatment. Similar late hyper-inflammatory trends were observed in *Srebf1*<sup>-/-</sup> macrophages stimulated with ligands for TLR2 and

results in increased recruitment of RNA polII to a subset of inflammatory response genes independent of changes in p65 binding activity. These results are consistent with prior studies suggesting that DHA, EPA, and 9Z-POA uncouple NFkB binding from its transcriptional output (Li et al., 2013). Because p65 binding itself is unchanged, the mechanism of inhibition is unlikely to

(F) UCSC genome browser image illustrating normalized tag counts for RNA-seq, RNA polII, and p65 ChIP-seq at loci of inflammatory genes in WT and *Srebf1*<sup>-/-</sup> BMDMs treated with KLA for the indicated times.

Values are expressed as mean ± SEM. \*p < 0.05 and \*\*p < 0.01.

See also Figure S5.

be through alterations in the I $\kappa$ B kinase cascade required for NF $\kappa$ B activation.

In concert, our findings provide evidence for a role of SREBP1 in promoting resolution of the transcriptional response of macrophages to TLR signaling by driving the synthesis of anti-inflammatory fatty acids. While we have shown that the SREBP1 pathway also influences resolution following activation of TLR2 and TLR3, the extent to which it is involved in resolution of responses to other pattern recognition receptors or cytokine-dependent inflammatory responses remains to be determined. It will therefore be of interest to investigate this pathway further with respect to control of the resolution phase of inflammation in response to infection and injury, as well as in disease contexts in which inflammation plays a pathogenic role.

## EXPERIMENTAL PROCEDURES

### Cell Culture

*LXR $\alpha$  $\beta^{-/-}$*  and *Srebf1 $^{-/-}$*  were generated as described previously (Repa et al., 2000b; Shimano et al., 1997). These mice were backcrossed to the C57BL/6J strain for more than ten generations. Mouse thioglycollate-elicited macrophages were isolated from male 6- to 9-week-old C57BL/6J (Charles River laboratories), *LXR $\alpha$  $\beta^{-/-}$* , and *Srebf1 $^{-/-}$*  mice and cultured as previously described (Spann et al., 2012). Peritoneal macrophages were harvested by lavage 3 days after intraperitoneal injection of 3 mL of 3% thioglycollate medium (<http://www.lipidmaps.org/protocols/>), overnight culture, and adherence selection. Bone marrow from mice were isolated by perfusion of the medullary cavity of femurs, tibias, and iliac bones and cultured in medium containing RPMI-1640, 10% FCS, and 20  $\mu$ g/mL M-CSF (R&D) for 6 days. RAW264.7 cells are maintained in the RPMI-1640 media supplemented with 10% FCS (Hyclone) and used between passage 5–10. For the fatty acid rescue experiments, cells were treated with fatty acids complexed with FA-free low-endotoxin BSA (Sigma, final FA:BSA molar ratio was 5:1).

### Animal Study

All mice used in this study have C57BL/6 background. Male, 8- to 11-week-old *Srebf1 $^{-/-}$*  mice and age-matched littermate control were individually housed in cages in a 12 hr/12 hr light/dark cycle with free access to food and water. For supplemental EPA administration study, mice were fed with fish meal-free diet (fish meal-free F1: 4.4% fat; Funabashi Farm) or fish meal-free diet supplemented with 5% EPA ethyl ester (v/v) for 7 days before single intraperitoneal injection of 5 mg/kg LPS ( $n = 5$ , each group). All animal procedures were in accordance with research guidelines for care and use of laboratory animals of Tokyo Medical and Dental University. Temporal changes of serum IL-6 and IL-1 $\alpha$  were quantified by ELISA (R&D).

### ChIP-Seq

ChIP from thioglycollate-elicited peritoneal macrophages or BMDMs was performed as described previously (Spann et al., 2012), with modifications as described in Supplemental Experimental Procedures. ChIP-seq libraries were prepared from ChIP DNA by blunting, A-tailing, adaptor ligation as previously described (Heinz et al., 2010) using barcoded adapters (NEXTflex, Bio Scientific). Libraries were PCR amplified for 12–15 cycles, size selected by gel extraction, and sequenced on either a Illumina Genome Analyzer II or HiSeq 2000 for 51 cycles.

### RNA-Seq

Total RNA was isolated from cells and purified using RNeasy columns and RNase-free DNase digestion according to the manufacturer's instructions (QIAGEN). RNA-seq libraries were prepared from poly(A)-enriched mRNA, either as previously described (Kaikkonen et al., 2013) or as detailed in Supplemental Experimental Procedures.

### High-Throughput Sequencing and Data Analysis

All sequencing was conducted using either Illumina Genome Analyzer II or HiSeq 2000 sequencers using single-end 50 bp reads. All data were aligned

to the mm9 assembly of the mouse genome, and all subsequent data analysis was performed using HOMER, and detailed instructions for analysis can be found at <http://homer.salk.edu/homer/> (Heinz et al., 2010). Each sequencing experiment was normalized to a total of  $10^7$  uniquely mapped tags by adjusting the number of tags at each position in the genome to the correct fractional amount given the total tags mapped. Sequence experiments were visualized by preparing custom tracks for the UCSC genome browser. Differentially expressed genes were identified using HOMER as described previously (Li et al., 2013). For SREBP1 ChIP-seq analysis, ChIP-seq peaks for each SREBP ChIP replicate were identified using Homer, and then the strength of binding at each loci was quantified as the position adjusted reads from the start of the peak region (Homer peak score). We calculated the IDR to measure the consistency between replicate experiments for the strength of binding at each loci and retained SREBP peaks with IDR < 0.05. For various ontology analyses, either HOMER or DAVID Bioinformatics Resources 6.7 was used. The accession number for the data from previously published GRO-seq and ChIP-seq experiments is GEO: GSE48759.

### Statistical Analyses

Statistical analyses were performed using Graph Pad Prism 5 software. The images were prepared using Adobe Illustrator CS5 or Photoshop CS5.1. Data are presented as the mean  $\pm$  SEM. For experiments involving two factors, data were analyzed by two-way ANOVA followed by Bonferroni post-tests. Individual pairwise comparisons were performed using Student's  $t$  test.  $p < 0.05$  was considered significant.

### ACCESSION NUMBERS

The accession number for the sequencing data reported in this paper is GEO: GSE79423.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cmet.2016.11.009>.

### AUTHOR CONTRIBUTIONS

Y.O., N.J.S., and C.K.G. conceived the project and designed experiments. Y.O., N.J.S., V.M.L., and J.T. analyzed data. Y.O., N.J.S., T.S., C.E., E.D.M., T.M., S.H., M.U.K., A.F.C., M.J.K., and M.T.L. performed experiments. H.S. provided mice. M.J.K., I.M., and A.S. provided technical expertise. Y.O., N.J.S., and C.K.G. interpreted data and wrote the manuscript.

### ACKNOWLEDGMENTS

We thank Dr. Timothy Osborne for making anti-SREBP1 antibodies available for these studies. We thank Dr. Kristen Jepsen, Mahdiah Khosroheidari, and Hiroko Matsui for assistance with Illumina sequencing; Jana Collier for technical assistance; Miguel D. Mooney for technical assistance; and Leslie Van Ael for assistance with manuscript preparation. Y.O. was supported by MEXT KAKENHI grant numbers 24641142 and 25H10 and grants from Naito Foundation, Japan Heart Foundation, Japan Foundation for Applied Enzymology, Ono Medical Foundation, and Takeda Science Foundation. M.U.K. was supported by a Fondation Leducq Career Development Award, Academy of Finland Research Fellowship, and grants from Sigrid Jusélius Foundation, ASLA-Fulbright, the Finnish Foundation for Cardiovascular Research, the Finnish Cultural Foundation, and Jane and Aatos Erkkö Foundation. T.S. was supported by funding from the Swedish Society for Medical Research. Studies were primarily supported by NIH grants DK091183, CA17390, HL088083, and DK063491.

Received: March 24, 2016

Revised: October 6, 2016

Accepted: November 25, 2016

Published: December 29, 2016

## REFERENCES

- A-Gonzalez, N., Bensinger, S.J., Hong, C., Beceiro, S., Bradley, M.N., Zelcer, N., Deniz, J., Ramirez, C., Díaz, M., Gallardo, G., et al. (2009). Apoptotic cells promote their own clearance and immune tolerance through activation of the nuclear receptor LXR. *Immunity* *31*, 245–258.
- Calkin, A.C., and Tontonoz, P. (2012). Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. *Nat. Rev. Mol. Cell Biol.* *13*, 213–224.
- Cao, H., Gerhold, K., Mayers, J.R., Wiest, M.M., Watkins, S.M., and Hotamisligil, G.S. (2008). Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. *Cell* *134*, 933–944.
- Castrillo, A., Joseph, S.B., Vaidya, S.A., Haberland, M., Fogelman, A.M., Cheng, G., and Tontonoz, P. (2003). Crosstalk between LXR and toll-like receptor signaling mediates bacterial and viral antagonism of cholesterol metabolism. *Mol. Cell* *12*, 805–816.
- Cildir, G., Akıncılar, S.C., and Tergaonkar, V. (2013). Chronic adipose tissue inflammation: all immune cells on the stage. *Trends Mol. Med.* *19*, 487–500.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* *107*, 21931–21936.
- Dennis, E.A., Deems, R.A., Harkewicz, R., Quehenberger, O., Brown, H.A., Milne, S.B., Myers, D.S., Glass, C.K., Hardiman, G., Reichart, D., et al. (2010). A mouse macrophage lipidome. *J. Biol. Chem.* *285*, 39976–39985.
- Feingold, K.R., Shigenaga, J.K., Kazemi, M.R., McDonald, C.M., Patzek, S.M., Cross, A.S., Moser, A., and Grunfeld, C. (2012). Mechanisms of triglyceride accumulation in activated macrophages. *J. Leukoc. Biol.* *92*, 829–839.
- Ghisletti, S., Huang, W., Jepsen, K., Benner, C., Hardiman, G., Rosenfeld, M.G., and Glass, C.K. (2009). Cooperative NCoR/SMRT interactions establish a corepressor-based strategy for integration of inflammatory and anti-inflammatory signaling pathways. *Genes Dev.* *23*, 681–693.
- Glass, C.K., and Natoli, G. (2016). Molecular control of activation and priming in macrophages. *Nat. Immunol.* *17*, 26–33.
- Goldstein, J.L., DeBose-Boyd, R.A., and Brown, M.S. (2006). Protein sensors for membrane sterols. *Cell* *124*, 35–46.
- He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., et al. (2010). Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* *42*, 343–347.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* *38*, 576–589.
- Hong, C., and Tontonoz, P. (2014). Liver X receptors in lipid metabolism: opportunities for drug discovery. *Nat. Rev. Drug Discov.* *13*, 433–444.
- Horton, J.D., Goldstein, J.L., and Brown, M.S. (2002). SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.* *109*, 1125–1131.
- Huang, S.C., Everts, B., Ivanova, Y., O'Sullivan, D., Nascimento, M., Smith, A.M., Beatty, W., Love-Gregory, L., Lam, W.Y., O'Neill, C.M., et al. (2014a). Cell-intrinsic lysosomal lipolysis is essential for alternative activation of macrophages. *Nat. Immunol.* *15*, 846–855.
- Huang, Y.L., Morales-Rosado, J., Ray, J., Myers, T.G., Kho, T., Lu, M., and Munford, R.S. (2014b). Toll-like receptor agonists promote prolonged triglyceride storage in macrophages. *J. Biol. Chem.* *289*, 3001–3012.
- Im, S.S., Yousef, L., Blaschitz, C., Liu, J.Z., Edwards, R.A., Young, S.G., Raffatellu, M., and Osborne, T.F. (2011). Linking lipid metabolism to the innate immune response in macrophages through sterol regulatory element binding protein-1a. *Cell Metab.* *13*, 540–549.
- Ito, A., Hong, C., Rong, X., Zhu, X., Tarling, E.J., Hedde, P.N., Gratton, E., Parks, J., and Tontonoz, P. (2015). LXRs link metabolism to inflammation through Abca1-dependent regulation of membrane composition and TLR signaling. *eLife* *4*, e08009.
- Iyer, S.S., Ghaffari, A.A., and Cheng, G. (2010). Lipopolysaccharide-mediated IL-10 transcriptional regulation requires sequential induction of type I IFNs and IL-27 in macrophages. *J. Immunol.* *185*, 6599–6607.
- Joseph, S.B., Castrillo, A., Laffitte, B.A., Mangelsdorf, D.J., and Tontonoz, P. (2003). Reciprocal regulation of inflammation and lipid metabolism by liver X receptors. *Nat. Med.* *9*, 213–219.
- 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., and Glass, C.K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* *51*, 310–325.
- Li, P., Spann, N.J., Kaikkonen, M.U., Lu, M., Oh, D.Y., Fox, J.N., Bandyopadhyay, G., Talukdar, S., Xu, J., Lagakos, W.S., et al. (2013). NCoR repression of LXRs restricts macrophage biosynthesis of insulin-sensitizing omega 3 fatty acids. *Cell* *155*, 200–214.
- Lumeng, C.N., DelProposto, J.B., Westcott, D.J., and Saltiel, A.R. (2008). Phenotypic switching of adipose tissue macrophages with obesity is generated by spatiotemporal differences in macrophage subtypes. *Diabetes* *57*, 3239–3246.
- Mantovani, A., Biswas, S.K., Galdiero, M.R., Sica, A., and Locati, M. (2013). Macrophage plasticity and polarization in tissue repair and remodelling. *J. Pathol.* *229*, 176–185.
- Medzhitov, R., and Horng, T. (2009). Transcriptional control of the inflammatory response. *Nat. Rev. Immunol.* *9*, 692–703.
- Oh, D.Y., Talukdar, S., Bae, E.J., Imamura, T., Morinaga, H., Fan, W., Li, P., Lu, W.J., Watkins, S.M., and Olefsky, J.M. (2010). GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell* *142*, 687–698.
- Osborn, O., and Olefsky, J.M. (2012). The cellular and signaling networks linking the immune system and metabolism in disease. *Nat. Med.* *18*, 363–374.
- 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.
- Raetz, C.R., Garrett, T.A., Reynolds, C.M., Shaw, W.A., Moore, J.D., Smith, D.C., Jr., Ribeiro, A.A., Murphy, R.C., Ulevitch, R.J., Fearn, C., et al. (2006). Kdo2-Lipid A of *Escherichia coli*, a defined endotoxin that activates macrophages via TLR-4. *J. Lipid Res.* *47*, 1097–1111.
- Reboldi, A., Dang, E.V., McDonald, J.G., Liang, G., Russell, D.W., and Cyster, J.G. (2014). Inflammation. 25-Hydroxycholesterol suppresses interleukin-1-driven inflammation downstream of type I interferon. *Science* *345*, 679–684.
- Repa, J.J., Liang, G., Ou, J., Bashmakov, Y., Lobaccaro, J.M., Shimomura, I., Shan, B., Brown, M.S., Goldstein, J.L., and Mangelsdorf, D.J. (2000a). Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRalpha and LXRbeta. *Genes Dev.* *14*, 2819–2830.
- Repa, J.J., Turley, S.D., Lobaccaro, J.A., Medina, J., Li, L., Lustig, K., Shan, B., Heyman, R.A., Dietschy, J.M., and Mangelsdorf, D.J. (2000b). Regulation of absorption and ABC1-mediated efflux of cholesterol by RXR heterodimers. *Science* *289*, 1524–1529.
- Rodríguez-Prados, J.C., Través, P.G., Cuenca, J., Rico, D., Aragonés, J., Martín-Sanz, P., Cascante, M., and Boscá, L. (2010). Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. *J. Immunol.* *185*, 605–614.
- Schultz, J.R., Tu, H., Luk, A., Repa, J.J., Medina, J.C., Li, L., Schwendner, S., Wang, S., Thoolen, M., Mangelsdorf, D.J., et al. (2000). Role of LXRs in control of lipogenesis. *Genes Dev.* *14*, 2831–2838.
- Shimano, H., Shimomura, I., Hammer, R.E., Herz, J., Goldstein, J.L., Brown, M.S., and Horton, J.D. (1997). Elevated levels of SREBP-2 and cholesterol synthesis in livers of mice homozygous for a targeted disruption of the SREBP-1 gene. *J. Clin. Invest.* *100*, 2115–2124.
- Spann, N.J., and Glass, C.K. (2013). Sterols and oxysterols in immune cell function. *Nat. Immunol.* *14*, 893–900.
- Spann, N.J., Garmire, L.X., McDonald, J.G., Myers, D.S., Milne, S.B., Shibata, N., Reichart, D., Fox, J.N., Shaked, I., Heudobler, D., et al. (2012). Regulated

accumulation of desmosterol integrates macrophage lipid metabolism and inflammatory responses. *Cell* 151, 138–152.

Tabas, I., and Glass, C.K. (2013). Anti-inflammatory therapy in chronic disease: challenges and opportunities. *Science* 339, 166–172.

Tannahill, G.M., Curtis, A.M., Adamik, J., Palsson-McDermott, E.M., McGettrick, A.F., Goel, G., Frezza, C., Bernard, N.J., Kelly, B., Foley, N.H., et al. (2013). Succinate is an inflammatory signal that induces IL-1 $\beta$  through HIF-1 $\alpha$ . *Nature* 496, 238–242.

Tencerova, M., Aouadi, M., Vangala, P., Nicoloso, S.M., Yawe, J.C., Cohen, J.L., Shen, Y., Garcia-Menendez, L., Pedersen, D.J., Gallagher-Dorval, K., et al. (2015). Activated Kupffer cells inhibit insulin sensitivity in obese mice. *FASEB J.* 29, 2959–2969.

Varma, V., Yao-Borengasser, A., Rasouli, N., Nolen, G.T., Phanavanh, B., Starks, T., Gurley, C., Simpson, P., McGehee, R.E., Jr., Kern, P.A., and Peterson, C.A. (2009). Muscle inflammatory response and insulin resistance: synergistic interaction between macrophages and fatty acids leads to impaired insulin action. *Am. J. Physiol. Endocrinol. Metab.* 296, E1300–E1310.

Wynn, T.A., Chawla, A., and Pollard, J.W. (2013). Macrophage biology in development, homeostasis and disease. *Nature* 496, 445–455.

Xu, H., Barnes, G.T., Yang, Q., Tan, G., Yang, D., Chou, C.J., Sole, J., Nichols, A., Ross, J.S., Tartaglia, L.A., and Chen, H. (2003). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* 112, 1821–1830.



# Sympathetic neuron–associated macrophages contribute to obesity by importing and metabolizing norepinephrine

Roksana M Pirzgalska<sup>1,11</sup>, Elsa Seixas<sup>1,11</sup>, Jason S Seidman<sup>2</sup>, Verena M Link<sup>2,3</sup>, Noelia Martínez Sánchez<sup>1</sup>, Inês Mahú<sup>1</sup>, Raquel Mendes<sup>1</sup>, Vitka Gres<sup>1</sup>, Nadiya Kubasova<sup>1</sup>, Imogen Morris<sup>1</sup>, Bernardo A Arús<sup>1,4</sup>, Chelsea M Larabee<sup>1</sup>, Miguel Vasques<sup>1,5</sup>, Francisco Tortosa<sup>6</sup>, Ana L Sousa<sup>7</sup>, Sathyavathy Anandan<sup>1</sup>, Erin Tranfield<sup>7</sup>, Maureen K Hahn<sup>8</sup>, Matteo Iannacone<sup>9</sup> , Nathanael J Spann<sup>2</sup>, Christopher K Glass<sup>2</sup> & Ana I Domingos<sup>1,10</sup>

The cellular mechanism(s) linking macrophages to norepinephrine (NE)-mediated regulation of thermogenesis have been a topic of debate. Here we identify sympathetic neuron–associated macrophages (SAMs) as a population of cells that mediate clearance of NE via expression of solute carrier family 6 member 2 (SLC6A2), an NE transporter, and monoamine oxidase A (MAOA), a degradation enzyme. Optogenetic activation of the sympathetic nervous system (SNS) upregulates NE uptake by SAMs and shifts the SAM profile to a more proinflammatory state. NE uptake by SAMs is prevented by genetic deletion of *Slc6a2* or inhibition of the encoded transporter. We also observed an increased proportion of SAMs in the SNS of two mouse models of obesity. Genetic ablation of *Slc6a2* in SAMs increases brown adipose tissue (BAT) content, causes browning of white fat, increases thermogenesis, and leads to substantial and sustained weight loss in obese mice. We further show that this pathway is conserved, as human sympathetic ganglia also contain SAMs expressing the analogous molecular machinery for NE clearance, which thus constitutes a potential target for obesity treatment.

Sympathetic innervation of adipose tissue promotes lipolysis and fat mass reduction via NE signaling<sup>1</sup>. In obesity, chronic local inflammation underlies adipose tissue dysfunction, and macrophages have been shown to play a central role<sup>1,2</sup>. The mechanism that links macrophages in white adipose tissue (WAT) to NE remains controversial. Some groups have reported that anti-inflammatory adipose tissue macrophages (ATMs) in the WAT produce NE to sustain thermogenesis and browning. In direct contradiction, other groups have reported that ATMs do not express a key enzyme required for NE production and that genetic deletion of this enzyme in mouse macrophages has no effect on thermogenesis and body weight<sup>3–6</sup>.

Here we identify a previously undescribed population of SAMs that import and degrade NE via specific proteins that are absent from ATMs. We found by transcriptional profiling of isolated SAMs that neural- and adrenergic-related genes are differentially expressed in these cells relative to other macrophage populations. SAMs accumulate intracellular NE despite lacking enzymes for NE biosynthesis. Using optogenetics, we demonstrate that SNS activity increases NE content and the proinflammatory state of SAMs. We functionally demonstrate that SAMs import and degrade NE via NE transporter

(SLC6A2) and degradation enzyme (MAOA), respectively. We further demonstrate that SAM-mediated clearance of extracellular NE contributes to obesity, as inhibiting NE import by SAMs ameliorates obesity, thermogenesis, and browning in mutant obese (*ob/ob*) mice and mice fed a high-fat diet (HFD). Finally, we demonstrate human relevance for this mechanism, as we found that SAMs are also present in human sympathetic ganglia and express similar molecular machinery as that observed in mice. Thus, the identification of SAMs represents a new contribution to the ongoing controversy surrounding the role of macrophages in thermogenesis and obesity while identifying an unforeseen immunological player in noradrenergic homeostasis with therapeutic potential for obesity.

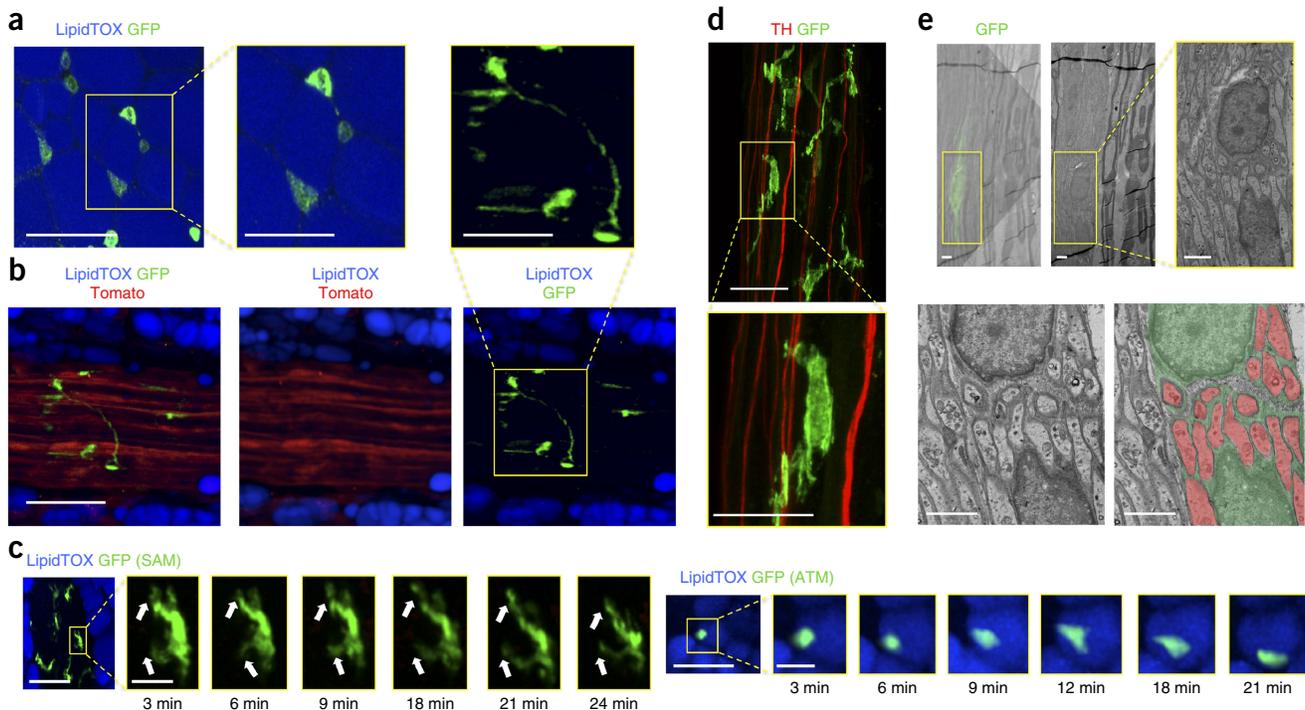
## RESULTS

### Specialized morphology and activation of SNS *Cx3cr1*-expressing cells

Our initial aim was to visualize the *in vivo* morphology of ATMs using two-photon and confocal microscopy in *Cx3cr1*<sup>GFP/+</sup> mice, in which macrophages are labeled with GFP. ATMs in fat parenchyma had a regular circular shape (Fig. 1a), whereas those located on sympathetic

<sup>1</sup>The Howard Hughes Medical Institute (HHMI) and Obesity Laboratory, Instituto Gulbenkian de Ciência, Oeiras, Portugal. <sup>2</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California, USA. <sup>3</sup>Faculty of Biology, Department II, Ludwig-Maximilians Universität München, Planegg-Martinsried, Germany. <sup>4</sup>Departamento de Bioquímica, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. <sup>5</sup>Department of Endocrinology, Curry Cabral Hospital, Centro Hospitalar de Lisboa Central, Lisbon, Portugal. <sup>6</sup>Department of Pathology, Centro Hospitalar Lisboa Norte, Hospital de Santa Maria, EPE, Lisbon, Portugal. <sup>7</sup>Electron Microscopy Unit, Instituto Gulbenkian de Ciência, Oeiras, Portugal. <sup>8</sup>Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. <sup>9</sup>Division of Immunology, Transplantation and Infectious Diseases, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) San Raffaele Scientific Institute, Milan, Italy. <sup>10</sup>The Howard Hughes Medical Institute (HHMI), New York, New York, USA. <sup>11</sup>These authors contributed equally to this work. Correspondence should be addressed to A.I.D. ([dominan@igc.gulbenkian.pt](mailto:dominan@igc.gulbenkian.pt)).

