bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genome-wide identification and analysis of transcribed enhancers during macrophage polarization

Elena Denisenko[1], Reto Guler[2,3], Musa Mhlanga[4,5,6], Harukazu Suzuki[7], Frank Brombacher[2,3], and Sebastian Schmeier[1*]

[1] Massey University, Institute of Natural and Mathematical Sciences, 0632 Albany, Auckland, New Zealand.

[2] Institute of Infectious Diseases and Molecular Medicine (IDM), Division of Immunology and South African Medical Research Council (SAMRC) Immunology of Infectious Diseases, Faculty of Health Sciences, University of Cape Town, Cape Town 7925, South Africa.

[3] International Centre for Genetic Engineering and Biotechnology (ICGEB), Cape Town Component, Cape Town 7925, South Africa.

[4] Gene Expression & Biophysics Group, CSIR Synthetic Biology ERA, Pretoria, 0001, South Africa.

[5] Institute of Infectious Diseases & Molecular Medicine (IDM), Division of Chemical Systems & Synthetic Biology, University of Cape Town, Cape Town 7925, South Africa.

[6] Gene Expression and Biophysics Unit, Instituto de Medicina Molecular, Faculdade de Medicina Universidade de Lisboa, Lisbon 1649-028, Portugal.

[7] Division of Genomic Technologies, RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan.

[*] Corresponding author: s.schmeier@massey.ac.nz

Email addresses:
Elena Denisenko: E.Denisenko@massey.ac.nz
Sebastian Schmeier: s.schmeier@massey.ac.nz
Reto Guler: Reto.Guler@uct.ac.za
Musa Mhlanga: musa@mhlangalab.org
Harukazu Suzuki: harukazu.suzuki@riken.jp
Frank Brombacher: frank.brombacher@icgeb.org

Abstract

Background

Macrophages are sentinel cells that play essential role in tissue homeostasis and host defence. Owing to their plasticity, macrophages acquire a range of functional phenotypes in response to microenvironmental stimuli, of which M1 (IFNγ) and M2 (IL-4/IL-13) phenotypes are well-known for their opposing pro- and anti-inflammatory roles. Enhancers have emerged as regulatory DNA elements crucial for transcriptional activation of gene expression, with recent studies highlighting their importance in macrophages.

Results

Using cap analysis of gene expression (CAGE) and epigenetic data, we identify on a large-scale transcribed enhancers in mouse macrophages, their time kinetics and target protein-coding genes. We observe an increase in target gene expression, concomitant with increasing numbers of associated enhancers and find that genes associated to many enhancers show a shift towards stronger enrichment for macrophage-specific biological

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

processes. We infer enhancers that drive transcriptional responses of genes upon cytokine-initiated M1 and M2 macrophage polarization and demonstrate stimuli-specificity of the regulatory associations. Finally, we show that enhancer regions are enriched for binding sites of inflammation-related transcription factors, suggesting a link between stimuli response and enhancer transcriptional control.

Conclusions

Our study provides new insights into genome-wide enhancer-mediated transcriptional control of macrophage genes, including those implicated in macrophage M1 and M2 polarization, and offers a detailed genome-wide catalogue to further elucidate enhancer regulation in macrophages.

Keywords

Transcriptional enhancers, eRNA, transcriptional regulation, transcriptome, macrophage polarization

Background

Macrophages are innate immune system sentinel cells that mediate homeostatic and protective functions, including host defence against invading pathogens [1]. Macrophages respond to a wide range of external stimuli by acquiring heterogeneous activation states that exert functional programs tailored for specific microenvironments [2]. Recently, a classification system for activated or polarized macrophages has been developed, but still remains controversial [2, 3]. Two functionally extreme subpopulations of activated macrophages are traditionally highlighted [4]. M1 or classically activated macrophages are pro-inflammatory macrophages that are polarized in response to particular stimuli including IFNγ, LPS and TNF [3]. M1 macrophages are characterized by efficient antigen presentation, high bactericidal activity and production of pro-inflammatory cytokines, reactive oxygen and nitrogen intermediates [3, 5]. M2 or alternatively activated macrophages are induced by stimuli such as IL-4 and IL-13 and are predominantly regulatory macrophages involved in homeostasis, angiogenesis, wound healing, tissue remodelling and parasitic and bacterial infection [1, 2, 6-8]. M2 macrophages release anti-inflammatory cytokines and show less efficient antigen presentation and decreased production of pro-inflammatory cytokines [2, 6]. Macrophage polarization towards M1 or M2 phenotypes is driven by specific transcriptional changes and is controlled by complex cellular mechanisms [3, 9].