Received 6 May; accepted 12 September; published online 9 October 2017; doi:10.1038/nm.4422



**Figure 1** Sympathetic neuron-associated *Cx3cr1*-GFP<sup>+</sup> cells exhibit differentiated morphology for specific association with SNS neurons. **(a)** Confocal images of WAT isolated from a *Cx3cr1*<sup>GFP/+</sup> mouse and stained using lipid stain LipidTOX (blue) and anti-GFP antibody (green). Images are representative of five similar experiments. **(b)** Confocal images of sympathetic nerve fibers in subcutaneous adipose tissue isolated from a cross of *TH-cre; LSL-Tomato* (red) and *Cx3cr1*<sup>GFP/+</sup> (green) mice. Adipocytes were stained using lipid stain LipidTOX (blue). Images are representative of three similar experiments. Scale bars in **a** and **b**, 50  $\mu$ m. The boxed regions in the main micrographs in **a** and **b** are shown at higher magnification; scale bars, 25  $\mu$ m. **(c)** Intravital multiphoton visualization of a neural–adipose connection in the inguinal fat pad of a live *Cx3cr1*<sup>GFP/+</sup> mouse; LipidTOX (blue) labels adipocytes. Images depict the morphological features and cell dynamics of *Cx3cr1*-GFP<sup>+</sup> cells associated with sympathetic nerve fibers (left) and *Cx3cr1*-GFP<sup>+</sup> cells in the parenchyma of subcutaneous fat (right). Images are representative of three similar experiments. Scale bars, 50  $\mu$ m. Boxed regions in the main micrographs are shown at higher magnification at the indicated time points; scale bars, 10  $\mu$ m. White arrows indicate dendritiform processes over time. **(d)** Confocal images of sympathetic nerve fibers isolated from the inguinal fat pad of a *Cx3cr1*<sup>GFP/+</sup> mouse and stained using anti-TH (red) and anti-GFP (green) antibodies. Images are representative of five similar experiments. Scale bar, 50  $\mu$ m. The boxed region in the main micrograph is shown at higher magnification below; scale bar, 25  $\mu$ m. **(e)** Correlative confocal and transmission electron microscopy of nerve fibers isolated from the subcutaneous fat pad of a *Cx3cr1*<sup>GFP/+</sup> mouse. Shown are an overlay of the GFP fluorescence (green) with the electron micrograph of the same section (upper left; the lower left image is a higher-magnification view of the boxed region), the electron micrograph alone (upper middle; the yellow boxed region is shown at higher magnification to the right), and the electron micrograph from the lower left with false coloring highlighting *Cx3cr1*-GFP<sup>+</sup> cells (green) and sympathetic nerves (red) (lower right). Images are representative of two similar experiments. Scale bars, 2  $\mu$ m.

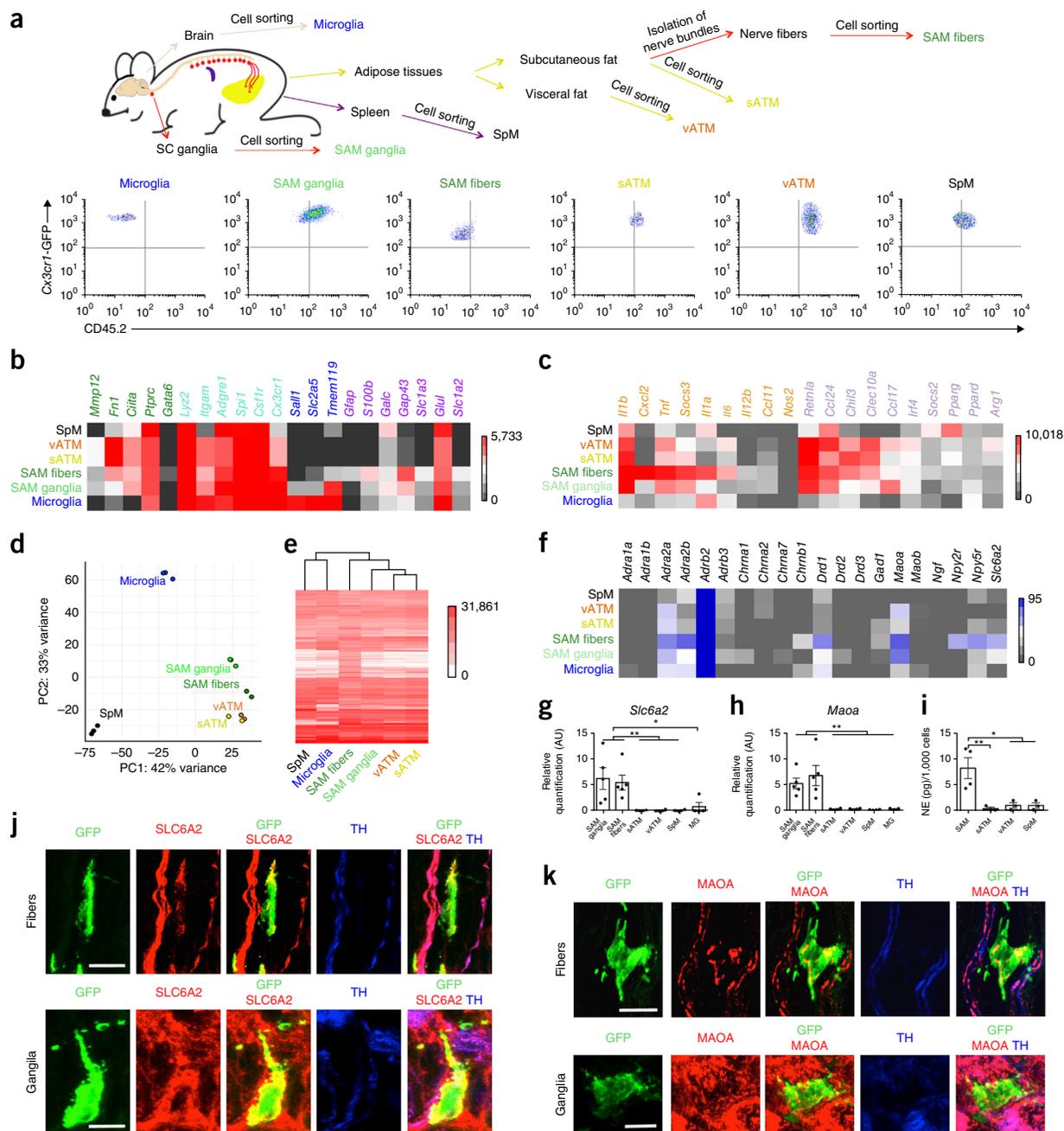
nerve bundles exhibited profuse pseudopodia that extended over a greater surface area (Fig. 1b and Supplementary Fig. 1a,b). Furthermore, we observed that sympathetic neuron-associated *Cx3cr1*-GFP<sup>+</sup> cells displayed dynamic extensions and retractions of dendritiform processes over time (Fig. 1c and Supplementary Video 1). In contrast, ATMs surrounding adipocytes displayed minimal temporal plasticity or displacement (Fig. 1c and Supplementary Video 2). Using correlative light–electron microscopy on WAT-derived nerve bundles, we confirmed that *Cx3cr1*-GFP<sup>+</sup> cells extended thin pseudopodial processes that enveloped nonmyelinated SNS axons (Fig. 1d,e and Supplementary Fig. 1c).

We then investigated whether sympathetic neuron-associated *Cx3cr1*-GFP<sup>+</sup> cells were present in other SNS compartments, such as paravertebral sympathetic ganglia. Through imaging superior cervical ganglia (SCGs) and thoracic chains, we visualized *Cx3cr1*-GFP<sup>+</sup> cells that were morphologically similar to those within WAT-derived SNS bundles (Supplementary Fig. 2). Owing to the established *ex vivo* explant potential of SCGs, we used them along with WAT-derived SNS nerve bundles as model systems for subsequent functional and molecular analyses.

### SNS *Cx3cr1*-expressing SAMs exhibit hematopoietic characteristics

Because nearly all *Cx3cr1*-GFP<sup>+</sup> cells isolated from sympathetic fibers expressed the immune marker CD45 (Supplementary Fig. 3) and macrophage marker F4/80 (Supplementary Fig. 4a), we designated these cells SAMs. Because of the specialized morphology and location of SAMs, we next explored how these cells compared to other tissue macrophages and brain microglia. We sorted F4/80<sup>+</sup>CD45<sup>+</sup> cells from the following tissues: sympathetic ganglia (SAM ganglia), sympathetic nerve fibers from inguinal fat (SAM fibers), neighboring subcutaneous fat (sATM), visceral fat (vATM), spleen (SpM), and brain (microglia) (Fig. 2a; gating details in Supplementary Fig. 3). The relative abundance of CD45<sup>high</sup>*Cx3cr1*-GFP<sup>+</sup> cells was nearly four times higher within nerve fibers (SAMs) than in subcutaneous WAT (sWAT) (sATMs; Supplementary Fig. 4b).

CD45 is highly expressed in hematopoietic cells but expressed at low levels in microglia. Flow cytometric analysis revealed that SAMs are CD45<sup>medium</sup> or CD45<sup>high</sup> (Supplementary Fig. 3), suggesting a hematopoietic origin for these cells. To test this hypothesis, we generated chimeras through transplantation of bone marrow from



**Figure 2** SAMs highly express macrophage-associated markers and possess the machinery for uptake and degradation of norepinephrine. **(a)** Top, schematic representation of tissue dissections and processing of macrophages isolated from the following tissues: brain, spleen, visceral fat, subcutaneous fat, sympathetic nerve fibers from subcutaneous fat, and superior cervical ganglia (SC ganglia). Bottom, representative flow cytometry dot plots indicating the CD45.2 status of macrophages from each tissue analyzed. **(b)** Heat map showing expression of genes associated with macrophage (green and cyan), microglial (blue and cyan), and glial (purple) profiles as determined by low-input RNA-seq. Values are in reads per kilobase of transcript per million mapped reads (RPKM). **(c)** Heat map showing expression of proinflammatory (orange) and anti-inflammatory (purple) genes as determined by low-input RNA-seq. Values are in RPKM. **(d)** PCA based on the top 500 genes with the most variable expression across SAM fibers (green), SAM ganglia (light green), vATM (orange), sATM (yellow), SpM (black), and microglia (blue). Each dot represents an independent experiment. **(e)** Heat map of transcript levels (RPKM values) based on the 5,000 genes most highly expressed by SAM fibers as determined by low-input RNA-seq. **(f)** Heat map showing expression of genes encoding neurotransmitter receptors, transporters, and catalytic enzymes. RPKM values were determined by low-input RNA-seq. Values in **b–f** represent three (SpM, microglia, and SAM ganglia) or two (vATM, sATM, and SAM fibers) independent experiments. **(g)** Expression of mRNA for *Slc6a2* as determined by qRT-PCR presented with normalization to *Gapdh* expression. Each data point represents tissues pooled from ten mice.  $n = 5$  experiments for SAM fibers and SAM ganglia and  $n = 4$  experiments for SpM, vATM, sATM, and microglia (MG). **(h)** Expression of mRNA for *Maoa* as determined by qRT-PCR with normalization to *Gapdh* expression. Each data point represents tissues pooled from ten mice.  $n = 5$  experiments for SAM fibers and SAM ganglia,  $n = 4$  experiments for SpM, vATM, and sATM, and  $n = 3$  experiments for microglia. **(i)** NE content in sorted CD45.2 (PE)+F4/80 (Alexa Fluor 647)<sup>+</sup> cells measured by NE ELISA. The numbers of cells used in NE assays were as follows:  $858 \pm 258$  for SAMs ( $n = 4$  experiments) and 1,000 cells for sATMs, vATMs, and SpMs ( $n = 3$  experiments). **(j, k)** Confocal images of sympathetic nerve fibers (top) and SCG (bottom) isolated from *Cx3cr1*<sup>GFP/+</sup> mice and stained using anti-GFP (green) and anti-TH (blue) antibodies together with anti-SLC6A2 (**j**) or anti-MAOA (**k**) antibody (red). Images are representative of two experiments. Scale bars, 10  $\mu$ m. Data in **g–i** were analyzed by one-way ANOVA followed by Tukey's multiple-comparisons test. Data are shown as average  $\pm$  s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ .

CD45.2 *Cx3cr1*<sup>GFP/+</sup> donors into irradiated CD45.1 recipient mice and observed complete repopulation of CD45<sup>+</sup> cells derived from the CD45.2 *Cx3cr1*<sup>GFP/+</sup> donors (Supplementary Fig. 4c). Eight weeks following transplantation, we established that CD45.2<sup>+</sup>*Cx3cr1*-GFP<sup>+</sup> SAMs repopulated sympathetic nerve bundles in WAT, whereas microglial repopulation in the brain did not occur (Supplementary Fig. 4d). This suggests that SAMs in sympathetic fibers have an origin similar to that of other hematopoietic macrophages rather than being from a microglial lineage.

### SAM expression profile is more macrophage- than glial-like

Considering the association of SAMs with neurons, we asked how the gene expression profile of SAMs compared to those of other tissue-resident macrophages and microglia (Fig. 2). We sorted macrophages from various tissues as described above (F4/80<sup>+</sup>CD45<sup>+</sup> cells designated as SAM ganglia, SAM fibers, sATMs, vATMs, SpMs, and microglia; Fig. 2a and Supplementary Fig. 3) and profiled gene expression by low-input RNA-seq (Fig. 2b–f). As expected, SAMs highly expressed markers common to both microglia and macrophages, such as *Adgre1*, *Csf1r*, and *Cx3cr1* (Fig. 2b). SAMs expressed macrophage-associated genes whose expression was excluded from microglia, such as *Fn1* and *Ciita* (Fig. 2b)<sup>7</sup>. Flow cytometric analysis showed that additional macrophage-specific markers whose expression was excluded from microglia (CD68, Ly6C, major histocompatibility complex II (MHCII), and CD11b) were also highly expressed in SAMs (Supplementary Fig. 5a,b). SAMs did not robustly express microglial- or glial-specific genes relative to macrophage-specific genes (Fig. 2b and Supplementary Fig. 5c)<sup>8–17</sup>. Expression of the *Sall1* gene, encoding a key microglial lineage-determining transcription factor, was strikingly absent from SAMs<sup>18</sup> (Fig. 2b).

Principal-component analysis (PCA) of the RNA-seq data showed tight clustering across replicates, indicating low contamination and high reproducibility (Fig. 2d). The absence of tyrosine hydroxylase (*Th*) expression in SAMs (Supplementary Fig. 5d) further excluded the possibility of contaminating cargo from neighboring cells, as *Th* was highly expressed in adjacent SNS neurons (Fig. 1b,d). PCA indicated that SAMs from fibers and ganglia were closely related, but both were distant from microglia and other macrophages (Fig. 2d). This was confirmed by phylogenetic analysis (Fig. 2e).

We hypothesized that the increased motility of SAMs (Fig. 1c) could indicate an activated, proinflammatory state. Therefore, we measured expression of a constellation of pro- and anti-inflammatory markers in SAMs by RNA-seq (Fig. 2c). Relative to other macrophage populations, SAMs highly expressed genes associated with macrophage activation, including *Cxcl2*, *Tnf*, *Socs3*, and *Il1a* (Fig. 2c), suggesting a constitutively proinflammatory steady state.

### SAMs are phylogenetically distinct from other macrophages

Consistent with the PCA results (Fig. 2d), Pearson correlation analyses of transcript levels indicated differential expression patterns across SAMs, sATMs, vATMs, SpMs, and microglia (Supplementary Fig. 6a,b). Adipose tissue macrophages (sATMs and vATMs) showed similar expression landscapes ( $R = 0.92$ ) that were distant from those of SAMs in fibers ( $R = 0.63$  for sATMs and  $R = 0.61$  for vATMs; Supplementary Fig. 6b). The expression landscapes of microglia and spleen macrophages were least correlated with other groups (Supplementary Fig. 6b).

Gene ontology analyses identified several biological processes associated with genes whose expression was enriched in SAMs relative to surrounding sATMs (Supplementary Fig. 6c). SAMs

preferentially expressed genes involved in synaptic signaling, cell–cell adhesion, and neuron development (Supplementary Fig. 6c), suggesting that these cells fulfill an intrinsic role in local neuronal maintenance. Taken together, these data demonstrate divergent gene expression patterns in SAMs and ATMs, constituting within-tissue macrophage specialization.

### SAMs import and degrade but do not synthesize NE

We next examined specific transcripts corresponding to the genes with divergent macrophage expression. The aforementioned populations of macrophages were sorted (Fig. 2a and Supplementary Fig. 3) for transcriptome analysis via low-input RNA-seq. Considering the gene ontology results (Supplementary Fig. 6c) and spatial proximity of SAMs to nerves (Fig. 1), we hypothesized that there would be differential expression of neurotransmitter receptors, transporters, or catalytic enzymes among these macrophage populations (Fig. 2f). In agreement with the Immunological Genome Project (ImmGen) database, we detected abundant expression of the *Adrb2* gene encoding  $\beta_2$  adrenergic receptor in all macrophage populations (Fig. 2f), which was confirmed by qRT-PCR (Supplementary Fig. 6d).

However, SAMs were the only population that expressed *Slc6a2*, the gene encoding the NE transporter (Fig. 2f). Similarly, *Maoa* was highly expressed in SAMs relative to the other macrophage types (Fig. 2f). Both results were validated by qRT-PCR (Fig. 2g,h and Supplementary Table 1). As SLC6A2 imports and MAOA degrades NE, we also tested for the presence of NE and detected it through enzyme-linked immunosorbent assay (ELISA) in sorted SAMs (Fig. 2i and Supplementary Fig. 6e). In agreement with our results, neither *Slc6a2* nor *Maoa* was substantially expressed in any macrophage population listed in the ImmGen database. Furthermore, we validated SLC6A2 and MAOA protein expression by immunofluorescence in SNS nerve fibers and SCG cryosections from *Cx3cr1*<sup>GFP/+</sup> mice (Fig. 2j,k). Representative photomicrographs show that GFP-expressing SAMs were double positive for membrane-bound SLC6A2 (Fig. 2j) and mitochondria-bound MAOA (Fig. 2k).

As SAMs, but not the other macrophage types assessed, possess the molecular machinery for import (Fig. 2f,g,j) and degradation (Fig. 2f,h,k) of NE, as well as considerably more NE than other macrophages (Fig. 2i and Supplementary Fig. 6e), we tested the possibility that SAMs synthesize NE. Through qRT-PCR of sorted SAMs, we did not detect expression of *Th*, which encodes an enzyme necessary for NE biosynthesis (Supplementary Fig. 5d). Taken together, these results indicate that SAMs have the molecular machinery for importing and degrading NE but not for biosynthesis of it.

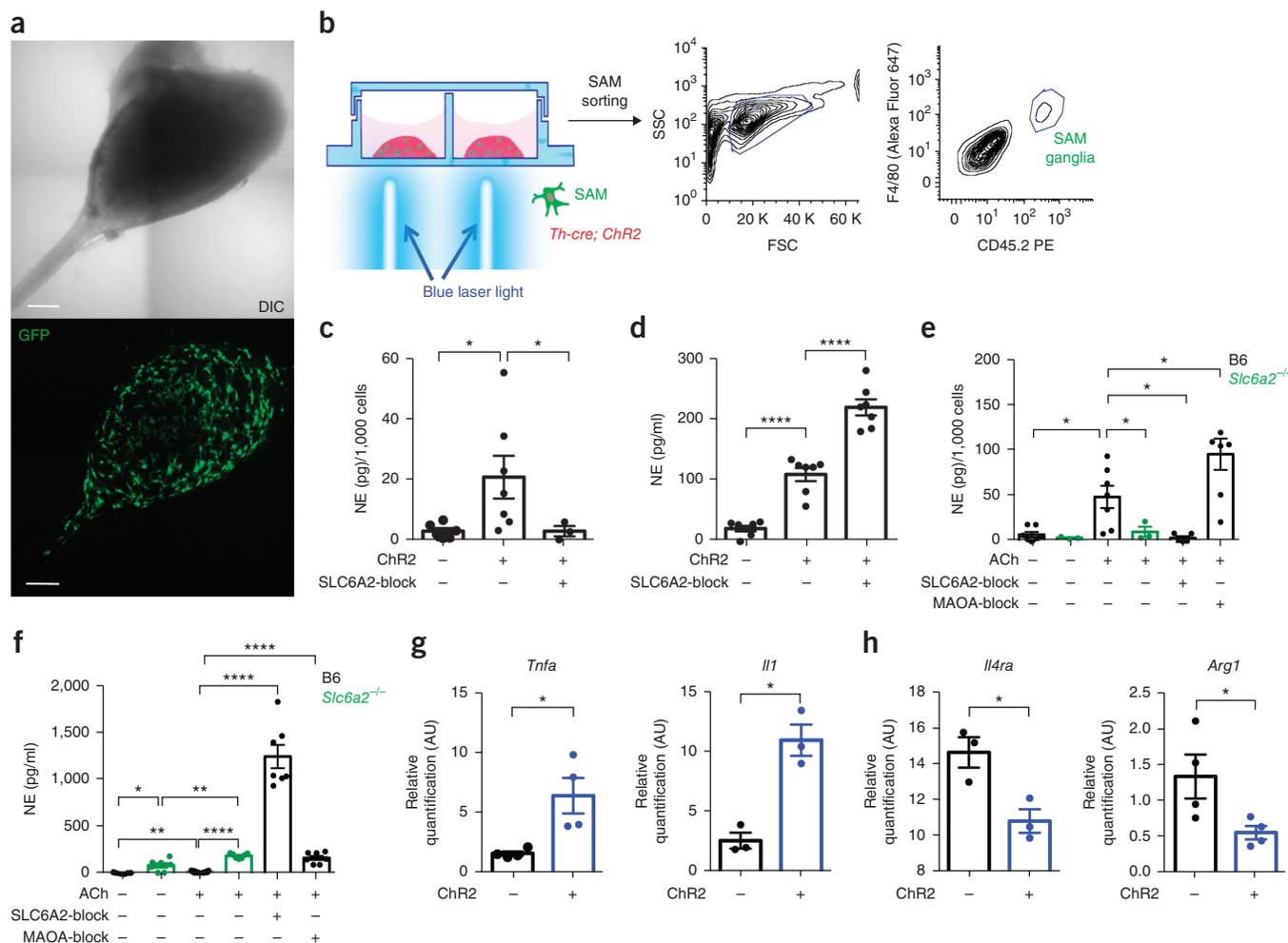
To explore the responsiveness of SAMs to NE, we optogenetically stimulated sympathetic neurons in SCG cultures from mice produced by crossing *Th-cre* mice with *loxP-STOP-loxP* (LSL)-*Chr2-YFP* mice<sup>1</sup>, which allowed us to visualize sympathetic neuron–macrophage interactions *ex vivo* (Fig. 3a,b). After optogenetic stimulation, we measured the NE content of sorted CD45<sup>+</sup>F4/80<sup>+</sup> cells. SAMs from channelrhodopsin-2 (Chr2)<sup>+</sup> cultures exhibited significantly higher NE levels (Fig. 3c) that were proportional to NE availability in the culture medium (Fig. 3d). NE release by Chr2<sup>+</sup> neurons was significantly higher than that from Chr2<sup>-</sup> neurons (Fig. 3d). Uptake of NE by SAMs was prevented by pharmacological blockade of SLC6A2 using the pharmacological inhibitor nisoxetine, despite the significant increase of NE in the culture medium (Fig. 3c,d).

To validate our optogenetic findings with a physiologically relevant stimulus, we activated SNS explants with acetylcholine (ACh), which is presynaptically released from spinal cord neurons to innervate SCGs.

ACh-treated CD45<sup>+</sup>F4/80<sup>+</sup> cells sorted from SCG explants contained significantly higher levels of NE than vehicle-treated controls (Fig. 3e). We validated that blockade of the NE importer SLC6A2 by nisoxetine prevented NE accumulation in SAMs (Fig. 3e). Co-incubation with ACh and nisoxetine further abolished NE uptake (Fig. 3e), despite the substantial increase in extracellular NE levels in the

culture medium (Fig. 3f). These results, along with the negligible expression levels of *Chrna1* (AChR) in SAMs (Supplementary Fig. 7a; also validated by qRT-PCR in Supplementary Fig. 7b), exclude a role for acetylcholine receptors (AChRs) in mediating NE import.

Next, we assessed the effect of blocking MAOA on NE content in CD45<sup>+</sup>F4/80<sup>+</sup> cells (Fig. 3e). Treatment with the MAOA inhibitor



**Figure 3** SAMs import and metabolize norepinephrine via SLC6A2 and MAOA, respectively, to regulate extracellular norepinephrine availability. **(a)** Representative images of *ex vivo* SCG explant cultures. Top, the area of the sympathetic ganglia is represented using the reflected-light differential interference contrast (DIC) channel. Bottom, *Cx3cr1*-GFP<sup>+</sup> cells in the same explant culture (GFP channel). Images are representative of 20 similar experiments. Scale bar, 100  $\mu$ m. **(b)** Schematic representation of optogenetic activation of sympathetic SCG explant culture (left) followed by CD45.2 (PE)+F4/80 (Alexa Fluor 647)<sup>+</sup> cell sorting (right). FSC, forward scatter; SSC, side scatter. **(c)** NE content in CD45.2<sup>+</sup>F4/80<sup>+</sup> cells isolated from SCG explant cultures from *Th*-*cre*; LSL-*ChR2*-YFP and LSL-*ChR2*-YFP mice after optogenetic activation. Each data point represents tissues pooled from six mice.  $n = 3$ –7 experiments. The following numbers of cells were used in NE assays (run in duplicate): 189  $\pm$  30 from *Th*-*cre*; LSL-*ChR2*-YFP SCG ( $n = 7$ ), 126  $\pm$  21 from LSL-*ChR2*-YFP SCG ( $n = 6$ ), and 159  $\pm$  19 from *Th*-*cre*; LSL-*ChR2*-YFP SCG stimulated with SLC6A2 blocker ( $n = 3$ ). **(d)** *Ex vivo* NE release upon optogenetic stimulation of SCG explants isolated from *Th*-*cre*; LSL-*ChR2*-YFP and LSL-*ChR2*-YFP mice. Each data point represents medium collected from one explant culture.  $n = 7$  per group. **(e)** NE content in CD45.2<sup>+</sup>F4/80<sup>+</sup> cells isolated from the SCG of either B6 or *Slc6a2*<sup>-/-</sup> mice and then incubated with ACh, ACh and SLC6A2 blocker, ACh and MAOA blocker, or culture medium. Each data point represents tissues pooled from six mice.  $n = 3$ –7 experiments. The following numbers of cells were used in NE assays (run in duplicate): 364  $\pm$  128 from B6 SCG ( $n = 7$ ), 238  $\pm$  55 from *Slc6a2*<sup>-/-</sup> SCG ( $n = 3$ ), 216  $\pm$  58 from B6 SCG incubated with ACh ( $n = 7$ ), 201  $\pm$  63 from *Slc6a2*<sup>-/-</sup> SCG incubated with ACh ( $n = 3$ ), 196  $\pm$  18 from B6 SCG incubated with ACh and SLC6A2 blocker ( $n = 5$ ), and 133  $\pm$  11 from B6 SCG incubated with ACh and MAOA blocker ( $n = 7$ ). **(f)** *Ex vivo* NE release from the SCG of either B6 or *Slc6a2*<sup>-/-</sup> mice after incubation with ACh, ACh and SLC6A2 blocker, ACh and MAOA blocker, or culture medium. Each data point represents medium collected from one explant culture.  $n = 7$  per group. **(g)** Expression of mRNA as determined by qRT-PCR relative to *Gapdh* expression for proinflammatory genes (*Tnfa* and *Il1*) in CD45.2<sup>+</sup>F4/80<sup>+</sup> cells isolated from SCG explant cultures from *Th*-*cre*; LSL-*ChR2*-YFP (blue) and LSL-*ChR2*-YFP (black) mice. Prior to cell sorting, SCG explants were optogenetically stimulated.  $n = 3$ –4 experiments (for *Tnfa*,  $n = 4$ ,  $P = 0.0467$ ; for *Il1*,  $n = 3$ ,  $P = 0.011$ ). **(h)** Expression of mRNA as determined by qRT-PCR relative to *Gapdh* expression for anti-inflammatory genes (*Il4ra* and *Arg1*) in CD45.2<sup>+</sup>F4/80<sup>+</sup> cells isolated from SCG explant cultures from *Th*-*cre*; LSL-*ChR2*-YFP (blue) and LSL-*ChR2*-YFP (black) mice. Prior to cell sorting, SCG explants were optogenetically stimulated.  $n = 3$ –4 experiments (for *Il4ra*,  $n = 3$ ,  $P = 0.0257$ ; for *Arg1*,  $n = 4$ ,  $P = 0.0497$ ). Data in **c**–**h** were analyzed by two-tailed unpaired Student's *t*-test and are shown as average  $\pm$  s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\*\* $P < 0.0001$ .

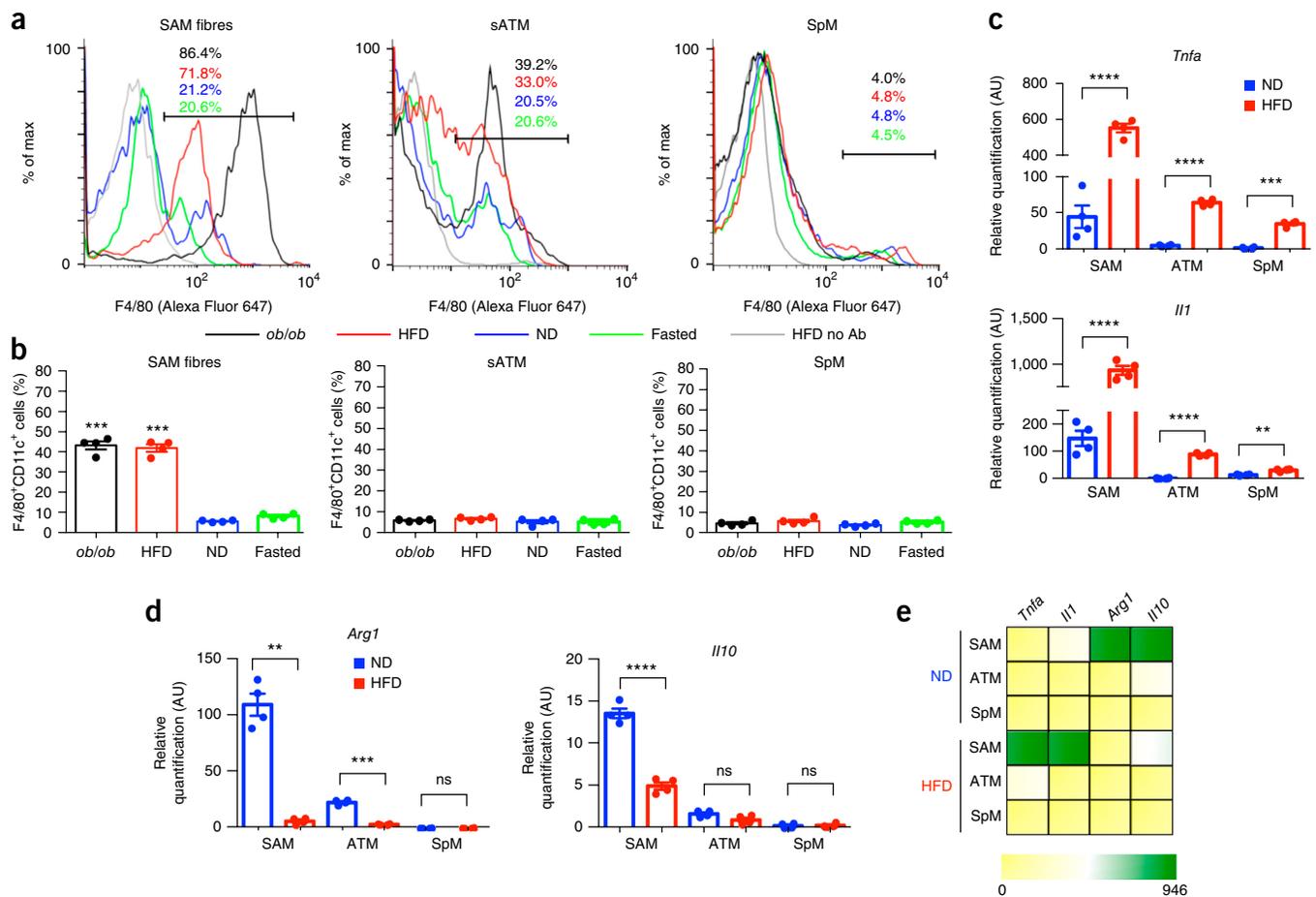
clorgyline was sufficient to nearly double intracellular NE levels in SAMs (Fig. 3e). In agreement with this finding, clorgyline increased NE levels in the medium (Fig. 3f), to which neuronal MAOA expression may also contribute. Genetic ablation of *Slc6a2* (in SCG cultures isolated from *Slc6a2*<sup>-/-</sup> mice) prevented NE uptake by SAMs regardless of NE availability in the culture medium (Fig. 3e,f). Finally, ATMs cultured *in vitro* with NE did not accumulate intracellular NE (Supplementary Fig. 7c), further demonstrating the specificity of NE uptake by SAMs. Altogether, our results indicate that *Slc6a2* is required for NE accumulation in SAMs.

We further probed whether the availability of NE, which can be manipulated *in vivo* by optogenetic activation of SNS neurons, changes the inflammatory profile of SAMs. We found that optogenetic stimulation of SCG explants correlated with an increase in proinflammatory gene expression, as measured by changes in expression of

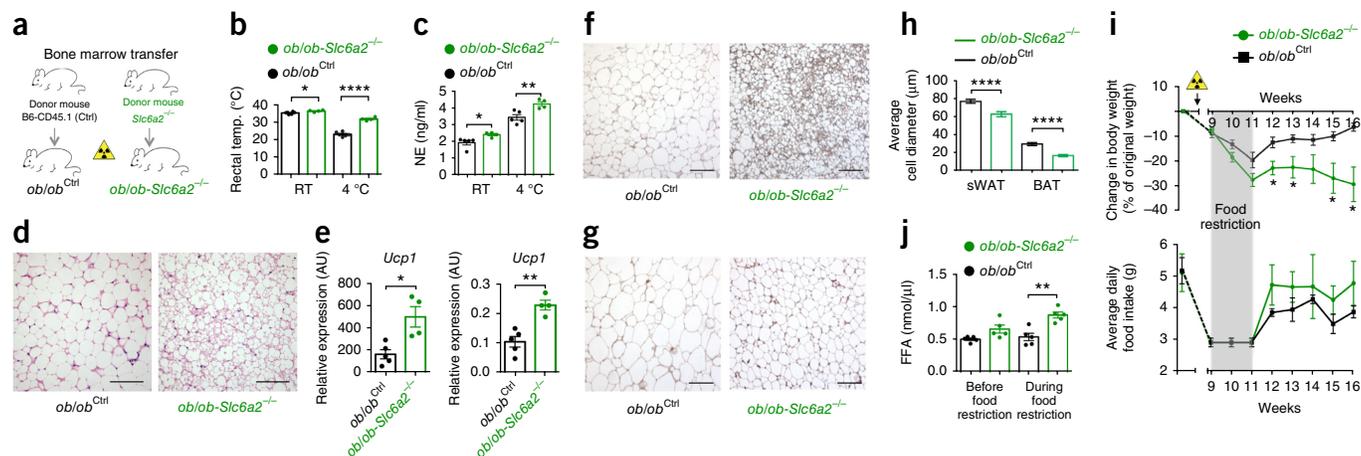
*Tnf* (*Tnfa*) and *Il1a* (*Il1l*) (Fig. 3g), and a decrease in the expression of anti-inflammatory genes, as measured by changes in expression of *Il4r* (*Il4ra*) and *Arg1* (Fig. 3h).

### SAMs are recruited and activated in obesity

We next used two mouse models to characterize the effect of obesity on tissue-specific functions of SAMs. In total, we employed four experimental groups: HFD, *Lep* (*Leptin*)-deficient (*ob/ob*), normal diet (ND), and 24-h-fasted ND mice. Flow cytometric analysis demonstrated that both obesity models (HFD and *ob/ob*) exhibited significantly higher percentages of SAMs than lean mice (ND) (Fig. 4a and Supplementary Fig. 8a). Furthermore, the acute metabolic challenge of fasting did not result in upregulation of SAMs, suggesting an obesity-specific causation of elevated macrophage content in sympathetic fibers (Fig. 4a and Supplementary Fig. 8a). Within the F4/80<sup>+</sup>



**Figure 4** Obesity-induced accumulation of SAMs. (a) Representative histograms showing percentages of F4/80 (Alexa Fluor 647)<sup>+</sup> cells in sympathetic nerve fibers (left), subcutaneous adipose tissue (middle), and spleen (right) in mice that were genetically obese (*ob/ob*; black), obese due to HFD (red), ND fed (blue), or fasted for 24 h (green). CD45.2 (PE)<sup>+</sup> cells were gated. Histograms are representative of four independent experiments. HFD no Ab, cells without antibody staining harvested from mice fed a HFD. Black lines indicate the region defining F4/80<sup>+</sup> cells. (b) Percentages of F4/80 (Alexa Fluor 647)<sup>+</sup>CD11c (FITC)<sup>+</sup> cells in sympathetic nerve fibers (left), subcutaneous adipose tissue (middle), and spleen (right) in mice that were genetically obese (*ob/ob*; black), obese due to HFD (red), ND fed (blue), or fasted for 24 h (green). CD45.2 (PE)<sup>+</sup> cells were gated. *n* = 4 experiments per group. Each data point represents one experiment. (c) Expression of mRNA as determined by qRT-PCR relative to *Gapdh* expression for proinflammatory genes (*Tnfa* and *Il1l*) in CD45.2<sup>+</sup>F4/80<sup>+</sup> cells in sympathetic nerve fibers (SAMs), subcutaneous adipose tissue (ATMs), and spleen (SpMs) isolated from mice that were fed either ND (blue) or HFD (red). *n* = 4 experiments per group. Each data point represents tissues pooled from ten mice. (d) Expression of mRNA as determined by qRT-PCR relative to *Gapdh* expression for anti-inflammatory genes (*Arg1* and *Il10*) in CD45.2<sup>+</sup>F4/80<sup>+</sup> cells including SAMs, ATMs, and SpMs isolated from mice that were fed either ND (blue) or HFD (red). *n* = 4 experiments per group. Each data point represents tissues pooled from ten mice. (e) Heat map showing the expression of pro- and anti-inflammatory genes as determined by the qRT-PCR analyses in c and d. Data in b were analyzed by one-way ANOVA followed by Bonferroni multiple-comparisons test with ND as the control group. Data in c and d were analyzed by two-tailed unpaired Student's *t*-test. Data are shown as average ± s.e.m. \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* < 0.0001; ns, not significant.



**Figure 5** Loss of *Slc6a2* function in SAMs rescues the thermogenic capacity of *ob/ob* mice. (a) Schematic representation of bone marrow transplant from either *Slc6a2*<sup>-/-</sup> or control B6 (CD45.1) mice into genetically obese *ob/ob* mice (*ob/ob-Slc6a2*<sup>-/-</sup> and *ob/ob*<sup>Ctrl</sup> chimeras, respectively). (b) Rectal temperature of *ob/ob*<sup>Ctrl</sup> (black) and *ob/ob-Slc6a2*<sup>-/-</sup> (green) chimeras was measured at room temperature (RT) and after 2 h of cold challenge (4 °C). Each data point represents one mouse. *n* = 4 *ob/ob-Slc6a2*<sup>-/-</sup> mice and *n* = 6 *ob/ob*<sup>Ctrl</sup> mice. \**P* = 0.025, \*\*\*\**P* < 0.0001. (c) Serum levels of NE in *ob/ob*<sup>Ctrl</sup> (black) and *ob/ob-Slc6a2*<sup>-/-</sup> (green) chimeras were measured at room temperature and after 2 h of cold exposure (4 °C). Each data point represents one mouse. *n* = 4 mice per group for *ob/ob-Slc6a2*<sup>-/-</sup> mice and *n* = 5 mice per group for *ob/ob*<sup>Ctrl</sup> mice. \**P* = 0.022, \*\**P* = 0.0072. (d) Optical micrographs of BAT removed from *ob/ob* chimeras following 2 h of cold challenge (4 °C) and stained with H&E. Left, BAT from an *ob/ob*<sup>Ctrl</sup> chimera. Right, BAT from an *ob/ob-Slc6a2*<sup>-/-</sup> chimera. Images are representative of fat organs collected from four *ob/ob*<sup>Ctrl</sup> and six *ob/ob-Slc6a2*<sup>-/-</sup> mice. (e) Expression of mRNA for *Ucp1* as determined by qRT-PCR relative to *Gapdh* expression in BAT (left) and sWAT (right) dissected after 2 h of cold challenge (4 °C). Each data point represents one mouse. *n* = 4 *ob/ob-Slc6a2*<sup>-/-</sup> mice (green) and *n* = 5 *ob/ob*<sup>Ctrl</sup> mice (black). \**P* = 0.0269, \*\**P* = 0.0015. (f) Optical micrographs of BAT dissected from *ob/ob*<sup>Ctrl</sup> (left) and *ob/ob-Slc6a2*<sup>-/-</sup> (right) chimeras following 2 h of cold challenge (4 °C) and stained with anti-UCP1 antibody. Images are representative of fat organs collected from four *ob/ob*<sup>Ctrl</sup> and six *ob/ob-Slc6a2*<sup>-/-</sup> mice. (g) Optical micrographs of sWAT dissected from *ob/ob*<sup>Ctrl</sup> (left) and *ob/ob-Slc6a2*<sup>-/-</sup> mice (right) following 2 h of cold challenge (4 °C) and stained with anti-UCP1 antibody. Images are representative of fat organs collected from four *ob/ob*<sup>Ctrl</sup> and six *ob/ob-Slc6a2*<sup>-/-</sup> mice. (h) Average adipocyte diameter quantified from optical micrographs of sWAT and BAT from *ob/ob* chimeras following 2 h of cold challenge (4 °C). Measurements are representative of four (*ob/ob-Slc6a2*<sup>-/-</sup>) and six (*ob/ob*<sup>Ctrl</sup>) independent micrographs. 18–34 measurements were obtained per micrograph. *n* = 169 cells for *ob/ob*<sup>Ctrl</sup> sWAT, *n* = 120 cells for *ob/ob-Slc6a2*<sup>-/-</sup> sWAT, *n* = 180 cells for *ob/ob*<sup>Ctrl</sup> BAT, *n* = 120 cells for *ob/ob-Slc6a2*<sup>-/-</sup> BAT. \*\*\*\**P* < 0.0001. (i) Body weight change (top) and daily food intake (bottom) of *ob/ob*<sup>Ctrl</sup> (*n* = 4 mice) and *ob/ob-Slc6a2*<sup>-/-</sup> (*n* = 6 mice) chimeras monitored for 7 weeks following 2 weeks of food intake normalization (0.06 g of food per 1 g of body weight per day; gray shading) that started 9 weeks after bone marrow transplant. The yellow triangle indicates when irradiation was performed. \**P* < 0.05. (j) Blood plasma nonesterified (free) fatty acid (FFA) concentration in *ob/ob*<sup>Ctrl</sup> and *ob/ob-Slc6a2*<sup>-/-</sup> chimeras measured 8 weeks after bone marrow transplant before and while mice were under a regimen of 0.06 g of food per 1 g of body weight per day. *n* = 5 mice per group. \*\**P* = 0.0022. Data in **b**, **c**, **e**, **h**, and **j** were analyzed by two-tailed unpaired Student's *t*-test and in **i** by multiple *t*-tests (one Student's *t*-test per row with correction for multiple comparisons using the Holm–Sidak method). Data are shown as average ± s.e.m. Scale bars in **d**, **f**, and **g**, 100 μm.

SAM fraction from HFD and *ob/ob* mice, we noted a high frequency of CD11c<sup>+</sup> cells (Fig. 4b), which are hallmarks of inflammation and insulin resistance in human obesity<sup>19</sup>. In contrast to SAM accumulation in SNS nerve fibers dissected from WAT, SAMs did not accumulate in SCG, which innervates neck structures such as salivary glands (Supplementary Fig. 8b).

The differential distribution of macrophages under conditions of obesity suggests that cytokine levels are also sensitive to obesity. Comparison of the anti- and proinflammatory gene profiles of SAMs, ATMs, and SpMs (Fig. 4c–e) revealed that obesity correlated with higher levels of proinflammatory genes (i.e., *Thfa* or *Il1*; Fig. 4c,e) and lower levels of anti-inflammatory genes (i.e., *Arg1* or *Il10*; Fig. 4d,e).

To determine whether local proliferation contributes to SAM accumulation, we measured the proliferation marker Ki-67 in SAMs by flow cytometry (Supplementary Fig. 8c,d). We observed that obesity (in the HFD and *ob/ob* models) did not substantially increase Ki-67<sup>+</sup> SAM percentage, whereas (in accordance with previous reports<sup>20</sup>) obesity increased Ki-67<sup>+</sup> ATMs from sWAT (Supplementary Fig. 8d).

### *Slc6a2* deletion in SAMs rescues obesity

We probed how ablating *Slc6a2* in SAMs affects obesity-associated pathology. We considered a Cre–loxP approach, but the established

macrophage Cre lines (*Cx3cr1-cre*<sup>21,22</sup> and *LyzM-cre*<sup>23</sup>) would not allow for specificity of ablation to SAMs. We thus took advantage of the cell-type specificity of *Slc6a2* expression, which is high in SAMs but negligible in other macrophage and hematopoietic cell populations (Fig. 2b,g and ImmGen<sup>24</sup>). We validated that there was not another population of hematopoietic origin expressing *Slc6a2* aside from SAMs; a rare population of CD45<sup>+</sup>F4/80<sup>-</sup> cells was present in SCG (Supplementary Fig. 9a) but did not express SLC6A2 (Supplementary Fig. 9b). SAM-specific genetic ablation of *Slc6a2* was achieved through bone marrow transfer from *Slc6a2*<sup>-/-</sup> mice<sup>25</sup> into genetically obese *ob/ob* recipients (*ob/ob-Slc6a2*<sup>-/-</sup>; Fig. 5a). Control chimeras consisted of *ob/ob* mice (*ob/ob*<sup>Ctrl</sup>) that received a bone marrow transfer from B6 CD45.1 mice. Chimeras recovered for 9 weeks following transplant to allow irradiation-induced inflammation to subside.

As cold temperature is a robust driver of SNS activity, we challenged mice for 2 h at 4 °C and observed that *ob/ob-Slc6a2*<sup>-/-</sup> chimeras displayed superior capacity for maintaining body temperature as compared to control *ob/ob*<sup>Ctrl</sup> chimeras (Fig. 5b). These thermogenic effects were accompanied by significant upregulation of NE levels in serum (Fig. 5c), rescue of BAT morphology (Fig. 5d), and browning of white fat, as measured by *Ucp1* mRNA and uncoupling protein 1 (UCP1) levels (Fig. 5e–g).

Transplant of bone marrow from *Slc6a2*<sup>-/-</sup> mice into *ob/ob* mice prevented obesity-induced hypertrophy of both BAT and WAT adipocytes (Fig. 5h) but did not affect total body weight (Fig. 5i). Because food-restriction challenge drives SNS activity and mobilizes lipid stores from adipose tissue, we normalized the daily food intake of the *ob/ob* chimeras for 2 weeks (Fig. 5i,j). After a dieting challenge, *ob/ob-Slc6a2*<sup>-/-</sup> mice lost nearly 30% of their body weight relative to their original body weight, after which their body weight was stable for up to 16 weeks, even after they were given *ad libitum* access to food (Fig. 5i). We also found that *ob/ob-Slc6a2*<sup>-/-</sup> mice exhibited higher lipid mobilization during food restriction relative to controls (Fig. 5j).

We analyzed wild-type B6 chimeras reconstituted with bone marrow from *Slc6a2*<sup>-/-</sup> mice relative to CD45.1 controls (B6-*Slc6a2*<sup>-/-</sup> and B6<sup>Ctrl</sup> chimeras, respectively) (Supplementary Fig. 9c). SAMs from B6-*Slc6a2*<sup>-/-</sup> chimeras did not accumulate NE (Supplementary Fig. 9d). In accordance with the results from *ob/ob* chimeras (Fig. 5), B6-*Slc6a2*<sup>-/-</sup> chimeras also exhibited increased serum levels of NE, thermogenesis, and lipolysis, as well as marked weight loss, relative to control mice (Supplementary Fig. 9e–i). Upon challenge with HFD, we observed weight-gain prevention in B6-*Slc6a2*<sup>-/-</sup> but not B6<sup>Ctrl</sup> mice (Supplementary Fig. 9g). These results indicate a notable anti-obesity effect of SAM-specific *Slc6a2* ablation.

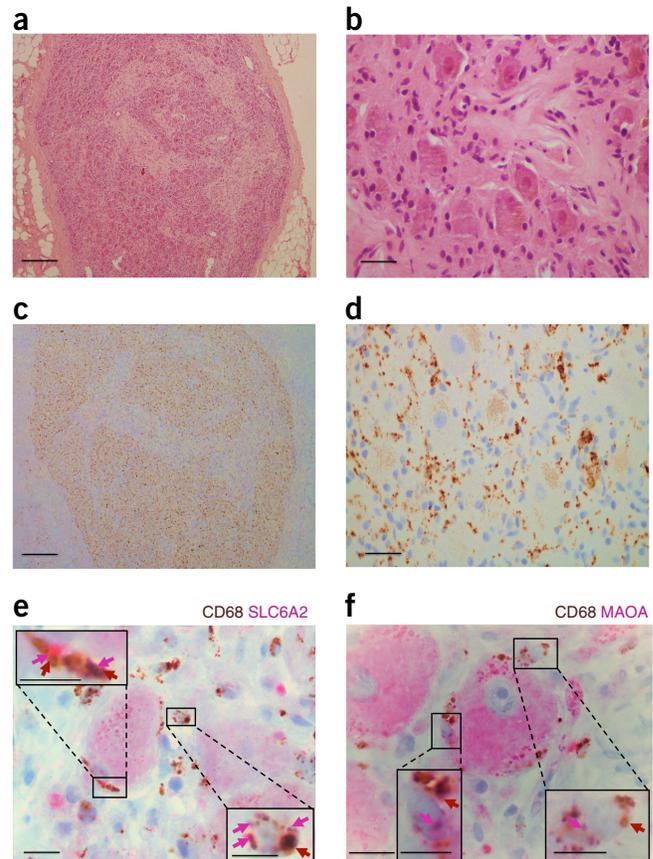
### SAMs are in BAT and act as a NE sink

In light of the enhanced thermogenic capacity of *ob/ob-Slc6a2*<sup>-/-</sup> chimeras, we questioned whether SAMs are present in BAT (Supplementary Fig. 10). BAT did contain *Cx3cr1*-GFP<sup>+</sup> cells (in accordance with a previous report<sup>19</sup>) that exhibited a morphology intermediate to those of SAMs (multiple pseudopodia) and ATMs (round) (Supplementary Fig. 10a as compared to Fig. 1c). Some of these cells appeared to make close contacts with thin TH<sup>+</sup> axons (Supplementary Fig. 10a). Because TH<sup>+</sup> nerve fibers in BAT are too delicate for dissection, we sorted macrophages from whole BAT for qRT-PCR analysis. *Slc6a2* and *Maoad* were expressed in BAT macrophages, although at lower levels than in SAMs isolated from dissected SNS nerve bundles in sWAT or SCG (Supplementary Fig. 10b,c). We also detected the presence of NE in BAT macrophages, although at lower levels than SAMs (Supplementary Fig. 10d). The lower levels of *Slc6a2*, *Maoad*, and NE content may reflect a dilution of BAT SAMs by BAT ATMs, as mixed (as opposed to isolated) populations were analyzed.

Finally, we used *LyzM-cre; Csf1r*-LSL-*DTR* mice with conditional expression of diphtheria toxin receptor (DTR) on macrophages to test whether macrophages serve as a sink for NE. After validating ablation of macrophages (Supplementary Fig. 11a,b), we observed a significant increase of NE in sWAT of *LyzM-cre; Csf1r*-LSL-*DTR* mice relative to *Csf1r*-LSL-*DTR* controls (Supplementary Fig. 11c). Note that, owing to constant hematopoietic input, it is practically impossible to completely deplete macrophages. This limitation notwithstanding, these results are consistent with a model in which macrophages act as a sink for NE.

### Human sympathetic ganglia also contain NE-degrading SAMs

Finally, we asked whether SAMs exist in humans. We obtained nine human excisional biopsies of SNS or thoracolumbar ganglia that were collected during sympathectomy and/or gangliotomy. We stained tissue sections with H&E (Fig. 6a,b) or an antibody against CD68, a human macrophage marker, and identified the presence of macrophages in SNS tissues (Fig. 6c,d and Supplementary Fig. 12).