Imbalance in M1 and M2 populations of macrophages with opposing pro- and anti-inflammatory roles has been implicated in disease progression [1]. Intracellular pathogen *Mycobacterium tuberculosis*, the causative agent of tuberculosis, interferes with M1 polarization of macrophages and promotes M2 polarization, thereby avoiding anti-bacterial action of M1 macrophages [10, 11]. Tumour microenvironments promote phenotypic switches from M1 to M2 macrophages, which might contribute to the tumour progression by inhibiting immune responses to tumour antigens [1, 2]. Conversely, the phenotypic switch from a M2 population of macrophages to M1 might contribute to obesity and metabolic syndrome [1, 2, 12]. Therefore, the development of techniques for manipulation and specific targeting of macrophage populations could ultimately improve diagnosis and treatment of inflammatory diseases. [2]. To advance this area of research, the cellular mechanisms responsible for macrophage polarization need to be further deciphered.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Gene expression in eukaryotic cells is a complex process guided by a multitude of mechanisms [13]. Precise regulation is required to ensure dynamic control of tissue-specific gene expression and to fine tune the responses to external stimuli [14]. One such level of control is facilitated via regulation of RNA transcription. This process is mediated by a complex transcriptional machinery with its components recognising specific regulatory regions of DNA. Promoters represent a better-characterized class of such regions from which RNA transcription is initiated [15, 16]. They act in concert with other cis-regulatory DNA elements, including enhancers, which are believed to play key roles in transcriptional regulation [17].

Enhancers are defined as regulatory DNA regions that activate transcription of target genes in a distance- and orientation-independent manner [17]. According to the dominant model, transcriptional regulation by enhancers is exerted via direct physical interaction between enhancer and target gene promoter mediated by DNA looping [17, 18]. Recent identification of distinct properties of enhancer regions enabled novel approaches to enhancer profiling [17]. Enhancer regions are often distinguished by a specific combination of chromatin marks present at these locations, such as H3K4me1 and H3K27ac [19, 20]. Enhancer sequences contain transcription factor binding sites (TFBS) that recruit transcription factors (TFs) to regulate target genes [21, 22]. In addition, enhancers are frequently bound by proteins such as histone acetyltransferase p300 and insulator-binding protein CTCF [20, 23-25]. Large-scale profiling of these enhancer-associated signatures by chromatin immunoprecipitation followed by sequencing (ChIP-seq) [25, 26] has greatly advanced enhancer identification and enabled systematic and genome-wide enhancer mapping [27, 28]. Another group of methods such as chromosome conformation capture (3C) [29] and its variant Hi-C [30] has been employed to profile physical DNA contacts, including those between promoters and enhancers [31, 32]. However, none of these methods has become a gold standard of enhancer detection, and the field is still actively developing.

Recent studies have led to the unexpected finding that most active enhancers recruit RNA polymerase II and are bi-directionally and divergently transcribed to produce RNA transcripts, referred to as eRNAs [33, 34]. While the functionality of eRNA remains controversial, a recent study by Hon et al. showed that many enhancers are transcribed into potentially functional long-noncoding RNAs (lncRNAs) playing a role in inflammation and immunity [35, 36]. Recently, quantification of eRNA transcription laid the foundation for a novel method of large-scale enhancer profiling [37]. In their study utilizing cap analysis of gene expression (CAGE) [38], Andersson et al. performed genome-wide mapping of transcriptional events followed by identification of enhancers based on co-occurrence of closely located divergent transcripts representing eRNAs [37]. The capacity of CAGE to simultaneously profile the expression of eRNAs and genes became an additional advantage, since eRNA production was shown to positively correlate with the production of mRNAs of target genes [33, 39].

These and other studies unravelled the fundamental importance of enhancer regions as DNA regulatory elements in multiple cell types, including macrophages [27, 28, 34, 39, 40]. Enhancers are extremely widespread, with an estimation of up to one million enhancers in mammalian genomes [19, 22, 23, 41]. They are major determinants of gene expression programs required for establishing cell type specificity and mediating response to extracellular signals [22, 42, 43]. Our current understanding of these elements, however, remains incomplete. High tissue-specificity of enhancers is a major hurdle towards establishing a comprehensive catalogue of the full enhancer population [22, 42]. Moreover, emerging evidence indicate that enhancers

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

selectively act in a stimuli- or condition-specific manner [44, 45]. A major challenge is, therefore, to catalogue enhancers active in different tissues and conditions and link them to target genes.

Recently, we investigated the transcriptional regulatory dynamics of protein-coding and lncRNA genes during M1 (IFNγ) and M2 (IL-4/IL-13) macrophage polarization using CAGE data [9]. We showed that particular TFs, such as Nfκb1, Rel, Rela, Irf1 and Irf2, drive macrophage polarization and are commonly activated but have distinct dynamics in M1 and M2 macrophages [9]. Here, we extended the former study to understand the regulatory influence of enhancers in the macrophage activation process. Our genome-wide *in silico* study aimed at characterizing the enhancer landscape in mouse macrophages and studying its dynamic changes during M1 and M2 polarization. We used CAGE data and enhancer-associated chromatin signature to identify enhancer regions. We inferred regulatory associations between enhancers and target protein-coding genes using their spatial organisation in topologically associating domains (TADs) [46] and correlation of CAGE-derived expression in our time-course. With these data, we established a catalogue of transcribed enhancer regions linked to their target genes. This catalogue provides insights into genome-wide enhancer-mediated regulation of transcription in mouse macrophages. Furthermore, we highlight the role of enhancers during macrophage polarization and report enhancers driving expression dynamics of known M1 and M2 activation marker genes.