**Figure 6** SAMs in the human sympathetic nervous system. (a,b) Optical micrograph of human ganglia from the thoracolumbar region stained with H&E (a) and a higher-magnification image (b). (c,d) Optical micrograph of human ganglia from the thoracolumbar region stained with an antibody against CD68 (c) and a higher-magnification image (d). (e) Optical micrograph of human ganglia from the thoracolumbar region stained with antibodies against CD68 and SLC6A2. Red arrows indicate CD68+SLC6A2<sup>+</sup> regions. Pink arrows indicate SLC6A2<sup>+</sup> regions. (f) Optical micrographs of human ganglia from the thoracolumbar region stained with antibodies against CD68 and MAOA. Red arrows indicate CD68+MAOA<sup>+</sup> regions. Pink arrows indicate MAOA<sup>+</sup> regions. Boxed regions in e and f represent higher-magnification images of the main micrographs. Scale bars: 1 mm (a,c), 100 μm (b,d), and 50 μm (e,f); for boxed regions, 25 μm. Images in a–f are representative of nine different human samples.

We next determined whether SAMs in human sympathetic ganglia also contain the machinery for uptake and degradation of NE (Fig. 6e,f and Supplementary Fig. 12). The CD68 macrophage marker colocalized with staining for SLC6A2 (Fig. 6e and Supplementary Fig. 12a) and MAOA (Fig. 6f and Supplementary Fig. 12b). Both SLC6A2<sup>+</sup> and MAOA<sup>+</sup> neurons existed, but the background levels were low relative to control human gut-associated lymphoid tissue (GALT) samples that also contained CD68<sup>+</sup> macrophages (Supplementary Fig. 12c,d).

### DISCUSSION

SAMs are a previously undescribed population of SNS-resident macrophages that import and degrade NE. To fulfill their function, SAMs express a dedicated molecular machinery that is, as best we can tell, absent from neighboring macrophages and other known macrophage populations (shown by our data and the ImmGen database). In SAMs, NE is imported by SLC6A2 and degraded by MAOA. This is a specialized

molecular mechanism for NE uptake with a role that is not fulfilled by the canonical phagocytic mechanisms generally present in macrophages<sup>26</sup>.

Unlike most other neurons, which exclusively release neurotransmitter at a terminal synapse, SNS neurons also release NE via varicosities distributed along axons that can extend for tens of centimeters<sup>27</sup>. SAMs possibly serve to prevent NE spillover into the bloodstream or neighboring tissues when SNS activity is high. Indeed, we demonstrate that, when SNS neurons are optogenetically activated, SAMs import increased levels of NE and become more polarized toward a proinflammatory phenotype. In this regard, NE can be considered a noxious stimulus that must be locally delivered in a controlled manner to a target tissue. Chronic and excessive systemic NE in serum, such as the levels present under chronic stress conditions or in medullary adrenal tumors, leads to hypertension and cardiopathy due to direct action in cardiovascular tissues<sup>28</sup>.

The activated polarization state of SAMs is consistent with a model in which these cells play a tissue-protective role by acting as a sentinel and scavenger of surplus levels of an endogenous neurotransmitter (i.e., NE) that, if released in excess from varicosities, could potentially be harmful. Tissue-protective immune cells have been documented in the brain and other, non-neuronal systems<sup>29–34</sup>. For instance, muscularis-resident macrophages in the gut induce rapid tissue-protective responses to potentially pathogenic insults via  $\beta 2$  adrenergic receptor signaling<sup>35</sup>. This mechanism and our study indicate specialization of macrophage populations for fulfillment of tissue-specific tasks in response to neuronal cues. Divergent gene expression landscapes across populations of resident macrophages isolated from different tissues support the idea of local macrophage adaptations<sup>22,36,37</sup>. In this study, we use transcriptional data to molecularly characterize SAMs alongside other macrophage populations. Our results suggest that macrophages associated with the SNS have specialized molecular programs, whose exploration might give further insight into mechanisms underlying SNS macrophage–neuron communication.

Although SAMs express common microglial genes and reside in proximity to nerve cells, SAM pseudopodia are morphologically distinct from the finely branching ramifications of resting microglia<sup>38,39</sup>. Moreover, SAMs are seemingly of hematopoietic origin, as suggested by our bone marrow chimera studies and their high expression of CD45 and macrophage markers. Future tracing studies are necessary to definitively determine SAM origin. To our knowledge, no reports exist on NE uptake by microglia, and we verified that the machinery for NE uptake is not expressed in these cells. In this regard, only one study has reported that NE can trigger microglia to import and degrade amyloid but not NE itself<sup>40</sup>. Neurotransmitter uptake has primarily been studied in astroglia, which are CX3CR1<sup>+</sup> (ref. 41).

Chimeric models require irradiation that generates inflammation. However, if given adequate recovery time (8 weeks), recruited macrophages dissipate from the brain, as represented in our chimeras by minimal residual *Cx3cr1*-GFP<sup>+</sup> microglia (0.06%). SAMs persist at levels that greatly surpass background irradiation-induced macrophage recruitment, and regenerated SAMs are seemingly identical to those in non-irradiated mice.

We show low expression of several astroglial markers in SAMs, raising the possibility of a hybrid peripheral cell type that unites some of the features of macrophages and glia. Alternatively, genes expressed in common by glial cells and SAMs may be attributable to the proximity of SAMs to neuron-derived signals, analogous to the observation that microglia, astrocytes, and neurons share the expression of certain

central nervous system–specific genes<sup>7,42</sup>. An alternative model is that SAMs share the lineage of satellite glial cells (SGCs), which are derived from embryonic neural crest<sup>11</sup> and also express canonical astroglial markers<sup>43</sup>. However, SGC import or degradation of NE has not been reported<sup>44</sup>.

Our study may fill a gap in the literature by demonstrating a cellular and molecular mechanism alternative to the proposed existence of NE-producing macrophages in WAT<sup>3</sup>. In this regard, our findings are consistent with other reports<sup>4–6</sup>, as we did not detect the NE biosynthetic machinery in SAMs nor in ATMs. The identification of SAMs sheds new light on this recent controversy by documenting how a particular population of macrophages can contain NE in the absence of its biosynthesis. We also document that BAT macrophages contain similar molecular machinery to that in SAMs for NE uptake, extending and validating the findings of our colleagues<sup>21</sup>. SAMs may play a tissue-protective role through regulation of regional NE levels by serving as a local sink that prevents the dangerous effects of chronically increased levels of systemic NE.

In sharp contrast to the anti-inflammatory state of intestinal nerve-associated *Cx3cr1*-GFP<sup>+</sup> macrophages<sup>35</sup>, SAMs exhibit a proinflammatory profile at steady state. This could be due to the constitutive presence of a danger signal—namely, NE. Whether this polarization is caused by NE import or by adrenergic signaling remains to be established. In this regard, polarization of enteric-associated macrophages has been linked to activation of  $\beta 2$  adrenergic receptor, which is also expressed in SAMs<sup>35</sup>. Regardless, our core message is relevant: SAMs are proinflammatory and act as a NE sink, and blocking NE uptake has an antiobesity effect. Our results support a model whereby SAMs pathologically accumulate in the SNS nerves of obese subjects in an organ-specific manner, thus explaining why we detect SAM accumulation in the WAT-associated SNS but not in SCG, which innervates salivary glands and other neck structures. The NE-scavenging role of SAMs may have become evolutionarily maladaptive, as in the past obesity was not a common physiological stress to which humans had to adapt. In modern times, the prevalence of overnutrition has created a need for increased lipolysis-inducing NE signaling to maintain fat stores, which is obstructed by the ‘original’ function of SAMs to limit NE levels.

Reduced NE availability in the adipose tissue is linked to blunted lipolysis and obesity. Very recently, our colleagues have shown that ATMs degrade NE during aging<sup>45</sup>. Whether this observation is also associated with the accumulation of SAMs in fat, as we observed in the two mouse models of obesity, remains to be established.

Our results demonstrate that SAM-specific *Slc6a2* ablation rescues BAT and adaptive thermogenesis in obese *ob/ob* mice, which in turn leads to sustained weight loss and lipid mobilization. We determine that blocking NE import into SAMs mitigates the recidivism of obesity that is typical after dieting. Overall, our results identify SAMs as a potential new molecular and cellular target for obesity therapy.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We would like to thank the Unit for Imaging and Cytometry at the Instituto Gulbenkian de Ciência (IGC) for assistance with flow cytometry, cell sorting, and multiphoton microscopy. We also want to thank the Antibody Service at the IGC

for the antibodies produced in house and the Histopathology facility at the IGC for tissue processing and histological assessment. This work was supported by the Fundação para a Ciência e Tecnologia (FCT), the European Molecular Biology Organization (EMBO), the Human Frontier Science Program (HFSP), Maratona da Saúde, and the US National Institutes of Health (NIH). R.M.P. was supported by FCT (SFRH/BD/88454/2012), J.S.S. was supported by the American Heart Association (16PRE30980030) and a training grant (T32DK007541), B.A.A. was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and N.M.-S. was supported by Xunta de Galicia (ED481B 2016/168-0). We thank M. Aouadi for helpful discussions.

#### AUTHOR CONTRIBUTIONS

A.I.D. conceptualized the study. R.M.P. performed two-photon and confocal microscopy. E.S. and R.M.P. performed flow cytometry. J.S.S. and R.M.P. performed low-input RNA-seq. V.M.L., J.S.S., and R.M.P. analyzed the RNA-seq data. M.I., A.L.S., S.A., and E.T. performed electron microscopy. E.S., R.M.P., N.M.S., I. Mahú, B.A.A., and C.M.L. performed functional tests. N.K., I. Morris, R.M., and V.G. performed related mouse husbandry and genotyping. F.T. and M.V. processed human ganglia. M.K.H. provided the *Slc6a2*<sup>-/-</sup> mice. N.J.S. developed the low-input RNA-seq protocols. A.I.D., C.K.G., and R.M.P. wrote the original draft of the manuscript. A.I.D., C.K.G., R.M.P., and C.M.L. reviewed and edited the final version of the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Zeng, W. *et al.* Sympathetic neuro-adipose connections mediate leptin-driven lipolysis. *Cell* **163**, 84–94 (2015).
- Mathis, D. Immunological goings-on in visceral adipose tissue. *Cell Metab.* **17**, 851–859 (2013).
- Nguyen, K.D. *et al.* Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis. *Nature* **480**, 104–108 (2011).
- Fischer, K. *et al.* Alternatively activated macrophages do not synthesize catecholamines or contribute to adipose tissue adaptive thermogenesis. *Nat. Med.* **23**, 623–630 (2017).
- Spadaro, O. *et al.* IGF1 shapes macrophage activation in response to immunometabolic challenge. *Cell Rep.* **19**, 225–234 (2017).
- Reitman, M.L. How does fat transition from white to beige? *Cell Metab.* **26**, 14–16 (2017).
- Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
- Anlauf, E. & Derouiche, A. Glutamine synthetase as an astrocytic marker: its cell type and vesicle localization. *Front. Endocrinol. (Lausanne)* **4**, 144 (2013).
- Bignami, A., Eng, L.F., Dahl, D. & Uyeda, C.T. Localization of the glial fibrillary acidic protein in astrocytes by immunofluorescence. *Brain Res.* **43**, 429–435 (1972).
- Chaudhry, F.A. *et al.* Glutamate transporters in glial plasma membranes: highly differentiated localizations revealed by quantitative ultrastructural immunocytochemistry. *Neuron* **15**, 711–720 (1995).
- Jessen, K.R. & Mirsky, R. The origin and development of glial cells in peripheral nerves. *Nat. Rev. Neurosci.* **6**, 671–682 (2005).
- Ludwin, S.K., Kosek, J.C. & Eng, L.F. The topographical distribution of S-100 and GFA proteins in the adult rat brain: an immunohistochemical study using horseradish peroxidase-labelled antibodies. *J. Comp. Neurol.* **165**, 197–207 (1976).
- Mearow, K.M., Mill, J.F. & Vitkovic, L. The ontogeny and localization of glutamine synthetase gene expression in rat brain. *Brain Res. Mol. Brain Res.* **6**, 223–232 (1989).
- Raff, M.C. *et al.* Galactocerebroside is a specific cell-surface antigenic marker for oligodendrocytes in culture. *Nature* **274**, 813–816 (1978).
- Regan, M.R. *et al.* Variations in promoter activity reveal a differential expression and physiology of glutamate transporters by glia in the developing and mature CNS. *J. Neurosci.* **27**, 6607–6619 (2007).
- Rusnakova, V. *et al.* Heterogeneity of astrocytes: from development to injury—single cell gene expression. *PLoS One* **8**, e69734 (2013).
- Sensenbrenner, M., Lucas, M. & Deloume, J.C. Expression of two neuronal markers, growth-associated protein 43 and neuron-specific enolase, in rat glial cells. *J. Mol. Med. (Berl.)* **75**, 653–663 (1997).
- Buttgereit, A. *et al.* Sall1 is a transcriptional regulator defining microglia identity and function. *Nat. Immunol.* **17**, 1397–1406 (2016).
- Wentworth, J.M. *et al.* Pro-inflammatory CD11c<sup>+</sup>CD206<sup>+</sup> adipose tissue macrophages are associated with insulin resistance in human obesity. *Diabetes* **59**, 1648–1656 (2010).
- Amano, S.U. *et al.* Local proliferation of macrophages contributes to obesity-associated adipose tissue inflammation. *Cell Metab.* **19**, 162–171 (2014).
- Wolf, Y. *et al.* Brown-adipose-tissue macrophages control tissue innervation and homeostatic energy expenditure. *Nat. Immunol.* **18**, 665–674 (2017).
- Gosselin, D. *et al.* An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, eaal3222 (2017).
- Clausen, B.E., Burkhardt, C., Reith, W., Renkawitz, R. & Förster, I. Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* **8**, 265–277 (1999).
- Merad, M. *et al.* Langerhans cells renew in the skin throughout life under steady-state conditions. *Nat. Immunol.* **3**, 1135–1141 (2002).
- Shirey-Rice, J.K. *et al.* Norepinephrine transporter variant A457P knock-in mice display key features of human postural orthostatic tachycardia syndrome. *Dis. Model. Mech.* **6**, 1001–1011 (2013).
- Aderem, A. & Underhill, D.M. Mechanisms of phagocytosis in macrophages. *Annu. Rev. Immunol.* **17**, 593–623 (1999).
- Stjärne, L. Basic mechanisms and local modulation of nerve impulse-induced secretion of neurotransmitters from individual sympathetic nerve varicosities. *Rev. Physiol. Biochem. Pharmacol.* **112**, 1–137 (1989).
- Schroeder, C. & Jordan, J. Norepinephrine transporter function and human cardiovascular disease. *Am. J. Physiol. Heart Circ. Physiol.* **303**, H1273–H1282 (2012).
- Filiano, A.J. *et al.* Unexpected role of interferon- $\gamma$  in regulating neuronal connectivity and social behaviour. *Nature* **535**, 425–429 (2016).
- Galle-Treger, L. *et al.* Nicotinic acetylcholine receptor agonist attenuates ILC2-dependent airway hyperreactivity. *Nat. Commun.* **7**, 13202 (2016).
- Ibiza, S. *et al.* Glial-cell-derived neuroregulators control type 3 innate lymphoid cells and gut defence. *Nature* **535**, 440–443 (2016).
- Kipnis, J. Multifaceted interactions between adaptive immunity and the central nervous system. *Science* **353**, 766–771 (2016).
- Louveau, A. *et al.* Structural and functional features of central nervous system lymphatic vessels. *Nature* **523**, 337–341 (2015).
- Rosas-Ballina, M. *et al.* Acetylcholine-synthesizing T cells relay neural signals in a vagus nerve circuit. *Science* **334**, 98–101 (2011).
- Gabanyi, I. *et al.* Neuro-immune interactions drive tissue programming in intestinal macrophages. *Cell* **164**, 378–391 (2016).
- Gautier, E.L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* **13**, 1118–1128 (2012).
- Okabe, Y. & Medzhitov, R. Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell* **157**, 832–844 (2014).
- Crotti, A. & Ransohoff, R.M. Microglial physiology and pathophysiology: insights from genome-wide transcriptional profiling. *Immunity* **44**, 505–515 (2016).
- Prinz, M. & Priller, J. Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nat. Rev. Neurosci.* **15**, 300–312 (2014).
- Kong, Y., Ruan, L., Qian, L., Liu, X. & Le, Y. Norepinephrine promotes microglia to uptake and degrade amyloid  $\beta$  peptide through upregulation of mouse formyl peptide receptor 2 and induction of insulin-degrading enzyme. *J. Neurosci.* **30**, 11848–11857 (2010).
- Kettenmann, H. & Ransom, B.R. *Neuroglia* (Oxford University Press, 2013).
- Butovsky, O. *et al.* Identification of a unique TGF- $\beta$ -dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
- Hanani, M. Satellite glial cells in sensory ganglia: from form to function. *Brain Res. Brain Res. Rev.* **48**, 457–476 (2005).
- Hanani, M. Satellite glial cells in sympathetic and parasympathetic ganglia: in search of function. *Brain Res. Rev.* **64**, 304–327 (2010).
- Camell, C.D. *et al.* Inflammation-driven catecholamine catabolism in macrophages blunts lipolysis during ageing. *Nature* <http://dx.doi.org/10.1038/nature24022> (2017).

## ONLINE METHODS

**General experimental approaches.** No samples, mice, or data points were excluded from the reported analyses. Samples were not randomized to experimental groups. Analyses were not performed in a blinded fashion. More detailed information can be found in the **Life Sciences Reporting Summary**.

**Antibodies, stain reagents, and drugs.** Antibodies were obtained from the following vendors: anti-F4/80–Alexa Fluor 647 (BioLegend, catalog no. 123122, clone BM8), anti-human CD68 (Dako, catalog no. M 0876, clone PG-M1), anti-human NE transporter (NET) (MAb Technologies, catalog no. NET17-1, clone 3-6C1 sc H10), anti-MAOA (Abcam, catalog no. ab126751, clone GR155892-5), anti-TH (Pel-Freez Biologicals, catalog no. P40101-150, lot 16736), anti-GFP (Abcam, catalog no. ab13970, lot GR279236-1), anti-TH (Aves Lab, catalog no. TYH, lot TH1205), anti-GFP (Invitrogen, catalog no. A11120, lot 1563696), anti-GFP (Abcam, catalog no. ab6556, lot GR292567-1), goat anti-chicken IgY (H+L) secondary antibody, Alexa Fluor 488 (Molecular Probes/Thermo Fisher Scientific, catalog no. A-11039, lot 1759025), goat anti-rabbit IgG (H+L) cross-adsorbed secondary antibody, Alexa Fluor 594 (Molecular Probes/Thermo Fisher Scientific, catalog no. A-11012, lot 1704538), anti-Ly6c–eFluor 405 (eBioscience, catalog no. 48-5932-82, clone HK1.4, lot 4306743), anti-CD11c-PE (BD Pharmingen, catalog no. 553802, clone HL3, lot 47030), anti-CD45.2-PE (BioLegend, catalog no. 109808, clone 104.2), anti-CD45.2-FITC (obtained from S. Kimura (Memorial Sloan Kettering Cancer Center); clone 104.2), anti-CD11b-FITC (ATCC, catalog no. TIB-128, clone M1/70), anti-MHCII-Bio (clone M5/114, ATCC, catalog no. TIB-120), SA<sub>v</sub>-APC/Cy7 (BioLegend, catalog no. 405208, lot B215107), anti-Ki-67–Alexa488 (BD Biosciences, catalog no. 558616, clone B56, lot 7138687), IgG–Alexa488, isotype control (BD Biosciences, catalog no. 557782, lot 7102576), anti-Siglec-F-BV421 (BD Biosciences, catalog no. 562681, lot 7047598), anti-CD68 (Bio-Rad, catalog no. MCA1957GA, clone FA-11), goat anti-rat IgG (H+L) cross-adsorbed secondary antibody, Alexa Fluor 594 (Invitrogen, catalog no. A-11007), goat anti-chicken IgY (H+L), Alexa Fluor 647 (Abcam, catalog no. ab150171), goat anti-rabbit IgG (H+L), Alexa Fluor 488, (Abcam, catalog no. ab150077), goat anti-mouse IgG (H+L), Alexa Fluor 488 (Sigma, catalog no. SAB4600387), anti-mouse IgG (whole molecule), Cy3 (Sigma, catalog no. C0992), rabbit anti-UCP1 (Abcam, catalog no. ab10983, lot GR249119-8), mouse anti-NET (MAb Technologies, catalog no. NET05-2, clone 2-3 B2 sc D7), SYTOX Blue dead cell stain (Molecular Probes/Thermo Fisher Scientific, catalog no. S34857, lot 1851462) was used to exclude dead cells. HCS LipidTOX Deep Red Neutral Lipid Stain (Molecular Probes/Thermo Fisher Scientific, catalog no. H34477) and HCS LipidTOX Red Neutral Lipid Stain (Molecular Probes/Thermo Fisher Scientific, catalog no. H34476) were used to stain lipids.

Acetylcholine chloride, nisoxetine hydrochloride, clorgyline, and NE were purchased from Sigma-Aldrich.

**Mice.** *Cx3cr1<sup>GFP/+</sup>* mice (*Cx3cr1<sup>tm1Litt</sup>/Litt*); stock no. 008451), *Th-cre* mice (stock no. 008601), *GFP-L10* mice (stock no. 024750), *LysM-cre* mice (stock no. 004781), *LSL-ChR2-YFP* mice (stock no. 012-569), *LSL-tdTomato* mice (stock no. 007909), *ob/ob* mice (stock no. 000632), and *Csf1r-LSL-DTR* mice (stock no. 024046) were purchased from the Jackson Laboratory (JAX). *NET<sup>fl/p</sup> (Slc6a2<sup>-/-</sup>)* mice were kindly provided by M. Hahn (Vanderbilt University). B6 (C57BL/6J) and B6-CD45.1 mice were purchased from Charles River and were bred and maintained at Instituto Gulbenkian de Ciência. Both males and females were used in this study. Mice were 4–10 weeks old (for details, see the **Life Sciences Reporting Summary**). Animal procedures were approved by the ethics committee of Instituto Gulbenkian de Ciência.

**Immunofluorescence and confocal microscopy.** Tissues were dissected and fixed in 4% paraformaldehyde for 2 h (at room temperature (RT), with agitation). For images in **Figure 2j,k**, we employed frozen sections and the fixation step was followed by cryoprotection in 30% sucrose (Alfa Aesar). 16- $\mu$ m sections were obtained in a Leica Cryostat CM3050 S. Both frozen sections and whole-mount tissues were incubated in a blocking and permeabilization solution (3% BSA, 2% goat serum, 0.1% Tween, and 0.1% sodium azide in 1 $\times$  PBS) for 1 h at RT with (whole mounts) or without (frozen sections) agitation. Incubations with primary antibodies were performed overnight at 4 °C with (whole mounts) or without

(frozen sections) agitation. The following dilutions of primary antibodies were used: anti-GFP (1:500), anti-TH (1:1,000), anti-SLC6A2 (1:500), anti-MAOA (1:100). Incubation with secondary antibodies was performed for 1–2 h at RT, with or without (in the case of frozen sections) agitation. Z-series stacks were acquired on a Leica TCS SP5 confocal inverted microscope. Analysis and quantification of images were performed in Fiji.

**In vivo two-photon microscopy.** Mice aged 2 months were kept anesthetized with 2% isoflurane. During surgery, body temperature was maintained at 37 °C with a warming pad. After application of local anesthetics (lidocaine), a sagittal incision of the skin was made above the suprapelvic flank to expose the subcutaneous inguinal fat pad. An imaging chamber was custom-built to minimize fat movement. Warm imaging solution (in mM: 130 NaCl, 3 KCl, 2.5 CaCl<sub>2</sub>, 0.6 MgCl<sub>2</sub>·6H<sub>2</sub>O, 10 HEPES without sodium, 1.2 NaHCO<sub>3</sub>, 10 glucose, pH 7.45 with NaOH) (37 °C) mixed with a fat dye (LipidTOX) was applied to label adipocytes, maintain tissue integrity, and allow the use of an immersion objective. Imaging experiments were performed under a two-photon laser-scanning microscope (Ultima, Prairie Instruments). Live images were acquired at 8–12 frames per second at depths below the surface ranging from 100–250  $\mu$ m, using an Olympus 20 $\times$  1.0 N.A. water-immersion objective with a laser tuned to a wavelength of 810–940 nm and emission filters at 525/50 nm and 595/50 nm for green and red fluorescence, respectively. Laser power was adjusted to be 20–25 mW at the focal plane (maximally 35 mW), depending on the imaging depth and levels of GFP expression and LipidTOX spread. Analysis and quantification of images were performed in Fiji.

**Electron microscopy.** Fresh tissue was perfused with 2% paraformaldehyde (Electron Microscopy Services (EMS)) and 0.2% glutaraldehyde (EMS) in 0.1 M phosphate buffer (PB; pH 7.4). After perfusion, fibers were isolated and immersion fixed for 2 h at RT in the same fixative. For quenching of autofluorescence from free aldehydes, nerves were washed with 0.15% glycine (VWR) in PB for 10 min at RT.

**Correlative light–electron microscopy.** After fixation, fibers were stabilized with 0.1% tannic acid (EMS) and embedded in 2% agarose (Omnipur) before cryoprotection in 30% sucrose (Alfa Aesar) overnight at 4 °C. Embedded samples were placed in optimal cutting temperature (OCT) compound (Sakura) and plunge frozen in liquid nitrogen. 10- $\mu$ m sections were obtained in a Leica Cryostat CM3050 S and placed on cover glasses coated with 2% (3-aminopropyl) triethoxysilane (Sigma-Aldrich) in acetone. Light microscopy imaging was performed in a Leica SP5 Live microscope after mounting the sections with PB. For electron microscopy processing, samples were washed ten times with PB and post-fixed in 1% osmium tetroxide (EMS) with 1% potassium hexacyanoferrate (Sigma-Aldrich) in PB for 30 min on ice. Dehydration was done in a graded ethanol series of 30%, 50%, 75%, 90%, and 100% ethanol for 10 min each. EPON resin (EMS) was used for embedding. 70-nm serial sections were obtained in a Leica EM UC7 and stained with 1% uranyl acetate and lead citrate for 5 min each. Electron microscopy images were acquired on a Hitachi H-7650 operating at 100 kV.

**Single-cell suspension.** Tissues were dissected from ten mice. Spleen, brain, visceral fat, and subcutaneous fat were excised and digested for 30 min with collagenase (Sigma) at 37 °C with shaking. Sympathetic nerve fibers were isolated from subcutaneous adipose tissues and digested for 30 min with hyaluronidase (Sigma) at 37 °C with shaking, washed, and further digested with collagenase for 15 min. SCGs were dissected and digested with collagenase for 10 min, washed, and further digested with trypsin (Biowest) for 30 min at 37 °C with shaking. Cell suspensions were filtered through a 70- $\mu$ m sieve and centrifuged at 450g for 5 min.

**Flow cytometry.** Flow cytometry data were acquired on an LSR Fortessa X-20 SORP (Becton–Dickinson), FACSCalibur (Becton–Dickinson), or CyAn ADP (Beckman Coulter) and analyzed using the FlowJo software package (Tree Star). Macrophages were sorted as live CD45 and F4/80 double-positive cells using a FACSAria IIu high-speed cell sorter (Becton Dickinson) or MoFlo High-Speed Cell Sorter produced by Dako Cytomation (now owned by Beckman Coulter).

**Bone marrow chimeras.** B6-CD45.1 mice (aged 8–10 weeks), B6 (C57BL/6J) mice (aged 8–10 weeks), and *ob/ob* mice (aged 8–10 weeks) were lethally irradiated (900 rad, 3.42 min,  $^{137}\text{Cs}$  source) (Gammacell 2000) and reconstituted with bone marrow cells from *Cx3cr1<sup>GFP/+</sup>* mice (aged 6 weeks), *Slc6a2<sup>-/-</sup>* mice (aged 6–8 weeks), B6 mice (aged 6–8 weeks), or B6-CD45.1 mice (aged 6–8 weeks). B6-CD45.1 mice and B6 mice were reconstituted with  $5 \times 10^6$  total bone marrow cells, and *ob/ob* mice were reconstituted with  $3 \times 10^7$  total bone marrow cells. Chimerism was assessed 8 weeks after reconstitution using flow cytometry.

**Low-input RNA-seq library preparation.** Sequencing libraries were prepared according to the Smart-seq2 method<sup>46</sup> with some modifications.  $1,715 \pm 115$  cells from nerve fibers,  $1,534 \pm 85$  cells from superior cervical ganglia, and 5,000 cells from other tissues (visceral fat, subcutaneous fat, spleen, and brain) were isolated as live CD45<sup>+</sup>F4/80<sup>+</sup> cells in TRIzol (Thermo Fisher) and were used as starting material. RNA was extracted with the Direct-zol MicroPrep kit (Zymo Research) with on-column DNase I treatment. 10  $\mu\text{l}$  of purified RNA was mixed with 5.5  $\mu\text{l}$  of SMARTScribe 5 $\times$  First-Strand Buffer (Clontech), 1  $\mu\text{l}$  of poly(T) primer for reverse transcription (2.5  $\mu\text{M}$ ; 5'-AAGCAGTGGTATCAACGCAGAGTAC(T<sub>30</sub>)VN-3'), 0.5  $\mu\text{l}$  of SUPERase IN (Ambion), 4  $\mu\text{l}$  of dNTP mix (10 mM; Invitrogen), 0.5  $\mu\text{l}$  of dithiothreitol (DTT) (20 mM; Clontech), and 2  $\mu\text{l}$  of betaine solution (5 M; Sigma), and samples were incubated at 50 °C for 3 min. 3.9  $\mu\text{l}$  of first-strand mix, containing 0.2  $\mu\text{l}$  of 1% Tween-20, 0.32  $\mu\text{l}$  of MgCl<sub>2</sub> (500 mM), 0.88  $\mu\text{l}$  of betaine solution (5 M; Sigma), 0.5  $\mu\text{l}$  of SUPERase IN (Ambion), and 2  $\mu\text{l}$  of SMARTScribe Reverse Transcriptase (100 U/ $\mu\text{l}$ ; Clontech), was added, and samples were incubated with one step at 25 °C for 3 min and one step at 42 °C for 60 min. 1.62  $\mu\text{l}$  of template-switch (TS) reaction mix containing 0.8  $\mu\text{l}$  of biotin-TS oligonucleotide (10  $\mu\text{M}$ ; biotin-5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3'), 0.5  $\mu\text{l}$  of SMARTScribe Reverse Transcriptase (100 U/ $\mu\text{l}$ ; Clontech), and 0.32  $\mu\text{l}$  of SMARTScribe 5 $\times$  First-Strand Buffer (Clontech) was added, and samples were then incubated at 50 °C for 2 min, 42 °C for 80 min, and 70 °C for 10 min. 14.8  $\mu\text{l}$  of second-strand synthesis, preamplification mix containing 1  $\mu\text{l}$  of preamplification oligonucleotide (10  $\mu\text{M}$ ; 5'-AAGCAGTGGTATCAACGCAGAGT-3'), 8.8  $\mu\text{l}$  of KAPA HiFi Fidelity Buffer (5 $\times$ ; KAPA Biosystems), 3.5  $\mu\text{l}$  of dNTP mix (10 mM; Invitrogen), and 1.5  $\mu\text{l}$  of KAPA HiFi HotStart DNA Polymerase (1U/ $\mu\text{l}$ ; KAPA Biosystems) was added, and samples were amplified by PCR: 95 °C for 3 min, 8 cycles at 98 °C for 20 s, 67 °C for 15 s, and 72 °C for 6 min, and a final extension step at 72 °C for 5 min. The synthesized double-stranded DNA (dsDNA) was purified using Sera-Mag SpeedBeads (Thermo Fisher Scientific) with final concentrations of 8.4% PEG 8000 and 1.1 M NaCl and then was eluted with 13  $\mu\text{l}$  of UltraPure water (Invitrogen). The product was quantified by Qubit dsDNA High Sensitivity Assay Kit (Invitrogen), and libraries were prepared using the Nextera DNA Sample Preparation Kit (Illumina). Tagmentation mix containing 11  $\mu\text{l}$  of 2 $\times$  Tagment DNA Buffer and 1  $\mu\text{l}$  of Tagment DNA Enzyme was added to 10  $\mu\text{l}$  of purified DNA, and samples were then incubated at 55 °C for 15 min. 6  $\mu\text{l}$  of Nextera Resuspension Buffer (Illumina) was added, and samples were incubated at RT for 5 min. Tagmented DNA was purified using Sera-Mag SpeedBeads (Thermo Fisher Scientific) with final concentrations of 7.8% PEG 8000 and 0.98 M NaCl and then eluted with 25  $\mu\text{l}$  of UltraPure water (Invitrogen). Final enrichment amplification was performed with Nextera primers, adding 1  $\mu\text{l}$  of Index 1 primers (100  $\mu\text{M}$ ; N7xx), 1  $\mu\text{l}$  of Index 2 primers (100  $\mu\text{M}$ ; N5xx) and 27  $\mu\text{l}$  of NEBNext High-Fidelity 2 $\times$  PCR Master Mix (New England BioLabs) and then amplifying samples by PCR: 72 °C for 5 min, 98 °C for 30 s, and 8–13 cycles of 98 °C for 10 s, 63 °C for 30 s, and 72 °C for 1 min. Libraries were size selected, quantified by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), and pooled and sequenced on a NextSeq 500 (Illumina) for 76 cycles at a depth of 25 to 30 million single-end reads per sample. To normalize for genomic DNA contamination, which occurred in some samples due to incomplete DNA removal during RNA isolation, the average intronic noise per base pair in all intronic regions per gene was calculated. The exonic reads were then normalized by subtracting the background noise per base pair for the complete length of the exonic regions. Genes without introns were not normalized, as these genes are the minority of genes and are typically short (code available at [https://github.com/vlink/DNA\\_contamination/](https://github.com/vlink/DNA_contamination/)).

Fastq files from sequencing experiments were mapped to the mouse mm10 genome using default parameters for STAR<sup>47</sup>. Mapped data were analyzed with HOMER<sup>48</sup> and custom R and Perl scripts.

**Superior cervical ganglion explant cultures.** SCGs were removed from mice aged 4–6 weeks under a stereomicroscope and placed in DMEM (Invitrogen, Carlsbad, CA, USA). Ganglia were cleaned from the surrounding tissue capsule and transferred into eight-well tissue culture chambers (Sarstedt, Nümbrecht, Germany) that were previously coated with poly-D-lysine (Sigma-Aldrich, Steinheim, Germany) in accordance with the manufacturer's instructions. Ganglia were then covered with 5  $\mu\text{l}$  of Matrigel (BD Bioscience, San Jose, CA, USA) and incubated for 7 min at 37 °C. DMEM without phenol red (Invitrogen) supplemented with 10% FBS (Invitrogen), 2 mM L-glutamine (Biowest, Nuaille, France), and nerve growth factor (Sigma-Aldrich) was subsequently added. 12 SCG explant cultures were prepared per condition. SCGs were cultured for a minimum of 24 h before further manipulation. The stimulation protocol in Figure 3 was performed for 2 h with the following concentrations of drugs: 10 mM acetylcholine chloride, 100 nM nisoxetine hydrochloride, and 100  $\mu\text{M}$  clorgylene.

**NE measurements after optogenetic stimulation *ex vivo*.** Depolarization of sympathetic neurons in *Th-cre*; *LSL-Chr2-YFP* explant cultures was performed on a Yokogawa CSU-X Spinning Disk confocal microscope using the 488-nm laser line pointing at the region of interest (ROI) for 200  $\mu\text{s}$ . Stimulation was repeated seven times using 40% laser intensity. NE content in the SCG explant culture medium and in sorted CD45<sup>+</sup>F4/80<sup>+</sup> cells was determined by NE ELISA kit (Labor Diagnostika Nord, Nordhorn, Germany, catalog no. BA E-5200). The same procedure was performed for *LSL-Chr2-YFP* control mice.

**NE measurements in macrophages from sWAT.** CD45.2 (PE)<sup>+</sup>F4/80 (Alexa Fluor 647)<sup>+</sup> cells from sWAT were sorted as live cells and incubated with 2  $\mu\text{M}$  NE for 2 h using the same culture conditions as those used for SCG explant cultures. Afterwards, cells were washed twice with 1 $\times$  PBS, and NE content was measured by NE ELISA kit (Labor Diagnostika Nord, Nordhorn, Germany, catalog no. BA E-5200).

**qPCR.** Total RNA from sorted cells was isolated using the RNeasy Plus Micro Kit (Qiagen, catalog no. 50974034). Total RNA from adipose tissues was isolated with the PureLink RNA Mini Kit (Ambion, Life Technologies, catalog no. 12183025). cDNA was reverse transcribed using SuperScript II (Invitrogen) and random primers (Invitrogen). qPCR was performed using SYBR Green (Applied Biosystems) in ABI QuantStudio 7 (Applied Biosystems). The *Gapdh* housekeeping gene was used to normalize samples. We used the following formula to calculate relative expression levels:  $\text{RQ} = 2^{-\Delta\text{Ct}} \times 100 = 2^{-(\text{Ct gene of interest} - \text{Ct Gapdh})} \times 100$ .

The primers used were as follows: *Lpl*-forward, 5'-CAGCTGGGCCTAACTTGAG-3'; *Lpl*-reverse, 5'-CCTCTCTGCAATCACACGAA-3'; *Pnpla2*-forward, 5'-CACTTTAGCTCCAAGGATGA-3'; *Pnpla2*-reverse, 5'-TGGTTTCAGTAGGCCATTCCT-3'; *Gfap*-forward, 5'-CCAGCTTCGAGCCAAGGA-3'; *Gfap*-reverse, 5'-GAAGCTCCGCTGGTAGACA-3'; *Gap43*-forward, 5'-AGCC AAGGAGGAGCCTAAAC-3'; *Gap43*-reverse, 5'-CTGTCCGGCACTTTCC TTAG-3'; *Ucp1*-forward, 5'-GTGAAGGTCAGAATGCAAGC-3'; *Ucp1*-reverse, 5'-AGGGCCCCCTTCATGAGGTC-3'; *Slc6a2*-forward, 5'-CAGGCACCTCCATTCTGTTT-3'; *Slc6a2*-reverse, 5'-GCGGCTTGAAGTTGATGATGCTG-3'; *Maoa*-forward, 5'-GCCAGTA TCACAGGCCAC 3'; *Maoa*-reverse, 5'-GTCCCACATAAGCTCCACCA-3'; *Chrm1*-forward, 5'-CATCCCAACATCACCGTCTT-3'; *Chrm1*-reverse, 5'-GAGAACGAAGGAAACCAACCAC-3'; *Chrm2*-forward, 5'-TGTCTCCAGTCTAGTGCAAGG-3'; *Chrm2*-reverse, 5'-CATTCTGA CCTGACGATCCAAC-3'; *Chrm4*-forward, 5'-GCCTTCATCC TCACCTGGAC-3'; *Chrm4*-reverse, 5'-AGTGGCATTGCAGAGTGCAT-3'; *Chrm5*-forward, 5'-CCA TGGACTGTGGGAAGTCA-3'; *Chrm5*-reverse, 5'-CAGCGTCC CATGAGGATGA-3'; *Chrna2*-forward, 5'-CTCCCATCCTGCTTTCCAG-3'; *Chrna2*-reverse, 5'-GTTTGAACAGCGGCTCCTC-3'; *Chrna3*-forward, 5'-GCGAACAGGTCACAGTTTATG-3'; *Chrna3*-reverse, 5'-GCATTTT TCTCTGGGTTTTCA-3'; *Chrna5*-forward, 5'-CGCTCTCTTCCACACACAA-3'; *Chrna5*-reverse, 5'-TAGGTCCACCGTCTTTCTCG-3'; *Chrna6*-forward, 5'-CTTTGTACAGCTGTCCAT-3'; *Chrna6*-reverse, 5'-GCCTCCT TTGTCTTGTC-3'; *Chrna7*-forward, 5'-ACAGTACTTCGCCAGCACCA-3'; *Chrna7*-reverse, 5'-AAACCATGCACACCAATTCA-3'; *Chrna9*-forward, 5'-ACAAGGCCACCAACTCCA-3'; *Chrna9*-reverse, 5'-ACCAACCCACTCCTCCTCTT-3'; *Chrna10*-forward, 5'-TCTGACCTCA

CAACCCACAA-3'; *Chrna10*-reverse, 5'-TCC TGTCTCAGCCTCCATGT-3'; *Chrnb2*-forward, 5'-GGGCAGGCA CACTATTCTTC-3'; *Chrnb2*-reverse, 5'-TCCAATCCTCCCTCACACTC-3'; *Chrnb3*-forward, 5'-CTCCTCAGACATT GGTTCACAGG-3'; *Chrnb3*-reverse, 5'-AATGAGG TCAACCATGGT-3'; *Chrnb4*-forward, 5'-TCTGGTTGCCTGACATCGTG-3'; *Chrnb4*-reverse, 5'-GGTTACACAAAGTACATGGA-3'; *Adrb2*-forward, 5'-GGTTATCGTCCTGG CCATCGTGTGG-3'; *Adrb2*-reverse, 5'-TGTTTCGTGAAGAAGTCACAGC AAGTCTC-3'; *Th*-forward, 5'-GGTATACGCCACGCTGAAGG-3'; *Th*-reverse, 5'-TAGCCACAGTACCGTTCAGG-3'; *Tnfa*-forward, 5'-ATGAG CACAGA AAGCATGATC-3'; *Tnfa*-reverse, 5'-TACAGGCTTGCTACTCGAATT-3'; *Il10*-forward, 5'-GCTCTTACTGACTGGCATGAG-3'; *Il10*-reverse, 5'-CGCAG CTCTAG GAGCATGTG-3'; *Il1*-forward, 5'-GAAGAAGAGCCCATCCT CTG-3'; *Il1*-reverse, 5'-TCATCTCGGAGCCTGTAGTG-3'; *Il4ra*-forward, 5'-TGACCTCACAGGAACCCAGGC-3'; *Il4ra*-reverse, 5'-GAACAGGC AAAACAACGGGAT-3'; *Gapdh*-forward, 5'-AACTTTGGCATTGTGGA AGG-3'; *Gapdh*-reverse, 5'-ACACATTGGGGGTAGGAACA-3'.

**Functional studies.** We measured rectal temperature with an electronic thermometer (Precision) when the mice were housed both at RT and at 4 °C with ND food and water *ad libitum*.

Free fatty acids were measured in blood plasma using the Free Fatty Acid Quantitation Kit (Sigma-Aldrich, catalog no. MAK044-1KT).

Serum levels of NE were determined by NE ELISA kit (Labor Diagnostika Nord, Nordhorn, Germany, catalog no. BA E-5200).

**High-fat diet challenge.** When B6 mice reached 8 weeks of age, we replaced ND with HFD (Ssniff, Spezialdiäten, Soest, Germany, catalog no. D12492). Analyses in **Figure 4** were performed when mice achieved a 40% increase in body weight after 3 months of a HFD.

**Intracellular staining for Ki-67.** Cells were surface-stained for 30 min. Subsequently, cells were washed and fixed with fixation and permeabilization buffer (eBioscience) and then permeabilized with permeabilization buffer (eBioscience). Following this process, cells were intracellularly stained with anti-Ki-67 or isotype control.

**Histopathological and immunohistochemical analyses.** Human and mouse tissues were fixed in buffered formalin, and inclusion in paraffin was done according to standard technical procedures. Histochemical and immunohistochemical studies were performed on formalin-fixed and paraffin-embedded (FFPE) tissue sections. Sections were 2 μm (human ganglia) or 3–6 μm (mouse tissues) thick for H&E or were 4 μm thick for immunohistochemical studies. The following markers were used for immunohistochemistry: aminoethylcarbazole (AEC) and 3,3'-diaminobenzidine (DAB) according to the usual technical procedure for

the marker. For the immunohistochemical studies, sections underwent antigenic recovery before incubation with primary antibodies: anti-CD68 (Dako, clone PG-M1; dilution 1:150) anti-human SLC6A2 (Mab Technologies, clone 3-6C1 sc H10; dilution 1:1,000), anti-MAOA (Abcam, clone GR155892-5; dilution 1:50), and anti-UCP1 (Abcam; dilution 1:500). Human tissues were analyzed under an optical microscope (Nikon Eclipse 50i), and iconography microscopic images were captured using a coupled digital camera (DS Camera Control Unit DS-L2). Mouse tissues were analyzed using a Leica DM LB2 microscope, and images were captured with a Leica DFC 250 camera.

**Diphtheria toxin-mediated macrophage depletion.** We used *LysM-cre*; *LSL-Csf1r-DTR* mice for this experiment and *LSL-Csf1r-DTR* mice as controls. Animals received injections of diphtheria toxin (DT) from *Corynebacterium diphtheria* (Calbiochem) once daily for four consecutive days. The first dose was 500 ng of DT in PBS per 20 g of body weight followed by three doses of 250 ng of DT in PBS per 20 g of body weight. Depletion was assessed by flow cytometry 12 h after the fourth injection. NE levels in adipose tissues were assayed by NE ELISA kit (Labor Diagnostika Nord, Nordhorn, Germany, catalog no. BA E-5200). Protein concentration was determined by the Bradford method.

**Statistics.** Statistical analyses were performed with GraphPad Prism software (San Diego, CA) using unpaired Student's *t*-test (two-tailed) when two groups were being compared or one-way ANOVA when several groups were being compared. One-way ANOVA was followed by Tukey's multiple-comparisons test, except for the data in **Figure 4b** and **Supplementary Figure 8a**, where it was followed by Bonferroni multiple-comparisons test with one group indicated as a control group.  $P < 0.05$  was considered statistically significant. Data are represented as mean ± s.e.m. Sample size was predetermined based on previous studies (for more information, see the **Life Sciences Reporting Summary**). Data displayed normal variance.

**Data availability.** The RNA-seq data sets are available at [GSE103847](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103847). The data that support the findings herein presented are available from the corresponding author upon reasonable request. A **Life Sciences Reporting Summary** is available.

46. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
47. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
48. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).



# Discussion

In this chapter, I will discuss the findings of chapters 1 to 5 and answer the overarching question of my thesis how the enhancer landscape in macrophages get selected and regulated in order to control gene expression. To answer this general question, I addressed three main points:

- What are the epigenetic mechanisms that are responsible for tissue-specific functions?
- How do complex stimuli change the epigenetic landscape of macrophages in comparison to single stimuli?
- How does natural genetic variation influence the epigenetic landscape and gene expression in macrophages?

Furthermore, I will put the answer of these questions in the bigger context of current scientific discoveries, as well as point out new challenges and questions based on these findings. Finally, I will give a short overall conclusion.

## Epigenetic mechanisms for tissue-specific functions

As discussed in the General Introduction, enhancers are bound by lineage-determining transcription factors (LDTFs), as well as signal-dependent transcription factors (SDTFs), and play an important role in the regulation of gene expression. It has been shown that PU.1 in combination with C/EBP establishes macrophage-specific enhancers, whereas PU.1 in combination with E2A and OCT-2 establishes enhancers specific to B cells (Heinz et al. (2010)). This provides a model for how cell type-specific enhancers are established. However, it does not account for the influence of tissue-specific environments on the enhancer landscape. We studied macrophages from different tissues to address this question (Gosselin, D., Link, V. M., Romanoski, C. E. et al. (2014)) (Chapter 1). We found an astonishing amount of gene expression differences in macrophages from very different environments of the body (i.e. the brain versus the peritoneal cavity). Different macrophage populations from the same biological environment show more similar gene expression profiles, but still experience differences. By evaluating the signature of primed and active enhancers in these different macrophage populations, we find a mostly shared histone modification profile at promoters, but vastly different signatures at enhancers. The histone modification signature at enhancers correlates well with the expression of the nearest

gene, confirming that enhancers are the main drivers of cell type-specific gene expression. Motif enrichment analysis of cell type-specific enhancers showed a general enrichment of the macrophage-specific transcription factor (TF) PU.1. However, the collaborative factors in the cells from different environments varied substantially (e.g. an enrichment of the GATA consensus binding motif in macrophages from the peritoneal cavity and an enrichment for myocyte enhancer factor-2 (MEF2) and mothers against decapentaplegic homolog (SMAD) consensus motifs for microglia – macrophages isolated from the brain). These vastly different collaborative binding partners for PU.1 were conserved in different mouse strains. To address the influence of the environment in more detail, cells were taken out of the brain and the peritoneal cavity and cultured for 7 days. Interestingly, the majority of cell type-specific genes were significantly downregulated after 7 days in culture, and many cell type-specific enhancers were lost. Culturing the cells for 7 days with supplemental transforming growth factor beta (TGF $\beta$ ) and retinoic acid to mimic an *in vivo* environment more closely restored gene expression of about half of the environment-dependent genes. This shows that the environment has a big influence on gene expression of different macrophage subpopulations. Furthermore, the environment the cells are exposed to is a complex combination of different signals and we have not yet managed to appropriately mimic it *in vitro*.

This study was co-published with a complementary study by Lavin et al. (2014). The authors independently confirmed the importance of MEF2 and SMAD for the enhancer landscape of microglia, as well as the importance of GATA for macrophages from the peritoneal cavity. Furthermore, the authors included monocytes and neutrophils in their analysis and showed that even between different cell types, the histone modifications at promoters are strikingly similar, but the patterns at enhancers are widely different. The authors used a different approach to evaluate the importance of the tissue environment on the enhancer landscape than we did. Instead of culturing the cells for 7 days, they depleted all immune cells from CD45.2 host mice and injected CD45.1 donor cells that allows tracking of the donor cells. After four months, the authors isolated macrophages from the lung, brain, kidney and peritoneal cavity and assessed the enhancer landscape of CD45.1 cells. They showed a striking similarity between the enhancer landscape of the original macrophages and the enhancer landscape of the CD45.1 macrophages. To further test the influence of the environment, the authors isolated macrophages from the peritoneal cavity and transferred them into the lung of a host mouse. After 15 days of differentiation the authors isolated the implanted macrophages and compared them to original lung macrophages. Interestingly, they adapted an enhancer landscape resembling the landscape of cells from this environment more closely than their original landscape. This shows that macrophages are not only influenced by their environment during differentiation, but they keep their plasticity and are able to adapt to their new host tissue.

These two papers showed the influence of the tissue-environment on the epigenetic landscape of macrophages independently. This work underlined the need for a deeper understanding of the influence of the tissue environment on cells.

Another study by Roberts et al. (2017) studied the influence of the tissue environment on the ability of macrophages to clear apoptotic cells. They found several populations of

macrophages including peritoneal macrophages and lung macrophages were able to engulf apoptotic cells, but lose this ability after removal from the tissue environment and putting them in culture. Cultured bone marrow–derived macrophages (BMDMs) that were implanted in either the peritoneal cavity or lung gained the ability to clear cells after several weeks in the tissue environment. Furthermore, the authors showed that the TFs Krüppel-like Factor (KLF) 2 and KLF 4 were involved in this process. We (Gosselin, D., Link, V. M., Romanoski, C. E. et al. (2014)), as well as Lavin et al. (2014) found KLF to be a TF important for macrophages from the peritoneal cavity.

The influence of the tissue environment was further shown by studying the transformation of circulating monocytes into Kupffer cells after niche availability (Scott et al. (2016)). After cell–specific depletion of Kupffer cells, the liver is repopulated by circulating monocytes about 48 hours after depletion. After 15 days in the liver, the infiltrated monocytes were not distinguishable from embryonic derived Kupffer cells by fluorescence activated cell sorting (FACS) and the most Kupffer cell–specific genes were expressed in the infiltrating monocytes. They also gained the capacity to self–renew, something that was believed to require genetic engineering of monocytes. This work is another powerful example of the influence of the tissue–environment on the gene expression profile of cells.

Interestingly, when we studied macrophages at sympathetic nerve bundles we found a previously undescribed population of sympathetic neuron-associated macrophages (SAMs) (Pirzgalska et al. (2017)). These macrophages populate a unique niche in the body and possess the ability to uptake norepinephrine, as well as synthesize it by expression of *solute carrier family 6 member 2 (Slc6a2)* and *monoamine oxidase A (MAOa)*. The expression of these two genes was shown by RNA–seq, as well as quantitative polymerase chain reaction (qPCR), and neither *Slc6a2* nor *MAOa* were substantially expressed in any macrophage population listed in the ImmGen database (Heng et al. (2008)). It is not clear which environmental signals produce the expression of this machinery, but this small population of macrophages is another impressive example of how the environment can influence phenotypes of cells.

The origin of most tissue–resident macrophages is not completely clear. Studies showed that most macrophages are derived from erythromyeloid progenitors (EMPs) directly from the yolk sac (Perdiguero et al. (2015)), but other studies showed evidence that tissue–resident macrophages are derived from classical hematopoietic stem cells with the exception of microglia and epidermal Langerhans cells (Sheng et al. (2015)). Although there is some controversy about the origin of tissue–resident macrophages, Mass et al. (2016) reported that EMPs are detected around embryonic day 8.5, start populating the liver at embryonic day 9.5 and populate the rest of the embryo by embryonic day 10.25. During this time all cells express a common macrophage–core program that is different from other cell types populating the embryo. Genes upregulated during this phase include *colony stimulating factor 1 receptor (Csf1r)*, *Maf*, *Basic leucine zipper transcriptional factor ATF–like (Batf)3*, and *PPAR $\gamma$* . At the same time, macrophages populating different tissues in the embryo start to lose expression of tissue–specific macrophage genes (e.g. *T–cell immunoglobulin and mucin domain containing 4 (Timd4)* expression is lost in all sub–populations except for Kupffer cells), as well as upregulate their respective signature genes (e.g. *spalt like*

*transcription factor (Sall)1* and *Sall3* in microglia, *nuclear receptor subfamily 1, group H, member 3 (Nr1h3)*, *DNA-binding protein inhibitor (Id)1*, and *Id3* in macrophages populating the liver and *aryl hydrocarbon receptor (Ahr)* in macrophages in the limb). This study suggests that tissue-specificity is established early on in embryogenesis, but it is based on a common macrophage core expression profile, explaining the high plasticity of macrophages even after fully differentiating.

A great model to study the influence of the changing tissue-environment over different developmental stages are microglia, because the brain-blood barrier is formed early in development. From this point on, there is no infiltration of other monocytes into the brain and the only immune cells are microglia. Matcovitch-Natan et al. (2016) studied the change in microglia during development of the brain and how these cells adapt to changes in environment. They showed that microglia undergo three main temporal developmental stages with different TFs and genes expressed at every stage (e.g. genes associated with cell cycle and differentiation were expressed in the first stage, whereas genes involved in neuronal development were expressed in a later stage). The authors also showed that the dynamics of the epigenetic landscape in microglia overlapped well with microglia gene expression across developmental stages. Furthermore, they showed that disturbance of microglia development due to viral infection shifts the cell toward a more advanced developmental stage, potentially explaining the influence of viral infections on brain development. This work suggested that not only the tissue-environment itself is influencing the enhancer landscape, but also changes during development play important roles.

Adam et al. (2015) showed that super enhancers (SEs) in transitamplifying cells (TACs) are different than SEs in hair follicle stem cells (HFSs), with TAC-specific genes exhibiting TAC-specific SEs and vice versa. They also found an enrichment of TAC-specific and HFS-specific TF binding motifs in their respective SEs. Furthermore, SEs specific to HFS genes were lost *in vitro*, but were restored after engraftment of these cells into donor mice. Interestingly, the authors also found new SEs *in vitro* that were close to genes implicated in wound healing, pointing out the plasticity of cells.

Many studies have focused on the influence of the environment on the epigenetic landscape in mice. However, to ultimately understand the influence of the tissue environment on human disease, it is necessary to study human cells directly. With the decrease of sequencing cost and the lower requirement of cells for next-generation sequencing assays, these studies became more feasible to do.

Much of the research focuses on the role of microglia in neurodegenerate diseases. One of the first studies to evaluate the environment on the TF network in human microglia was performed by Gosselin et al. (2017). The authors reported gene expression data from healthy brain tissue of 19 individuals including different sexes, ages and disease diagnosis. They identified a common mRNA profile in human microglia. Many of these genes have been previously associated with neurodegenerative diseases. They identified consensus binding motifs in microglia enhancers, confirming the previously reported roles of MEF2 and SMAD in mouse microglia biology. However, they also showed substantial differences in the expression of many TFs between mouse and human, most notably the lack of expression of *SALL2*, *SALL3*, as well as *SMAD1* in human microglia. By culturing human

microglia for 7 days, they observed a similar effect on the epigenetic landscape, as well as the transcriptome, than previously observed in mice (Gosselin, D., Link, V. M., Romanoski, C. E. et al. (2014)). Interestingly, many genes downregulated in culture were upregulated during microglia development. Furthermore, the expression of about half of the genes that were highly expressed in microglia and also associated with neurodegenerative disease were impacted by a transition of the cell from the *in vivo* environment into culture.

We published one of the first studies (Gosselin, D., Link, V. M., Romanoski, C. E. et al. (2014)) demonstrating the importance of the tissue environment for the regulation of the epigenetic landscape of cells and subsequently the regulation of gene expression. This study was an important first step to realize that studying cells in culture is not sufficient in order to understand regulation *in vivo*. More studies have been published since that (e.g. Roberts et al. (2017), Scott et al. (2016), Pirzgalska et al. (2017)), underlining the importance of the tissue environment, as well as demonstrating the plasticity of cells and their ability to adapt to new environments. These initial studies were performed in mice and built an important first step in deciphering the influence of the tissue environment on gene regulation. In recent years studies in human became more feasible and are now seen as crucial tools in order to understand the interplay between tissue environment, regulation of gene expression and disease.

## The influence of complex stimuli on the epigenetic landscape in macrophages in comparison to single stimuli

Macrophages are innate immune cells and respond to diverse inflammatory signals, as well as damage-associated molecular patterns (DAMP) or microbe-associated molecular patterns (MAMP). In order to understand the regulation of macrophages after stimulation, studies often utilize bone marrow-derived macrophages (BMDMs), or thioglycollate-elicited peritoneal macrophages (TGEMs) and stimulate them with diverse inflammatory signals (e.g. TLR4 agonists, IFN $\gamma$ , TNF) or alternate activation signals like IL-4 or IL-13. However, very little is known about *in vivo* responses to stimuli, where cells experience more than one signal at a time. This is challenging to study, because of the low numbers of cells that experience the signal. To simulate a more closely assembling *in vivo* environment, we (Eichenfield, D. Z., Troutman, D. T., Link, V. M. et al. (2016)) stimulated BMDMs with homogenized skin (tissue homogenate) to mimic a wound environment (Chapter 2). Comparing the transcriptome of BMDMs stimulated with tissue homogenate showed marked differences from BMDMs stimulated with single stimuli like polyinosinic-polycytidylic acid (Poly I:C) for activation of TLR3, Kdo2-lipid A (KLA) for activation of TLR4, Pam3CSK4 - a synthetic triacylated lipopeptide (Pam3) for activation of TLR1 and TLR2, or a co-activation with KLA and IFN $\gamma$  to induce pro-inflammatory gene signatures, as well as IL-4 or TGF $\beta$  to achieve alternately activated and de-activated gene profiles. Cells from *in vivo* wounds were extracted and RNA-seq was performed. The transcriptome profile of cells from the wound overlapped mostly with the transcriptome

profile of BMDMs stimulated with tissue homogenate, showing that tissue homogenate is a good way to mimic a wound environment and present a cell with several different stimuli at the same time (DAMP, MAMP, as well as factors residing in the extracellular matrix like TGF $\beta$ ). To better understand the regulation of gene expression after tissue homogenate exposure, we performed ChIP-seq for the LDTFs PU.1 and FOS, for the KLA-induced factor P65, the TGF $\beta$ -induced factor SMAD3, and the histone activation mark H3K27ac under vehicle and tissue homogenate treated conditions. We found that binding of the factors was highly associated with tissue homogenate induced active enhancers. We overall observed a strong co-occurrence of P65 and SMAD3 at sites already bound by FOS and PU.1. Additionally, these sites overlapped with binding of REV-ERB, a factor involved in wound healing. Interestingly, we found enhancers that were bound by SMAD3 after TGF $\beta$  treatment, bound by P65 after KLA treatment and bound simultaneously by both factors after tissue homogenate stimulation. Tissue homogenate treatment established binding sites for P65 and SMAD3 that were not observed following treatment with KLA or TGF $\beta$ , respectively, and also showed binding of REV-ERB. Therefore, the complex tissue homogenate signal drove substantial co-localization of P65 with SMAD3 that was not observed following selective treatment with single stimuli. We furthermore found nuclear factor (erythroid-derived 2)-like (NRF)2 as the most enriched TF binding motif at SMAD binding sites after tissue homogenate treatment. ChIP-seq experiments demonstrated that tissue homogenate increased the genome-wide binding of NRF2 at thousands of genomic locations, a substantial fraction of which were observed to overlap with the tissue homogenate induced binding sites for P65, FOS, and SMAD3. We found an eight-fold increase in co-localization of these four factors at enhancers after tissue homogenate treatment suggesting that the combination of signals present in tissue homogenate induce co-binding of multiple TFs to enhancers that mediate the tissue injury response. This study provides first insights into how more complex stimuli regulate gene expression that cannot be observed under single stimuli. This stressed the importance of studying epigenetic regulation in cells that experience more complex signals. With the advances in technology it will become more and more feasible to study cells from *in vivo* environments directly under healthy and diseased states.

In cancer, mutations often cause dysregulation of gene expression programs in cells, leading to changing signals in the cell's environment. Therefore, tumors provide a good model to study how several different stimuli together influence the epigenetic landscape in cells. Roe et al. (2017) developed an *in vitro* model that closely mimics the *in vivo* environment of healthy cells and metastatic cancer cells. Comparing these, the authors found several thousand enhancers gained in the cancer cells that were also mostly present in human metastatic cells. The authors were able to show that the expression of the TF *FOXA1* was increased in tumor cells and further elevated in metastatic tumor cells. FOXA1 is a pioneer TF that was bound to the newly gained enhancers in metastatic cells in collaborative binding with the upregulated TFs GATA binding factor-5 (GATA5), BATF2, paired mesoderm homeobox protein 2 (PRRX2), and PAX9. The authors conclude that their findings suggest that a FOXA1-dependent enhancer reprogramming promotes progression of the cancer and metastasis *in vivo*.

Yang et al. (2015) also studied the influence of a unique tumor–environment in squamous cell carcinoma (SCC) on the epigenetic landscape of stem cells *in vivo*. By analyzing super enhancers (SEs) in healthy cells in comparison to tumor cells, the authors found many different SEs. They found an enrichment of SOX binding sites in tumor–related SEs, as well as AP–1 binding sites. Surprisingly, they also found E–twenty–six (ETS) consensus sites as enriched in tumor–associated SEs, although ETS proteins have not been previously implicated in SCC. However, they also showed new SEs at the *ETS domain–containing protein 3 (ELK3)* and *ETS2* genes and provided evidence that these factors drive hyper proliferation and SCC progression, as well as that *ELK3* and *ETS2* expression correlates with poor prognosis in human SCCs.

Our previous study demonstrated *in vitro* the importance to study the regulation of complex stimuli. Therefore, using the tumor environment to understand regulation of the epigenetic landscape is a logical step forward. Cancer is not only one of the most deadly diseases in the world, but it also provides a unique insight into how complex signals affect the enhancer landscape of cells.

Much work has been put into understanding how immune cells get activated and respond to different signals in the surrounding environment. However, to maintain a healthy tissue homeostasis, cells need to go back to their unstimulated state after the signal is resolved. Macrophages are good models to study this, because they respond to different signals, like inflammation or tissue injury, but they also return to baseline after the signal is resolved. It has been shown that macrophages synthesize anti–inflammatory fatty acids in an LXR–dependent manner after TLR4 stimulation (Oh et al. (2010), Spann et al. (2012), Li et al. (2013a)), but whether the production of these species contributes to the resolution phase of inflammatory responses has not been established. We (Oishi et al. (2017)) demonstrated a reciprocal relationship between levels of anti–inflammatory fatty acids and the expression of pro–inflammatory genes after TLR4 activation. Anti–inflammatory fatty acids that have the ability to suppress NF–kB are rapidly downregulated following stimulation of BMDMs with KLA and activation of the TLR4 response. However, 12 – 24 hours after stimulation, fatty acid levels rise again, initiating the resolution phase. The upregulation of anti–inflammatory fatty acids was independent of LXR, but instead sterol regulatory element–binding protein 1 (SREBP1)–driven. SREBP1 binding in the resolution phase is correlated with increased enhancer activity at its binding sites. Furthermore, SREBP1 knockout mice lack the ability to induce the production of fatty acids. Therefore, we show that SREBP1 and the production of anti–inflammatory fatty acids play crucial roles in the resolution of the inflammatory response of macrophages.

This important mechanism was recently confirmed by Körner et al. (2018) who showed a decrease of inflammation in sepsis after treating mice with omega–3 lipid emulsions. They observed a significant reduction in pro–inflammatory macrophages and an increase in alternately activated macrophages. Although the authors did not study the underlying transcriptional mechanisms for their observation, it seems plausible that the omega–3 fatty acids activate the same pathways as SREBP1. Furthermore, the authors observe an increased level of IL–10 and TGFβ. These cytokines are known to induce anti–inflammatory gene expression programs in macrophages.

In summary, complex *in vivo* stimuli might activate TF binding programs that cannot be observed in tissue cultures using single stimuli. We provided substantial evidence for this by using a signal mimicking an *in vivo* wound environment and observe combinations of TFs binding at enhancers that cannot be observed following single stimuli. Especially in the context of tissue–environment, it is important to analyze the response of tissue–resident cells to signals rather than removing them from their natural environment and stimulate them with single stimuli *in vitro*. However, this does not mean that tissue–culture studies with single stimuli are obsolete. In order to understand the precise molecular mechanisms at play, these studies can provide unprecedented insight into the effect of TF binding on the expression of genes. Only *in vitro* studies allow a controlled environment to understand signal processing. Using this controlled environment we were able to show that the resolution phase in macrophages after stimulation is partly regulated by SREBP1, in an LXR–independent manner. This new finding helps to understand how macrophages reach baseline gene expression again, after exposure to an acute inflammatory signal. To more precisely understand regulation of cells during disease, however, studying the cells in their natural environment during health and diseases states is absolutely essential.

## The influence of natural genetic variation on the epigenetic landscape and gene expression in macrophages

It has been a long–standing goal in biomedical research to understand the interplay between gene expression and disease. After sequencing the human genome (International Human Genome Sequencing Consortium (2001), Venter et al. (2001)), genome–wide association studies (GWAS) became possible, which allow the association of single–nucleotide polymorphisms (SNPs) to human traits or disease phenotypes. Most GWAS focused on common diseases like cancer (e.g. Broderick et al. (2007), Hunter et al. (2007), Eeles et al. (2008), Shiraishi et al. (2016)), type II diabetes mellitus (Diabetes Genetics Initiative et al. (2007), Scott et al. (2007), Hara et al. (2014)), neurodegenerative disease (Fung et al. (2006), Simon-Sanchez et al. (2007), Reiman et al. (2007), Vojinovic et al. (2015)), heart disease (Larson et al. (2007), Wellcome Trust Case Control Consortium (2007), Wang et al. (2016)), or auto–immune diseases (Yamazaki et al. (2005), Rioux et al. (2007), Raj et al. (2016), Sulem et al. (2011)). However, it is becoming clear that most SNPs associated with traits can be found in non–coding regions of the genome, rather than in coding sequences. This makes an interpretation of the results often times very hard. This also means that most consequential genetic variation is within regulatory regions of the genome rather than in coding regions that translate into proteins. To understand how these SNPs affect the host, a detailed knowledge of the affected cell type, as well as the epigenetic landscape within this cell type is necessary. Therefore, the research focus has shifted to understanding the direct influence of genetic variation on gene expression by epigenetically profiling different tissues and activation states.

One early study in humans investigated the relationship between common genetic polymor-

phisms and differences in gene expression in endothelial cells at baseline and after exposure to an oxidized phospholipid species to induce early atherosclerotic lesions (Romanoski et al. (2010)). The authors found significant interactions between genotype and response to phospholipids for about one-third of the most highly regulated genes. Much of this regulation is dependent on distal regulatory elements. This study therefore showed that genetic variation in non-coding regions of the genome can affect gene expression and alter the response to stimuli.

At about the same time another study (Kasowski et al. (2010)) investigated the influence of genetic variation on TF binding and gene expression in ten human samples at baseline. The authors showed that 7.5% of NF- $\kappa$ B binding sites and 25% of RNA polymerase II (Pol II) binding sites differed significantly between at least two individuals and that these binding sites were enriched for SNPs affecting the underlying binding motif. About 30% of differences in NF- $\kappa$ B binding could be explained by SNPs, however that also means 70% of differences in binding remained elusive. The authors observed a correlation between differences in binding and gene expression, but binding differences greatly exceeded transcriptome differences, hinting to a potential buffering system in the cells. TF binding itself, however, is often not sufficient for biological output. A better measurement is chromatin accessibility which can be measured by DNase I hypersensitivity sites sequencing (DNase-seq) and usually correlates well with nearby gene expression. Degner et al. (2012) performed DNase-seq in 70 Yoruba lymphoblastoid cell lines and showed that common genetic variants affect chromatin accessibility at thousands of hypersensitive regions across the human genome. They found that causal variants often lie within or very near to hypersensitivity regions, often affecting the binding affinity of TFs. Therefore, measuring chromatin accessibility rather than TF binding might be a better way to study the impact of genetic variation on gene expression.

An even more direct readout of chromatin activity is measuring histone modifications. As discussed in the General Introduction, activity of enhancers and promoters can be directly measured by histone marks. In 2013, Science published three back-to-back articles that investigated the effect of natural genetic variation on histone modifications (Kilpinen et al. (2013), McVicker et al. (2013), Kasowski et al. (2013)). Kilpinen et al. (2013) investigated differences in histone modifications, TF binding and gene expression in two parent-offspring trios, as well as eight unrelated individuals. They found that allele-specific activity across all regulatory layers is largely transmitted from parents to children. Their studies also suggest that TF binding, histone modifications, and transcription operate within the same allelic framework. Changes in TF binding showed correlation with gene expression, whereas changes in histone modifications did not, suggesting that changes in TF binding are often causal to changes in gene expression. This finding was supported by McVicker et al. (2013) who found that histone modifications are directed by sequence-specific TFs and that homozygous high-expression genotypes are more enriched in DNase I hypersensitivity and active histone marks, as well as depleted in repressive histone marks than heterozygous sites or homozygous low-expression genotypes. Therefore, their study suggests that single genetic variants can affect multiple aspects of chromatin states including histone activity marks, DNase I hypersensitivity and chromatin accessibility. Interestingly, Kasowski et al.

(2013) showed that despite the strong relationship between genetic variation and histone activity, there is not always an effect on gene expression. They suggest that variability in several enhancers is necessary to affect gene expression. They also found many genes that were equally expressed, but had one differentially activated enhancer, further supporting their hypothesis that several enhancers together regulate gene expression.

The observation that regions of differentially bound TFs have more impact on gene expression than differences in histone modifications, as well as the fact that only about 30% of the binding differences can be explained by direct mutations of the TF binding motif, raised the question how the majority of differently bound TFs are regulated. Heinz et al. (2013) tried to answer this question by studying binding of LDTFs and SDTFs in macrophages from two commonly used inbred mouse strains. They observed that only a fraction of TF binding could be explained by mutations in the respective consensus motif, however mutations in nearby TF motifs of collaborative factors increased the percentage of explained binding differences to about 60%. This was especially notable for the SDTF NF- $\kappa$ B, where only about 10% of differential binding could be explained by differences in the NF- $\kappa$ B motif, which increased to 35% when considering the binding motifs of the LDTFs PU.1 and C/EBP. Interestingly, the authors also showed that mutations in less conserved nucleotides in the consensus binding motif have less impact on TF binding than mutations in highly conserved residues. Furthermore, they showed that by swapping the enhancer sequence from one strain into the other, they also swapped the binding profile at this locus.

Although this study improved the number of loci that could be explained, a lot of differential binding remained unexplained. This led me to think that by only looking at three TFs a big part of the collaborative factors and motifs were missed, accounting for the gap in explaining the binding profiles. To address this question we developed Mutation Analysis for Regulatory Genomic Elements (MARGE) (Link et al. (2018b)), a software pipeline that allows the analysis of TF binding in several individuals at the same time by leveraging a linear mixed effects model. To avoid biases due to mapping errors caused by natural genetic variation, MARGE allows mapping of the data to custom made genomes, as well as shifting the data to the respective reference coordinates for downstream analysis. The software can analyze every kind of TF binding, open chromatin or enhancer activity data. To investigate the effect of natural genetic variation, each locus is annotated with all possible motifs. The binding of each factor is then modeled as the fixed effect motif existence/score with random effects locus and genotype for each motif. Motifs that impact binding of the factor will therefore show a significant association between the motif existence and the binding of a factor.

To investigate the scope of TFs that are involved in collaborative binding in macrophages, we took advantage of the big number of natural genetic variation between commonly used laboratory mouse strains and wild-derived mice (Link et al. (2018a)). We observed substantial strain-specific differences in gene expression in which increased genetic differences between mouse strains led to a substantial but non-linear increase in divergent gene expression. This was observed on the level of mRNA, as well as on the level of nascent transcription. The majority of the effects of genetic variation on nascent transcription

map to distal *cis*-regulatory elements, which was shown by using filial 1 hybrid (F1) mice. By applying MARGE to the TF binding data sets of four TFs in five different strains of mice, we found evidence for roles of about 100 TFs, increasing the percentage of explained differential binding sites to 70%. By only considering the 500 most differently bound loci, this fraction increased further to up to 90%. However, a substantial fraction of the strain-specific binding of these factors cannot be explained by local mutations. Investigation of the basis for this discrepancy led to the identification of highly interconnected clusters of TFs that reside within topologically associating domains (TADs). These connected regulatory domains (CRDs) are highly associated with strain-specific TF binding, enhancer activity, as well as nearest gene expression. They additionally show significant enrichment in three-dimensional interactions, further supporting the notion that CRDs present a new and important layer of transcriptional regulation.

Recently, two studies made similar observations. Waszak et al. (2015) profiled histone modifications, binding of PU.1 and Pol II, as well as gene expression in lymphoblastoid cell lines from 47 whole-genome sequenced individuals and observed coordinated chromatin variation across individuals, which they named variable chromatin modules (VCMs). VCMs were enriched for TF bound regions and in the majority of cases causal for observed changes in gene expression. Cheng et al. (2017) performed ATAC-seq, as well as gene expression profiling of T-cells from 105 healthy donors at basal state and after *in vitro* stimulation. They observed co-accessible ATAC-seq peaks across loci at kilobase and megabase scales, consistent with three dimensional chromosome organization patterns. They furthermore found enrichment for significant associations between gene expression and co-accessibility of ATAC-seq peaks.

Cheng et al. (2017) found enrichment of co-accessible ATAC-seq peaks in super enhancers (SEs), whereas we found CRDs to be distinct from SEs. This led to the conclusion that CRDs and co-accessible ATAC-seq peaks are different biological phenomena. However, the concept of chromatin accessibility that is concordantly regulated over several loci, as well as within TADs seems to be a powerful biological regulatory mechanism to control gene expression.

It is important to keep in mind that the effect of genetic variation on gene expression is dependent on the cell type (Ackermann et al. (2013)). SNPs that are completely silent in one specific cell type and do neither affect chromatin states nor gene expression can have dramatic effects in another cell type. Furthermore, many SNPs are silent in basal state and are associated with different responses to stimuli. Lee et al. (2014) showed that of 264 loci with genetic variants associated with changes in gene expression in human dendritic cells, 121 of them were only observed after one or more stimuli without any effect at basal state. This study underscored the need to study genetic variation in different cell types and under different stimuli in order to capture the whole impact of these variants on gene expression. More efforts have been put into studying the relationship between genetics and environment, in developing new computational methods (Knowles et al. (2017)), as well as in discovering the impact of natural genetic variation on gene expression between different tissues and stimuli (Li et al. (2017), Chen et al. (2016)). However, many questions remain unanswered and more effort is needed in order to understand the exact

mechanisms that regulate gene expression. Our work provided evidence that about 100 TFs confidently expressed in macrophages are involved in establishing the epigenetic landscape. This work should be expanded to other cell types to decipher the regulation of enhancers on a organism-wide scale. Furthermore, we provided evidence for an additional layer of transcriptional regulation by discovering connected regulatory domains (CRDs). This phenomenon has been undescribed so far and the underlying regulatory mechanisms remain unclear. Uncovering these mechanisms will provide the next step in understanding how enhancers regulate gene expression.

## Conclusion and further directions

This thesis studied the regulation of the epigenetic landscape and subsequently gene expression in different macrophages from different tissues and after exposure to different stimuli. Furthermore, we investigated the effect of natural genetic variation on mRNA expression, nascent transcription, TF binding, as well as histone modifications. We and others showed the influence of the tissue environment on the epigenetic landscape of the cell. With these studies we showed that the impact of the environmental factors plays a crucial role for cell identity and tissue culture systems are not sufficient to model these complex interactions. Furthermore, a cell experiences different signals at the same time, so studying regulation of gene expression *in vitro* using a single stimuli allows to study basic mechanisms, but does not even closely cover the processes observed *in vivo*. We showed that studying natural genetic variation is a powerful tool to learn more about regulation of TF binding and the epigenetic landscape. By studying it we further discovered a new layer of transcriptional regulation. In order to gain additional insight into the selection of enhancers, we developed Mutation Analysis for Regulatory Genomic Elements (MARGE), a new software that analyzes the effect of natural genetic variation on TF binding. With this software, we found evidence for the involvement of 100 confidently expressed TFs in macrophages in the selection of enhancers. However, this study was performed *in vitro* in order to minimize differences in gene expression and the epigenetic landscape due to different *in vivo* signals. This strategy proved to be useful in uncovering the scope of TFs involved in collaborative binding, as well as the impact of genetic variation on gene expression. However, this study allows only limited conclusion for the detailed mechanisms observed *in vivo*. To really uncover the regulation of gene expression, these strategies need to be combined. The next steps in advancing our current knowledge is to study different populations of *in vivo* cells from different individuals at baseline and after exposure to several different stimuli. Efforts are ongoing in the Glass laboratory and many other laboratories to set up systems and study the gene environment interaction in more detail. Cancer systems might provide a powerful resource, as the tumor environment differs substantially from healthy tissue. With the decreasing demand for cells to run NGS assays, as well as the better collaboration between surgeons and researchers, these studies become more and more feasible. Recently, several groups, including us, have described new layers of transcriptional regulation by showing correlated regions of open chromatin

or TF binding. The underlying mechanisms for this phenomenon are completely unknown yet. Several possible explanations are plausible. For example, it is possible that one SNP causes the disruption of Pol II, which then in turn cannot remodel the chromatin anymore and the binding sites remain inaccessible for the TFs. Another possible explanation is that rearrangement of chromatin due to long non-coding RNAs can remodel the nucleosomes and therefore make the binding site accessible. More research will be necessary in the future to understand these large-scale differences that seem to have the biggest impact on gene expression.



# Bibliography

- Ackermann, M., Sikora-Wohlfeld, W., and Beyer, A. (2013). Impact of Natural Genetic Variation on Gene Expression Dynamics. *PLoS Genet.*, 9:e1003514.
- Adam, R. C., Yang, H., Rockowitz, S., Larsen, S. B., Nikolova, M., Oristian, D. S., Polak, L., Kadaja, M., Asare, A., Zheng, D., and E., F. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature*, 521:366370.
- Adams, M. D., Kelly, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252:1651–1656.
- Asmann, Y. W., Necela, B. M., Kalari, K. R., Hossain, A., Baker, T. R., Carr, J. M., Davis, C., Getz, J. E., Hostetter, G., Li, X., McLaughlin, S. A., Radisky, D. C., Schroth, G. P., Cunliffe, H. E., Perez, E. A., and Thompson, E. A. (2012). Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res.*, 72:1921–1928.
- Auerbach, R. K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M., and Snyder, M. (2009). Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. U.S.A.*, 106:14926–14931.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a  $\beta$ -globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27:299–308.
- Barish, G. D., Yu, R. T., Karunasiri, M., Ocampo, C. B., Dixon, J., Benner, C., Dent, A. L., Tangirala, R. K., and Evans, R. M. (2010). Bcl-6 and NF-kappaB cistromes mediate opposing regulation of the innate immune response. *Genes Dev*, 24:2760–2765.
- 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.
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129:823–837.

- Benjamini, Y. and Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res*, 40:e72.
- Bonn, S., Zinzen, R. P., Girardot, C., Gustafson, E. H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E. E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.*, 44:148–156.
- Bossard, P. and Zaret, K. S. (1998). GATA transcription factors as potentiators of gut endoderm differentiation. *Development*, 125:4909–4917.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., DuBridg, R. B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, 18:630–634.
- Broderick, P., Carvajal-Carmona, L., Pittman, A. M., Webb, E., Howarth, K., Rowan, A., Lubbe, S., Spain, S., Sullivan, K., Fielding, S., Jaeger, E., Vijayakrishnan, J., Kemp, Z., Gorman, M., Chandler, I., Papaemmanuil, E., Penegar, S., Wood, W., Sellick, G., Qureshi, M., Teixeira, A., Domingo, E., Barclay, E., Martin, L., Sieber, O., CORGI Consortium, Kerr, D., Gray, R., Peto, J., Cazier, J. B., Tomlinson, I., and Houlston, R. S. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.*, 39:1315–1317.
- Budry, L., Balsalobre, A., Gauthier, Y., Khetchoumian, K., L'honoé, A., Vallette, S., Brue, T., Figarella-Branger, D., Meij, B., and Drouin, J. (2012). The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev.*, 26:2299–2310.
- 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. *Nature Methods*, 10:1213–1218.
- Carr, A. and Biggin, M. D. (1999). A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeo-protein DNA binding in *Drosophila* embryos. *EMBO J.*, 18:1598–1608.
- Chen, L., Ge, B., Casale, F. P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., Datta, A., Richardson, D., Burden, F., Mead, D., Mann, A. L., Fernandez, J. M., Rowlston, S., Wilder, S. P., Farrow, S., Shao, X., Lambourne, J. J., Redensek, A., Albers, C. A., Amstislavskiy, V., Ashford, S., Berentsen, K., Bomba, L., Bourque, G., Bujold, D., Busche, S., Caron, M., Chen, S. H., Cheung, W., Delaneau, O., Dermitzakis, E. T., Elding, H., Colgiu, I., Bagger, F. O., Flicek, P., Habibi, E., Iotchkova, V., Janssen-Megens, E., Kim, B., Lehrach, H., Lowy, E., Mandoli, A., Matarese, F.,

- Maurano, M. T., Morris, J. A., Pancaldi, V., Pourfarzad, F., Rehnstrom, K., Rendon, A., Risch, T., Sharifi, N., Simon, M. M., Sultan, M., Valencia, A., Walter, K., Wang, S. Y., Frontini, M., Antonarakis, S. E., Clarke, L., Yaspo, M. L., Beck, S., Guigo, R., Rico, D., Martens, J. H. A., Ouwehand, W. H., Kuijpers, T. W., Paul, D. S., Stunnenberg, H. G., Stegle, O., Downes, K., Pastinen, T., and Soranzo, N. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell*, 167:1398–1414.
- Chen, Y., Negre, N., Li, Q., Mieczkowska, J. O., Slattery, M., Liu, T., Zhang, Y., Kim, T. K., He, H. H., Zieba, J., Ruan, Y., Bickel, P. J., Myers, R. M., Wold, B. J., White, K. P., Lieb, J. D., and Liu, X. S. (2010). Systematic evaluation of factors influencing ChIP-seq fidelity. *Nature Methods*, 9:609–614.
- Cheng, C. S., Gate, R. E., Aiden, A. P., Siba, A., Tabaka, M., Lituiev, D., Machol, I., Subramaniam, M., Shammim, M., Hougen, K. L., Wortman, I., Huang, S., Durand, N. C., Feng, T., De Jager, P. L., Chang, H. Y., Lieberman Aiden, E., Benoist, C., Beer, M. A., Ye, C. J., and Regev, A. (2017). Genetic determinants of co-accessible chromatin regions in T cell activation across humans. *BioRxiv*.
- Cheng, Y., Ma, Z., Kim, B. H., Wu, W., Cayting, P., Boyle, A. P., Sundaram, V., Xing, X., Dogan, N., Li, J., Euskirchen, G., Lin, S., Lin, Y., Visel, A., Kawli, T., Yang, X., Patacsil, D., Keller, C. A., Giardine, B., mouse ENCODE Consortium., Kundaje, A., Wang, T., Pennacchio, L. A., Weng, Z., Hardison, R. C., and Snyder, M. P. (2014). Principles of regulatory information conservation between mouse and human. *Nature*, 515:371–375.
- Cheung, M. S., Down, T. A., Latorre, I., and Ahringer, J. (2011). Systematic bias in high-throughput sequencing data and its correction by BEADS. *Nucleic Acid Res*, 39:e103.
- Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., McKernan, K. J., and Grimmond, S. M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*, 5:613–619.
- Colla, S., Ong, D., Ogoti, Y., Marchesini, M., Mistry, N. A., Clise-Dwyer, K., Ang, S. A., Storti, P., Viale, A., Giuliani, N., Ruisaardi, K., Ganan Gomez, I., Bristow, C. A., Estecio, M., Weksberg, D. C., Ho, Y. W., Hu, B., Genovese, G., Pettazzoni, P., Murtani, A. S., Jiang, S., Hua, S., Ryan, M. C., Carugo, A., Nezi, L., Wei, Y., Yang, H., D’Anca, M., Zhang, L., Gaddis, S., Gong, T., Horner, J. W., Heffernan, T. P., Jones, P., Cooper, L. J., Liang, H., Kantarjian, H., Wang, Y. A., Chin, L., Bueso-Ramos, C., Garcia-Manero, G., and DePinho, R. A. (2015). Telomere dysfunction drives aberrant hematopoietic differentiation and myelodysplastic syndrome. *Cancer Cell*, 27:644–657.
- Core, L. J., Waterfall, J. J., and Lis, J. T. (2010). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, 2010:1845–1848.

- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., and Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.*, 107:21931–21936.
- Deardorff, M. A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., Cole, K. E., De Baere, E., Decroos, C., Di Donato, N., Ernst, S., Francey, L. J., Gyftodimou, Y., Hirashima, K., Hullings, M., Ishikawa, Y., Jaulin, C., Kaur, M., Kiyono, T., Lombardi, P. M., Magnaghi-Jaulin, L., Mortier, G. R., Nozaki, N., Petersen, M. B., Seimiya, H., Siu, V. M., Suzuki, Y., Takagaki, K., Wilde, J. J., Willems, P. J., Prigent, C., Gillessen-Kaesbach, G., Christianson, D. W., Kaiser, F. J., Jackson, L. G., Hirota, T., Krantz, I. D., and Shirahige, K. (2012). HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature*, 489:313–317.
- Degner, J. F., Pai, A. A., Pique-Regi, R., Veyrieras, J., Gaffney, D. J., Pickrell, J. K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G. E., Stephens, M., Gilad, Y., and Pritchard, J. K. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*, 482:390–394.
- D’haeseleer, P. (2006). What are DNA sequence motifs? *Nat Biotechnol*, 24:423–425.
- Diabetes Genetics Initiative, o., Saxena, R., Voight, B. F., Lyssenko, V., Burt, N. P., de Bakker, P. I., Chen, H., Roix, J. J., Kathiresan, S., Hirschhorn, J. N., Daly, M. J., Hughes, T. E., Groop, L., Altshuler, D., Almgren, P., Florez, J. C., Meyer, J., Ardlie, K., Bengtsson Boström, K., Isomaa, B., Lettre, G., Lindblad, U., Lyon, H. N., Melander, O., Newton-Cheh, C., Nilsson, P., Orho-Melander, M., Råstam, L., Speliotes, E. K., Taskinen, M. R., Tuomi, T., Guiducci, C., Berglund, A., Carlson, J., Gianniny, L., Hackett, R., Hall, L., Holmkvist, J., Laurila, E., Sjögren, M., Sterner, M., Surti, A., Svensson, M., Svensson, M., Tewhey, R., Blumenstiel, B., Parkin, M., Defelice, M., Barry, R., Brodeur, W., Camarata, J., Chia, N., Fava, M., Gibbons, J., Handsaker, B., Healy, C., Nguyen, K., Gates, C., Sougnez, C., Gage, D., Nizzari, M., Gabriel, S. B., Chirn, G. W., Ma, Q., Parikh, H., Richardson, D., Ricke, D., and Purcell, S. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 315:1331–1336.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29:15–21.
- Eeles, R. A., Kote-Jarai, Z., Giles, G. G., Olama, A. A., Guy, M., Jugurnauth, S. K., Mulholland, S., Leongamornlert, D. A., Edwards, S. M., Morrison, J., Field, H. I., Southey, M. C., Severi, G., Donovan, J. L., Hamdy, F. C., Dearnaley, D. P., Muir, K. R., Smith, C., Bagnato, M., Ardern-Jones, A. T., Hall, A. L., O’Brien, L. T., Gehr-Swain, B. N., Wilkinson, R. A., Cox, A., Lewis, S., Brown, P. M., Jhavar, S. G., Tymrakiewicz,

- M., Lophatananon, A., Bryant, S. L., UK Genetic Prostate Cancer Study Collaborators, British Association of Urological Surgeons' Section of Oncology, UK ProtecT Study Collaborators, Horwich, A., Huddart, R. A., Khoo, V. S., Parker, C. C., Woodhouse, C. J., Thompson, A., Christmas, T., Ogden, C., Fisher, C., Jamieson, C., Cooper, C. S., English, D. R., Hopper, J. L., Neal, D. E., and Easton, D. F. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.*, 40:316–321.
- Egelhofer, T. A., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwierz, P., Alekseyenko, A. A., Cheung, M., Day, D. S., Gadel, S., Gorchakov, A. A., Gu, T., Kharchenko, P. V., Kuan, S., Latorre, I., Linder-Basso, D., Luu, Y., Ngo, W., Perry, M., Rechtsteiner, A., Riddle, N. C., Schwartz, Y. B., Shanower, G. A., Vielle, A., Ahringer, J., Elgin, S. C. R., Kuroda, M. I., Pirrotta, V., Ren, B., Strome, S., Park, P. J., Karpen, G. H., Hawkins, R. D., and Lieb, J. D. (2011). An assessment of histone-modification antibody quality. *Nat. Struct. Mol. Biol.*, 18:91–93.
- Eichenfield, D. Z., Troutman, D. T., Link, V. M., Lam, M. T., Cho, H., Gosselin, D., Spann, N. J., Lesch, H. P., Tao, J., Muto, J., Gallo, R. J., Evans, R. M., and Glass, C. K. (2016). Tissue damage drives co-localization of NF- $\kappa$ B, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages. *eLife*, 5.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489:57–74.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shores, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., and Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473:43–49.
- Fung, H. C., Scholz, S., Matarin, M., Simón-Sánchez, J., Hernandez, D., Britton, A., Gibbs, J. R., Langefeld, C., Stiegert, M. L., Schymick, J., Okun, M. S., Mandel, R. J., Fernandez, H. H., Foote, K. D., Rodríguez, R. L., Peckham, E., De Vrieze, F. W., Gwinn-Hardy, K., Hardy, J. A., and Singleton, A. (2006). Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol*, 5:911–916.
- Geissmann, F., Manz, M. G., Jung, S., Sieweke, M. H., Merad, M., and Ley, K. (2010). Development of monocytes, macrophages, and dendritic cells. *Science*, 327:656–661.
- Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., Yip, K. Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R. K., Feng, X., Leng, J., Vielle, A., Niu, W., Rhrissorrakrai, K., Agarwal, A., Alexander, R. P., Barber, G., Brdlik, C. M., Brennan, J., Brouillet, J. J., Carr, A., Cheung, M. S., Clawson, H., Contrino, S., Dannenberg, L. O., Dernburg, A. F., Desai, A., Dick, L., Dosé, A. C., Du, J., Egelhofer, T., Ercan, S., Euskirchen, G., Ewing, B., Feingold, E. A., Gassmann, R., Good, P. J., Green, P., Gullier, F., Gutwein, M.,

- Guyer, M. S., Habegger, L., Han, T., Henikoff, J. G., Henz, S. R., Hinrichs, A., Holster, H., Hyman, T., Iniguez, A. L., Janette, J., Jensen, M., Kato, M., Kent, W. J., Kephart, E., Khivansara, V., Khurana, E., Kim, J. K., Kolasinska-Zwierz, P., Lai, E. C., Latorre, I., Leahey, A., Lewis, S., Lloyd, P., Lochovsky, L., Lowdon, R. F., Lubling, Y., Lyne, R., MacCoss, M., Mackowiak, S. D., Mangone, M., McKay, S., Mecnas, D., Merrihew, G., Miller, D. M. r., Muroyama, A., Murray, J. I., Ooi, S. L., Pham, H., Phippen, T., Preston, E. A., Rajewsky, N., Rättsch, G., Rosenbaum, H., Rozowsky, J., Rutherford, K., Ruzanov, P., Sarov, M., Sasidharan, R., Sboner, A., Scheid, P., Segal, E., Shin, H., Shou, C., Slack, F. J., Slightam, C., Smith, R., Spencer, W. C., Stinson, E. O., Taing, S., Takasaki, T., Vafeados, D., Voronina, K., Wang, G., Washington, N. L., Whittle, C. M., Wu, B., Yan, K. K., Zeller, G., Zha, Z., Zhong, M., Zhou, X., modENCODE, C., Ahringer, J., Strome, S., Gunsalus, K. C., Mickle, G., Liu, X. S., Reinke, V., Kim, S. K., Hillier, L. W., Henikoff, S., Piano, F., Snyder, M., Stein, L., Lieb, J. D., and Waterston, R. H. (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science*, 330:1775–1787.
- Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C., Ragoussis, J., and Natoli, G. (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity*, 32:317–328.
- Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.*, 17:877–885.
- Gorden, S. (2003). Alternative activation of macrophages. *Nat. Rev. Immunol.*, 3:23–35.
- Gordon, S., Plüddemann, A., and Martinez Estrada, F. (2014). Macrophage heterogeneity in tissues: phenotypic diversity and functions. *Immunol Rev*, 262:36–55.
- Gosselin, D., Skola, D., Coufal, Holtman, I. R., Schlachetzki, J. C. M., Sajti, E., Jaeger, B. N., OConnor, C., Fitzpatrick, C., Pasillas, M. P., Pena, M., Adair, A., Gonda, D. G., Levy, M. L., Ransohoff, R. M., Gage, F. H., and Glass, C. K. (2017). An environment-dependent transcriptional network specifies human microglia identity. *Science*, 356:eaal3222.
- Gosselin, D., Link, V. M., Romanoski, C. E., Fonseca, G. J., Eichenfield, D. Z., Spann, N. J., Stender, J. D., Chun, H. B., Garner, H., Geissmann, F., and Glass, C. K. (2014). Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell*, 159:1327–40.
- Griffith, M., Griffith, O. L., Mwenifumbo, J., Goya, R., Morrissy, A. S., Morin, R. D., Corbett, R., Tang, M. J., Hou, Y. C., Pugh, T. J., Robertson, G., Chittaranjan, S., Ally, A., Asano, J. K., Chan, S. Y., Li, H. I., McDonald, H., Teague, K., Zhao, Y., Zeng, T.,

- Delaney, A., Hirst, M., Morin, G. B., Jones, S. J., Tai, I. T., and Marra, M. A. (2010). Alternative expression analysis by RNA sequencing. *Nature Methods*, 7:843–847.
- Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., and Kraus, W. T. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, 145:622–634.
- Hara, K., Fujita, H., Johnson, T. A., Yamauchi, T., Yasuda, K., Horikoshi, M., Peng, C., Hu, C., Ma, R. C., Imamura, M., Iwata, M., Tsunoda, T., Morizono, T., Shojima, N., So, W. Y., Leung, T. F., Kwan, P., Zhang, R., Wang, J., Yu, W., Maegawa, H., Hirose, H., DIAGRAM consortium, Kaku, K., Ito, C., Watada, H., Tanaka, Y., Tobe, K., Kashiwagi, A., Kawamori, R., Jia, W., Chan, J. C., Teo, Y. Y., Shyong, T. E., Kamatani, N., Kubo, M., Maeda, S., and Kadowaki, T. (2014). Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum. Mol. Genet.*, 23:239–246.
- He, H. H., Meyer, C. A., Shin, H., Bailey, S. T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., Mieczkowski, P., Lieb, J. D., Zhao, K., Brown, M., and Liu, X. S. (2010). Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.*, 42:343–347.
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, 39:311–318.
- Heinz, S., Benner, C., Spann, N. J., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, 38:576–589.
- Heinz, S., Romanoski, C. E., Benner, C., Allison, K. A., Kaikkonen, M. U., Orozco, L. D., and Glass, C. K. (2013). Effect of natural genetic variation on enhancer selection and function. *Nature*, 503:487–492.
- Heng, T. S. P., Painter, M. W., and The Immunological Genome Project Consortium (2008). The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.*, 9:1091–1094.
- Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., and Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell*, 155:934–947.
- Hsieh, C. L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C. J., Lee, G. S., Chen, S., Balk, S. P., Liu, X. S., Brown, M., and Kantoff, P. W. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. U S A*, 111:7319–7324.

- Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys, S. S., McCarty, C. A., Feigelson, H. S., Calle, E. E., Thun, M. J., Hayes, R. B., Tucker, M., Gerhard, D. S., Fraumeni, J. F. J., Hoover, R. N., Thomas, G., and Chanock, S. J. (2007). A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, 39:870–874.
- International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, 409:860–921.
- Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, 409:533–538.
- Johnson, S. D., Mortazavi, A. M., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316:1497–1502.
- 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., and Glass, C. K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell*, 51:310–325.
- Kasowski, M., Kyriazopoulou-Panagiotopoulou, S., Grubert, F., Zaugg, J. B., Kundaje, A., Liu, Y., Boyle, A. P., Zhang, Q. C., Zakharia, F., Spacek, D. V., Li, J., Xie, D., Olarerin-George, A., Steinmetz, L. M., Hogenesch, J. B., Kellis, M., Batzoglou, S., and Snyder, M. (2013). Extensive Variation in Chromatin States Across Humans. *Science*, 342:750–752.
- Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S. M., Habegger, L., Rozowsky, J., Shi, M., Urban, A. E., Hong, M., Karczewski, K. J., Huber, W., Weissman, S. M., Gerstein, M. B., Korbel, J. O., and Snyder, M. (2010). Variation in Transcription Factor Binding Among Humans. *Science*, 328:232–235.
- Kazemian, M., Pham, H., Wolfe, S. A., Brodsky, M. H., and Sinha, S. (2013). Widespread evidence of cooperative DNA binding by transcription factors in *Drosophila* development. *Nucleic Acids Res*, 41:8237–8252.
- Khrameeva, E. E. and Gelfand, M. S. (2012). Biases in read coverage demonstrated by interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. *BMC Bioinformatics*.
- Kieffer-Kwon, K. R., Tang, Z., Mathe, E., Qian, J., Sung, M. H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., Vian, L., Nelson, S., Zare, H., Hakim, O., Reyon, D., Yamane, A., Nakahashi, H., Kovalchuk, A. L., Zou, J., Joung, J. K., Sartorelli, V., Wei, C. L., Ruan, X., Hager, G. L., Ruan, Y., and Casellas, R. (2013). Interactome maps

- of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell*, 155:1507–1520.
- Kilpinen, H., Waszak, S. M., Gschwind, A. R., Raghav, S. K., Witwicki, R. M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N. I., Yurovsky, A., Lappalainen, T., Romano-Palumbo, L., Planchon, A., Bielser, D., Bryois, J., Padioleau, I., Udin, G., Thurnheer, S., Hacker, D., Core, L. J., Lis, J. T., Hernandez, N., Reymond, A., Deplancke, B., and Dermitzakis, E. T. (2013). Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription. *Science*, 342:744–747.
- Knowles, D. A., Davis, J. R., Edgington, H., Raj, A., Favé, M., Zhu, X., Potash, J. B., Weissman, M. M., Shi, J., Levinson, D. F., Awadalla, P., Mostafavi, S., Montgomery, S. B., and Battle, A. (2017). Allele-specific expression reveals interactions between genetic variation and environment. *Nature Methods*, 14:699–704.
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., Hayashizaki, Y., and Carninci, P. (2006). CAGE: cap analysis of gene expression. *Nature Methods*, 3:211–222.
- Körner, A., Schlegl, M., Theurer, J., Frohnmeyer, H., Adolph, M., Heinink, M., Giera, M., Rosenberger, P., and Mirakaj, V. (2018). Resolution of inflammation and sepsis survival are improved by dietary  $\omega$ -3 fatty acids. *Cell Death Differ*, 25:421–431.
- Lam, M. T., Cho, H., Lesch, H. P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M. U., Kim, A. S., Kosaka, M., Lee, C. Y., Watt, A., Grossman, T. R., Rosenfeld, M. G., Evans, R. M., and Glass, C. K. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 498:511–515.
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9:357–359.
- Larson, M. G., Atwood, L. D., Benjamin, E. J., Cupples, L. A., D’Agostino, R. B. S., Fox, C. S., Govindaraju, D. R., Guo, C. Y., Heard-Costa, N. L., Hwang, S. J., M., M. J., Newton-Cheh, C., O’Donnell, C. J., Seshadri, S., Vasan, R. S., Wang, T. J., Wolf, P. A., and Levy, D. (2007). Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med. Genet.*, 19:S5.
- Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Sshaul, H., Mered, M., Jung, S., and Amit, I. (2014). Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment. *Cell*, 159:1312–1326.
- Lee, C. S., Friedman, J. R., Fulmer, J. T., and Kaestner, K. H. (2005). The initiation of liver development is dependent on Foxa transcription factors. *Nature*, 435:944–947.

- Lee, M. N., Ye, C., Villani, A., Raj, T., Li, W., Eisenhaure, T. M., Imboywa, S. H., Chipendo, P. I., Ran, F. A., Slowikowski, K., Ward, L. D., Raddassi, K., McCabe, C., Lee, M. H., Frohlich, I. Y., Hafler, D. A., Kellis, M., Raychaudhuri, S., Zhang, F., Stranger, B. E., Benoist, C. O., De Jager, P. L., Regev, A., and Hacohen, N. (2014). Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Science*, 343:1246980.
- Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25:1754–1760.
- Li, P., Spann, N. J., Kaikkonen, M. U., Lu, M., Oh, D. Y., Fox, J. N., Bandyopadhyay, G., Talukdar, S., Xu, J., Lagakos, W. S., Patsouris, D., Armando, A., Quehenberger, O., Dennis, E. A., Watkins, S. M., Auwerx, J., Glass, C. K., and Olefsky, J. M. (2013a). NCoR repression of LXRs restricts macrophage biosynthesis of insulin-sensitizing omega 3 fatty acids. *Cell*, 155:200–214.
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H. S., Glass, C. K., and Rosenfeld, M. G. (2013b). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498:516–520.
- Li, X., Kim, Y., Tsang, E. K., Davis, J. R., Damani, F. N., Chiang, C., Hess, G. T., Zappala, Z., Strober, B. J., Scott, A. J., Li, A., Ganna, A., Bassik, M. C., Merker, J. D., GTEx Consortium, ., Hall, I. M., Battle, A., and B., M. S. (2017). The impact of rare variation on gene expression across tissues. *Nature*, 550:293–243.
- Link, V. M., Duttke, S. H., Chun, H. B., Holtman, I. R., Westin, E., Hoeksema, M., Abe, Y., Skola, D., Romanoski, C. E., Tao, J., Fonseca, G. J., Troutman, D. T., Spann, N. J., Strid, T., Sakai, M., Yu, M., Hu, R., Fang, R., Metzler, D., Ren, B., and Glass, C. K. (2018a). Transcription Factor Landscapes in Macrophages from Genetically Diverse Mice Reveal Connected Regulatory Domains. *Cell - under review*.
- Link, V. M., Gosselin, D., and Glass, C. K. (2015). Mechanisms Underlying the Selection and Function of Macrophage-Specific Enhancers. *Cold Spring Harb Symp Quant Biol*, 80:213–221.
- Link, V. M., Romanoski, C. E., Metzler, D., and Glass, C. K. (2018b). MARGE: Mutation Analysis for Regulatory Genomic Elements. *in preparation*.
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, 133:523–526.
- Liu, Y. C., Zou, X. B., Chai, Y. F., and Yao, Y. M. (2014). Macrophage polarization in inflammatory diseases. *Int. J. Biol. Sci.*, 10:520–529.

- Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, 13:1675–1680.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2002). *Molecular Cell Biology. 4th edition.* W H Freeman & Co (Sd).
- Lucas, T., Waisman, A., Ranjan, R., Roes, J., Krieg, T., Müller, W., Roers, A., and Eming, S. (2010). Differential roles of macrophages in diverse phases of skin repair. *J. Immunol.*, 184:3964–3977.
- Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage specific transcription. *Cell*, 132:958–970.
- Maher, C. A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., and Chinnaiyan, A. M. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature*, 548:97–101.
- Mass, E., Ballesteros, I., Farlik, M., Halbritter, F. and Günther, P., Crozet, L., Jacome-Galarza, C. E., Händler, K. Klughammer, J., Kobayashi, Y., Gomez-Perdiguero, E., Schultze, J. L., Beyer, M., Bock, C., and F., G. (2016). Specification of tissue-resident macrophages during organogenesis. *Science*, 353:aaf4238.
- Matcovitch-Natan, O., Winter, D. R., Giladi, A., Aguilar, S. V., Spinrad, A., Sarrazin, S., Ben-Yehuda, H., David, E., González, F. G., Perrin, P., Keren-Shaul, H., Gury, M., Lara-Astaiso, D., Thaïss, C. A., Cohen, M., Halpern, K. B., Baruch, K., Deczkowska, A., Lorenzo-Vivas, E., Itzkovitz, S., Elinav, E., Sieweke, M. H., Schwartz, M., and Amit, I. (2016). Microglia development follows a stepwise program to regulate brain homeostasis. *Science*, 23:aad8670.
- McConnell, S. C., Huo, Y., Liu, S., and Ryan, T. M. (2011). Human globin knock-in mice complete fetal-to-adult hemoglobin switching in postnatal development. *Mol. Cell. Biol.*, 31:876–883.
- McPherson, C. E., Shim, E. Y., Friedman, D. S., and Zaret, K. S. (1993). An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell*, 75:387–398.
- McVicker, G., van de Geijn, B., Degner, J. F., Cain, C. E., Banovich, N. E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J. K. (2013). Identification of Genetic Variants That Affect Histone Modifications in Human Cells. *Science*, 342:747–749.

- Medzhitov, R. and Horng, T. (2009). Transcriptional control of the inflammatory response. *Nature Review Immunology*, 9:692–703.
- Mendoza-Parra, M. A., Pattabhiraman, S., and Gronemeyer, H. (2012). Sequential chromatin immunoprecipitation protocol for global analysis through massive parallel sequencing (reChIP-seq). *Protocol Exchange*.
- Mikkelsen, T. S., Xu, Z., Zhang, X., Wang, L., Gimble, J. M., Lander, E. S., and Rosen, E. D. (2010). Comparative epigenomic analysis of murine and human adipogenesis. *Cell*, 143:156–169.
- Minoche, A. E., Dohm, J. C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.*, 12:R112.
- modENCODE, C., Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., Landolin, J. M., Bristow, C. A., Ma, L., Lin, M. F., Washietl, S., Arshinoff, B. I., Ay, F., Meyer, P. E., Robine, N., Washington, N. L., Di Stefano, L., Berezikov, E., Brown, C. D., Candéias, R., Carlson, J. W., Carr, A., Jungreis, I., Marbach, D., Sealfon, R., Tolstorukov, M. Y., Will, S., Alekseyenko, A. A., Artieri, C., Booth, B. W., Brooks, A. N., Dai, Q., Davis, C. A., Duff, M. O., Feng, X., Gorchakov, A. A., Gu, T., Henikoff, J. G., Kapranov, P., Li, R., MacAlpine, H. K., Malone, J., Minoda, A., Nordman, J., Okamura, K., Perry, M., Powell, S. K., Riddle, N. C., Sakai, A., Samsonova, A., Sandler, J. E., Schwartz, Y. B., Sher, N., Spokony, R., Sturgill, D., van Baren, M., Wan, K. H., Yang, L., Yu, C., Feingold, E., Good, P., Guyer, M., Lowdon, R., Ahmad, K., Andrews, J., Berger, B., Brenner, S. E., Brent, M. R., Cherbas, L., Elgin, S. C., Gingeras, T. R., Grossman, R., Hoskins, R. A., Kaufman, T. C., Kent, W., Kuroda, M. I., Orr-Weaver, T., Perrimon, N., Pirrotta, V., Posakony, J. W., Ren, B., Russell, S., Cherbas, P., Graveley, B. R., Lewis, S., Micklem, G., Oliver, B., Park, P. J., Celniker, S. E., Henikoff, S., Karpen, G. H., Lai, E. C., MacAlpine, D. M., Stein, L. D., White, K. P., and Kellis, M. (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*, 330:1787–1797.
- Moore, K. J. and Tabas, I. (2011). Macrophages in the pathogenesis of atherosclerosis. *Cell*, 145:341–355.
- Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., McDonald, H., Varhol, R., Jones, S., and Marra, M. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques*, 45:81–94.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5:621–628.
- Mosser, D. M. and Edwards, J. P. (2008). Exploring the full spectrum of macrophage activation. *Nat. Rev. Immunol.*, 8:958–969.

- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320:1344–1349.
- Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M. C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., and Kanaya, S. (2011). Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res*, 39:e90.
- Noy, R. and Pollard, J. W. (2014). Tumor-associated macrophages: from mechanisms to therapy. *Immunity*, 41:49–61.
- Odegaard, J. I., Ricardo-Gonzalez, R. R., Goforth, M. H., Morel, C. R., Subramanian, V., L., M., A., R. E., D., V., F., B., AW., F., and A., C. (2007). Macrophage-specific PPARgamma controls alternative activation and improves insulin resistance. *Nature*, 447:1116–1120.
- Oh, D. Y., Talukdar, S., Bae, E. J., Imamura, T., Morinaga, H., Fan, W., Li, P., Lu, W. J., Watkins, S. M., and Olefsky, J. M. (2010). GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell*, 42:687–698.
- Oishi, Y., Spann, N. J., Link, V. M., Muse, E. D., Strid, T., Edillor, C., Kolar, M. J., Matsuzaka, T., Hayakawa, S., Tao, J., Kaikkonen, M. U., Carlin, A. F., Lam, M. T., Manabe, I., Shimano, H., Saghatelian, A., and Glass, C. K. (2017). SREBP1 Contributes to Resolution of Pro-inflammatory TLR4 Signaling by Reprogramming Fatty Acid Metabolism. *Cell Metab.*, 25:412427.
- 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.
- Ozsolak, F., Platt, A. R., Jones, D. R., Reifemberger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M., and Milos, P. M. (2009). Direct RNA sequencing. *Nature*, 461:814–818.
- Paolicelli, R. C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., Giustetto, M., Ferreira, T. A., Guiducci, E., Dumas, L., Ragozzino, D., and Gross, C. T. (2011). Synaptic pruning by microglia is necessary for normal brain development. *Science*, 333:1456–1458.
- Perdiguero, E. G., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M. F., Geissmann, F., and Rodewald, H. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature*, 518:547551.

- Peterson, K. R., Clegg, C. H., Huxley, C., Josephson, B. M., Haugen, H. S., Furukawa, T., and Stamatoyannopoulos, G. (1993). Transgenic mice containing a 248-kb yeast artificial chromosome carrying the human beta-globin locus display proper developmental control of human globin genes. *Proc. Natl. Acad. Sci. U.S.A.*, 90:7593–7597.
- Pirzgalska, R. M., Seixas, E., Seidman, J. S., Link, V. M., Martínez Sánchez, N., Mahú, I., Mendes, R., Gres, V., Kubasova, N., Morris, I., Arús, B. A., Larabee, C. M., Vasques, M., Tortosa, F., Sousa, A. L., Anandan, S., Tranfield, E., Hahn, M. K., Iannacone, M., Spann, N. J., Glass, C. K., and Domingos, A. I. (2017). Sympathetic neuron-associated macrophages contribute to obesity by importing and metabolizing norepinephrine. *Nat. Med.*, 23:1309–1321.
- Pollard, J. W. (2004). Tumour-educated macrophages promote tumour progression and metastasis. *Nat. Rev. Cancer*, 3:71–78.
- Qin, F., Song, Z., Babiceanu, M., Song, Y., Facemire, L., Singh, R., Adli, M., and Li, H. (2015). Discovery of CTCF-sensitive Cis-spliced fusion RNAs between adjacent genes in human prostate cells. *PLoS Genet.*, 11.
- Qiu, Y., Nguyen, K. D., Odegaard, J. I., Cui, X., Tian, X., Locksley, R. M., Palmiter, R. D., and Chawla, A. (2014). Eosinophils and type 2 cytokine signaling in macrophages orchestrate development of functional beige fat. *Cell*, 157:1292–1308.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470:279–283.
- Raj, P., Rai, E., Song, R., Khan, S., Wakeland, B. E., Viswanathan, K., Arana, C., Liang, C., Zhang, B., Dozmorov, I., Carr-Johnson, F., Mitrovic, M., Wiley, G. B., Kelly, J. A., Lauwerys, B. R., Olsen, N. J., Cotsapas, C., Garcia, C. K., Wise, C. A., Harley, J. B., Nath, S. K., James, J. A., Jacob, C. O., Tsao, B. P., Pasare, C., Karp, D. R., Li, Q. Z., Gaffney, P. M., and Wakeland, E. K. (2016). Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. *eLife*, 15:e12089.
- Reiman, E. M., Webster, J. A., Myers, A. J., Hardy, J., Dunckley, T., Zismann, V. L., Joshipura, K. D., Pearson, J. V., Hu-Lince, D., Huentelman, M. J., Craig, D. W., Coon, K. D., Liang, W. S., Herbert, R. H., Beach, T., Rohrer, K. C., Zhao, A. S., Leung, D., Bryden, L., Marlowe, L., Kaleem, M., Mastroeni, D., Grover, A., Heward, C. B., Ravid, R., Rogers, J., Hutton, M. L., Melquist, S., Petersen, R. C., Alexander, G. E., Caselli, R. J., Kukull, W., Papassotiropoulos, A., and Stephan, D. A. (2007). GAB2 alleles modify Alzheimer’s risk in APOE epsilon4 carriers. *Neuron*, 54:713–720.
- Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. H., Wilson, C. J., Bell, S. P., and Young, R. A. (2000). Genome-Wide Location and Function of DNA Binding Proteins. *Science*, 290:2306–2309.

- Rioux, J. D., Xavier, R. J., Taylor, K. D., Silverberg, M. S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M. M., Datta, L. W., Shugart, Y. Y., Griffiths, A. M., Targan, S. R., Ippoliti, A. F., Bernard, E. J., Mei, L., Nicolae, D. L., Regueiro, M., Schumm, L. P., Steinhardt, A. H., Rotter, J. I., Duerr, R. H., Cho, J. H., Daly, M. J., and Brant, S. R. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.*, 39:596–604.
- Roberts, A. W., Lee, B. L., Deguine, J., John, S., Shlomchik, M. J., and Barton, G. M. (2017). Tissue-Resident Macrophages Are Locally Programmed for Silent Clearance of Apoptotic Cells. *Immunity*, 47:810–812.
- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O. L., He, A., Marra, M., Snyder, M., and Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods*, 4:651–657.
- Roe, J., Hwang, C., Somerville, T. D. D., Milazzo, J. P., Lee, E. J., Da Silva, B., Maiorino, L., Tiriach, H., Young, C. M., Miyabayashi, K., Filippini, D., Creighton, B., Burkhart, R. A., Buscaglia, J. M., Kim, E. J., L., G. J., Lazenby, A. J., Grunkemeyer, J. A., Hollingsworth, M. A., Grandgenett, P. M., Egeblad, M., Park, Y., Tuveson, D. A., and Vakoc, C. R. (2017). Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell*, 170:823–825.
- Romanoski, C. E., Lee, S., Kim, M. J., Ingram-Drake, L., Plaisier, C. L., Yordanova, R., Tilford, C., Guan, B., He, A., Gargalovic, P. S., Kirchgessner, T. G., Berliner, J. A., and Lusk, A. J. (2010). Systems Genetics Analysis of Gene-by-Environment Interactions in Human Cells. *Am. J. Hum. Genet.*, 86:399–410.
- Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., Weaver, M., Shafer, A., Lee, K., Neri, F., Humbert, R., Singer, M. A., Richmond, T. A., Dorschner, M. O., McArthur, M., Hawrylycz, M., Green, R. D., Navas, P. A., Noble, W. S., and Stamatoyannopoulos, J. A. (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nature Methods*, 3:511–518.
- Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. *Genome Res.*, 22:1748–1759.
- Schaukowitch, K., Joo, J. Y., Liu, X., Watts, J. K., Martinez, C., and Kim, T. K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell*, 56:29–42.
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270:467–470.

- Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132:887–898.
- Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*, 352:1225–1228.
- Schwartz, S., Oren, R., and Ast, G. (2011). Detection and removal of biases in the analysis of next-generation sequencing reads. *PLoS ONE*, 6:e16685.
- Scott, C. L., Zheng, F., De Baetselier, P., Martens, L., Saeys, Y., De Prijck, S., Lippens, S., Abels, C., Schoonooghe, S., Raes, G., Devoogdt, N., Lambrecht, B. N., Beschin, A., and Guilliams, M. (2016). Bone marrow-derived monocytes give rise to self-renewing and fully differentiated Kupffer cells. *Nat. Commun.*, 7.
- Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., Erdos, M. R., Stringham, H. M., Chines, P. S., Jackson, A. U., Prokunina-Olsson, L., Ding, C. J., Swift, A. J., Narisu, N., Hu, T., Pruim, R., Xiao, R., Li, X. Y., Conneely, K. N., Riebow, N. L., Sprau, A. G., Tong, M., White, P. P., Hetrick, K. N., Barnhart, M. W., Bark, C. W., Goldstein, J. L., Watkins, L., Xiang, F., Saramies, J., Buchanan, T. A., Watanabe, R. M., Valle, T. T., Kinnunen, L., Abecasis, G. R., Pugh, E. W., Doheny, K. F., Bergman, R. N., Tuomilehto, J., Collins, F. S., and Boehnke, M. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, 316:1341–1345.
- Sheng, J., Ruedl, C., and Karjalainen, K. (2015). Most Tissue-Resident Macrophages Except Microglia Are Derived from Fetal Hematopoietic Stem Cells. *Immunity*, 43:382–393.
- Shiraishi, K., Okada, Y., Takahashi, A., Kamatani, Y., Momozawa, Y., Ashikawa, K., Kunitoh, H., Matsumoto, S., Takano, A., Shimizu, K., Goto, A., Tsuta, K., Watanabe, S., Ohe, Y., Watanabe, Y., Goto, Y., Nokihara, H., Furuta, K., Yoshida, A., Goto, K., Hishida, T., Tsuboi, M., Tsuchihara, K., Miyagi, Y., Nakayama, H., Yokose, T., Tanaka, K., Nagashima, T., Ohtaki, Y., Maeda, D., Imai, K., Minamiya, Y., Sakamoto, H., Saito, A., Shimada, Y., Sunami, K., Saito, M., Inazawa, J., Nakamura, Y., Yoshida, T., Yokota, J., Matsuda, F., Matsuo, K., Daigo, Y., Kubo, M., and Kohno, T. (2016). Association of variations in HLA class II and other loci with susceptibility to EGFR-mutated lung adenocarcinoma. *Nat. Commun.*, 7:12451.
- Simon-Sanchez, J., Scholz, S., Fung, H. C., Matarin, M., Hernandez, D., Gibbs, J. R., Britton, A., de Vrieze, F. W., Peckham, E., Gwinn-Hardy, K., Crawley, A., Keen, J. C., Nash, J., Borgaonkar, D., Hardy, J., and Singleton, A. (2007). Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum. Mol. Genet.*, 16:1–14.

- Smale, S. T. (2012). Transcriptional regulation in the innate immune system. *Curr. Opin. Immunol.*, 24:51–57.
- Soccio, R. E., Chen, E. R., Rajapurkar, S. R., Safabakhsh, P., Marinis, J. M., Dispirito, J. R., Emmett, M. J., Briggs, E. R., Fang, B., and Everett, L. J. (2015). Genetic variation determines PPAR $\gamma$  function and anti-diabetic drug response in vivo. *Cell*, 162:33–44.
- Soufi, A., Donahue, G., and Zaret, K. S. (2012). Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell*, 151:994–1004.
- Spann, N. J., Garmire, L. X., McDonald, J. G., Myers, D. S., Milne, S. B., Shibata, N., Reichart, D., Fox, J. N., Shaked, I., Heudobler, D., Raetz, C. R., Wang, E. W., Kelly, S. L., Sullards, M. C., Murphy, R. C., Merrill, A. H. J., Brown, H. A., Dennis, E. A., Li, A. C., Ley, K., Tsimikas, S., Fahy, E., Subramaniam, S., Quehenberger, O., Russell, D. W., and Glass, C. K. (2012). Regulated accumulation of desmosterol integrates macrophage lipid metabolism and inflammatory responses. *Cell*, 151:138–152.
- Step, S. E., Lim, H. W., Marinis, J. M., Prokesch, A., Steger, D. J., You, S. H., Won, K. L., and Lazar, M. A. (2014). Anti-diabetic rosiglitazone remodels the adipocyte transcriptome by redistributing transcription to PPAR $\gamma$ -driven enhancers. *Genes Dev.*, 28:1018–1028.
- Stergachis, A. B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A. P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., Thurman, R. E., Vong, S., Bates, D., Neri, F., Diegel, M., Giste, E., Dunn, D., Vierstra, J., Hansen, R. S., Johnson, A. K., Sabo, P. J., Wilken, M. S., Reh, T. A., Treuting, P. M., Kaul, R., Groudine, M., Bender, M. A., Borenstein, E., and Stamatoyannopoulos, J. A. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature*, 515:365–370.
- Sulem, P., Gudbjartsson, D. F., Walters, G. B., Helgadottir, H. T., Helgason, A., Gudjonsson, S. A., Zanon, C., Besenbacher, S., Bjornsdottir, G., Magnusson, O. T., Magnusson, G., Hjartarson, E., Saemundsdottir, J., Gylfason, A., Jonasdottir, A., Holm, H., Karason, A., Rafnar, T., Stefansson, H., Andreassen, O. A., Pedersen, J. H., Pack, A. I., de Visser, M. C., Kiemenev, L. A., Geirsson, A. J., Eyjolfsson, G. I., Olafsson, I., Kong, A., Masson, G., Jonsson, H., Thorsteinsdottir, U., Jonsdottir, I., and Stefansson, K. (2011). Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat. Genet.*, 43:1127–1130.
- Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keeffe, S., Haas, S., Vingron, M., Lehrach, H., and Yaspo, M. L. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 321:956–960.

- Takeda, K. and Akira, S. (2000). STAT family of transcription factors in cytokine-mediated biological responses. *Cytokine Growth Factor Rev.*, 11:199–207.
- Takeuchi, O. and Akira, S. (2010). Pattern recognition receptors and inflammation. *Cell*, 140:805–820.
- Teitelbaum, S. (2000). Bone resorption by osteoclasts. *Science*, 289:1504–1508.
- Trapnell, C., Prachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25:1105–1111.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, 28:511–515.
- Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, 270:484–487.
- Velusamy, T., Palanisamy, N., Kalyana-Sundaram, S., Sahasrabudhe, A. A., Maher, C. A., Robinson, D. R., Bahler, D. W., Cornell, T. T., Wilson, T. E., Lim, M. S., Chinnaiyan, A. M., and Elenitoba-Johnson, K. S. (2013). Recurrent reciprocal RNA chimera involving YPEL5 and PPP1CB in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.*, 110:3035–3040.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, K., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A.,

- Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The Sequence of the Human Genome. *Science*, 291:1304–1351.
- Vojinovic, D., Adams, H. H., van der Lee, S. J., Ibrahim-Verbaas, C. A., Brouwer, R., van den Hout, M. C., Oole, E., van Rooij, J., Uitterlinden, A., Hofman, A., van IJcken, W. F., Aartsma-Rus, A., van Ommen, G. B., Ikram, M. A., van Duijn, C. M., and Amin, N. (2015). The dystrophin gene and cognitive function in the general population. *Eur. J. Hum. Genet.*, 23:837–843.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M. U., Ohgi, K. A., Glass, C. K., Rosenfeld, M. G., and Fu, X. D. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, 474:390–394.
- Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., and Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456:470–476.
- Wang, Y., Hamet, P., Thorin, E., Tremblay, J., Raelson, J., Wu, Z., Luo, H., Jin, W., Lavoie, J. L., Peng, J., Marois-Blanchet, F. C., Tahir, M. R., Chalmers, J., Woodward, M., Harrap, S., Qi, S., Li, C. Y., and Wu, J. (2016). Reduced blood pressure after smooth muscle EFNB2 deletion and the potential association of EFNB2 mutation with human hypertension risk. *Eur. J. Hum. Genet.*, 24:1817–1825.
- Waszak, S. M., Delaneau, O., Gschwind, A. R., Kilpinen, H., Raghav, S. K., Witwicki, R. M., Orioli, A., Wiederkehr, M., Panousis, N. I., Yurovsky, A., Romano-Palumbo, L.,

- Planchon, A., Bielser, D., Padioleau, I., Udin, G., Thurnheer, S., Hacker, D., Hernandez, N., Reymond, A., Deplancke, B., and Dermitzakis, E. T. (2015). Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell*, 162:1039–1050.
- Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447:661–678.
- Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153:307–319.
- Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., and Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 453:1239–1243.
- Wright, J. R. (1990). Clearance and recycling of pulmonary surfactant. *Am J Physiol.*, 259:L1–L12.
- Wynn, T. A., Chawla, A., and Pollard, J. W. (2013). Macrophage biology in development, homeostasis, and disease. *Nature*, 496:445–455.
- Xu, S., Grullon, S., Ge, K., and Peng, W. (2014). Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) to Map Regions of Histone Methylation Patterns in Embryonic Stem Cells. *Methods Mol. Biol.*, 1150:97–111.
- Yamazaki, K., McGovern, D., Ragoussis, J., Paolucci, M., Butler, H., Jewell, D., Cardon, L., Takazoe, M., Tanaka, T., Ichimori, T., Saito, S., Sekine, A., Iida, A., Takahashi, A., Tsunoda, T., Lathrop, M., and Nakamura, Y. (2005). Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn’s disease. *Hum. Mol. Genet.*, 14:3499–3506.
- Yang, A., Zhu, Z., Kapranov, P., McKeon, F., Church, G. M., Gingeras, T. R., and Struhl, K. (2006). Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol. Cell*, 24:593–602.
- Yang, H., Schramek, D., Adam, R. C., Keyes, B. E., Wang, P., Zheng, D., and Fuchs, E. (2015). ETS family transcriptional regulators drive chromatin dynamics and malignancy in squamous cell carcinomas. *eLife*, 4:e10870.
- Yin, L. and Lazar, M. A. (2005). The orphan nuclear receptor Rev-erb $\alpha$  recruits the N-CoR/histone deacetylase 3 corepressor to regulate the circadian Bmal1 gene. *Mol. Endocrinol.*, 19:1452–1459.
- Zamir, I., Harding, H. G., Atkins, G. B., Hörlein, A., Glass, C. K., Rosenfeld, M. G., and Lazar, M. A. (1996). A nuclear hormone receptor corepressor mediates transcriptional silencing by receptors with distinct repression domains. *Mol. Cell. Biol.*, 16:5458–5465.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, B., Li, W., and Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.*, 9:R137.



# Glossary

|           |   |
|-----------|---|
| Ahr       | aryl hydrocarbon receptor.  |
| AP-1      | activator protein 1.  |
| ATAC-seq  | assay for transposase-accessible chromatin using sequencing.            |
| Batf      | Basic leucine zipper transcriptional factor ATF-like.                   |
| BMDM      | bone marrow-derived macrophage.   |
| C/EBP     | CCAAT-enhancer-binding protein.   |
| CAGE      | cap analysis of gene expression.  |
| cDNA      | complementary DNA.  |
| ChIP-chip | chromatin immunoprecipitation coupled with microarrays.                 |
| ChIP-seq  | chromatin immunoprecipitation coupled to massively parallel sequencing. |
| CRD       | connected regulatory domain.  |
| Csf1r     | colony stimulating factor 1 receptor.                                   |
| Cx3cr1    | CX3C chemokine receptor 1.  |
| DAMP      | damage-associated molecular patterns.                                   |
| DNA       | deoxyribonucleic acid.  |
| DNase I   | deoxyribonuclease I.  |
| DNase-seq | DNase I hypersensitivity sites sequencing.                              |
| EBF1      | early B-cell factor 1.  |
| ELK3      | ETS domain-containing protein 3.  |
| EMP       | erythromyeloid progenitor.  |
| eRNA      | enhancer RNA.   |
| EST       | expressed sequence tag.   |
| ETS       | E-twenty-six.   |
| F1        | filial 1 hybrid.  |

---

|              |   |
|--------------|---|
| FACS         | fluorescence activated cell sorting.                    |
| FAIRE        | formaldehyde-assisted isolation of regulatory elements. |
| FOXA1        | forkhead box protein A1.                                |
| GATA5        | GATA binding factor-5.                                  |
| GRO-seq      | global run-on sequencing.                               |
| GWAS         | genome-wide association studies.                        |
| H3K27ac      | acetylation of histone H3 lysine 27.                    |
| H3K27me3     | trimethylation of histone H3 on lysine 27.              |
| H3K4me1      | monomethylation of histone H3 lysine 4.                 |
| H3K4me2      | dimethylation of histone H3 lysine 4.                   |
| H3K4me3      | trimethylation of histone H3 lysine 4.                  |
| HDAC3        | histone deacetylase 3.                                  |
| HFS          | hair follicle stem cell.                                |
| HOMER        | Hypergeometric Optimization of Motif Enrichment.        |
| Id           | DNA-binding protein inhibitor.                          |
| IFN $\gamma$ | interferon gamma.                                       |
| IL           | interleukin.  |
| IRF          | interferon regulatory factor.                           |
| KLA          | Kdo2-lipid A.   |
| KLF          | Krüppel-like Factor.                                    |
| LDTF         | lineage-determining transcription factor.               |
| LPS          | lipopolysaccharide.                                     |
| LXR          | liver X receptor.                                       |
| MACS2        | Model-based Analysis of ChIP-Seq 2.                     |
| MAMP         | microbe-associated molecular patterns.                  |
| MAOa         | monoamine oxidase A.                                    |
| MARGE        | Mutation Analysis for Regulatory Genomic Elements.      |
| MEF2         | myocyte enhancer factor-2.                              |
| Mmp9         | matrix metalloproteinase 9.                             |
| MPSS         | massively parallel signature sequencing.                |
| mRNA         | messenger RNA.  |
| NCoR         | nuclear receptor co-repressor 1.                        |

---

|                |   |
|----------------|---|
| NF- $\kappa$ B | nuclear factor kappa-light-chain-enhancer of activated B cells. |
| NGS            | next-generation sequencing.                                     |
| Nr1h3          | nuclear receptor subfamily 1, group H, member 3.                |
| NRF            | nuclear factor (erythroid-derived 2)-like.                      |
| OCT            | octamer transcription factor.                                   |
| Pam3           | Pam3CSK4 - a synthetic triacylated lipopeptide.                 |
| PAX            | paired box protein.   |
| Pol II         | RNA polymerase II.  |
| Poly I:C       | polyinosinic-polycytidylic acid.                                |
| PPAR $\gamma$  | peroxisome proliferator-activated receptor gamma.               |
| PRRX2          | paired mesoderm homeobox protein 2.                             |
| qPCR           | quantitative polymerase chain reaction.                         |
| RNA            | ribonucleic acid.   |
| RNA-seq        | RNA sequencing.   |
| RNase H        | ribonuclease H.   |
| rRNA           | ribosomal RNA.  |
| SAGE           | serial analysis of gene expression.                             |
| Sall           | spalt like transcription factor.                                |
| SAM            | sympathetic neuron-associated macrophage.                       |
| SCC            | squamous cell carcinoma.  |
| SDTF           | signal-dependent transcription factor.                          |
| SE             | super enhancer.   |
| Slc6a2         | solute carrier family 6 member 2.                               |
| SMAD           | mothers against decapentaplegic homolog.                        |
| SNP            | single-nucleotide polymorphism.                                 |
| Sono-seq       | sonication of cross-linked chromatin sequencing.                |
| SOX            | sex determining region Y-box.                                   |
| SREBP1         | sterol regulatory element-binding protein 1.                    |
| STAT           | signal transducer and activator of transcription.               |
| TAC            | transit-amplifying cell.  |
| TAD            | topologically associating domain.                               |

|             |  |
|-------------|--|
| TF          | transcription factor.                                |
| TGEM        | thioglycollate-elicited peritoneal macrophage.       |
| TGF $\beta$ | transforming growth factor beta.                     |
| Timd4       | T-cell immunoglobulin and mucin domain containing 4. |
| TLR         | toll-like receptor.                                  |
| TNF         | tumor necrosis factor.                               |
| TT-seq      | transient transcriptome sequencing.                  |
| VCM         | variable chromatin module.                           |