Results

Identification of transcribed mouse macrophage enhancers

Active enhancers were shown to be bi-directionally transcribed in mammals [33, 34], and eRNAs profiled by CAGE technology [38] were used before to reliably infer enhancer regions in human [37]. To identify transcribed enhancers in mouse tissue, we used the FANTOM5 collection of CAGE mouse samples [16] and a similar strategy as developed before [37] (see Methods). This approach yielded 42,470 mouse enhancers, with 17,752 enhancers deemed transcribed in our macrophage samples (Fig. 1a, Methods). To refine this set, we sub-selected 11,216 (63%) transcribed enhancers that carry enhancer-specific chromatin signatures (Fig. 1a), as determined by ChIP-seq in mouse macrophages [47] (see Methods). Notably, of all mouse enhancers not transcribed in macrophages, only 19% carry macrophage enhancer chromatin signatures, highlighting the specificity of enhancers in mouse tissues.

Macrophage enhancer-gene interactome

We aimed at studying enhancers that regulate expression of protein-coding genes in macrophages. We first identified pairs of enhancers and promoters located within TADs [46], since this regulation is thought to be exerted via direct enhancer-promoter contact [17, 18]. Thereafter, we refined these pairs using CAGE expression data based on the observation that eRNA and their target expression are positively correlated [33] (see Methods). This yielded 222,870 TAD-based enhancer-promoter (E-P) pairs, with 64,891 pairs showing significant positive correlation of expression in macrophages (Fig. 1a). These correlation-based regulatory associations formed the basis for our further analyses and included 8,667 enhancers deemed active in mouse macrophages. Interestingly, most of the TAD-based E-P pairs showed positive expression correlation (Additional file 1: Figure S1a), which supports the definition of a TAD as a structural unit favouring internal regulatory interactions [48]. Our filtering approach further selected regulatory associations with the highest correlation (Additional file 1: Figure S1a), which we considered more reliable. The median distance between

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

enhancers and promoters in the correlation-based E-P pairs was significantly smaller at 191,033nt as compared to 278,735nt for all TAD-based pairs (Additional file 1: Figure S1b).

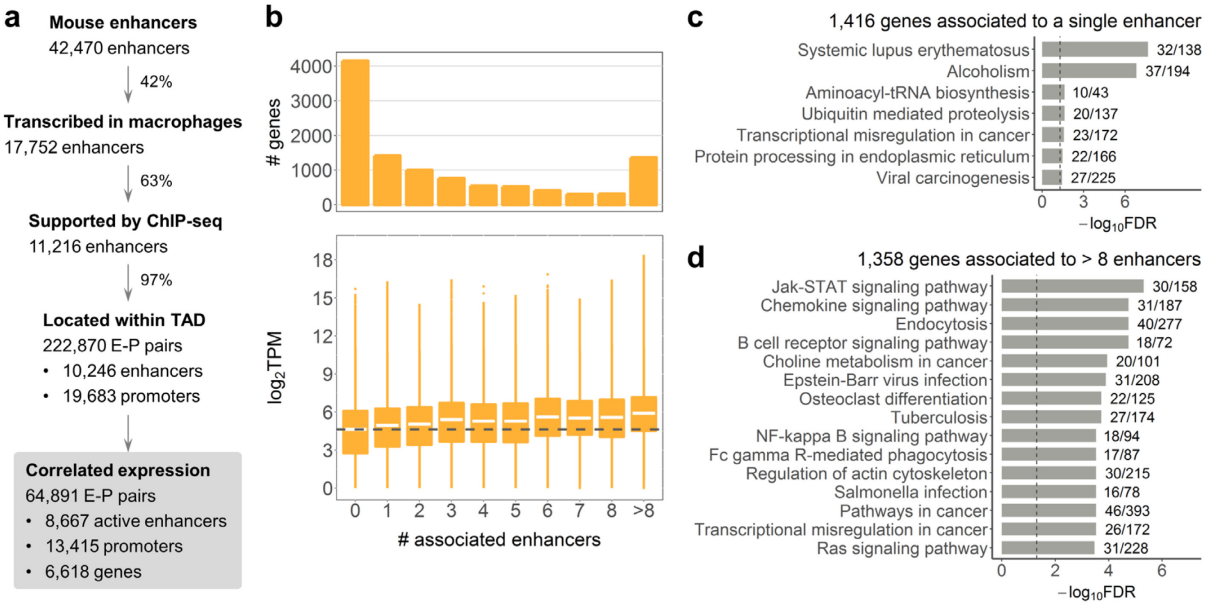

Fig. 1 Macrophage enhancer-gene interactome. **a** Pipeline for identification of enhancers and enhancer-promoter associations. **b** Number and expression of genes associated to different number of enhancers. Dashed line shows median expression of genes not associated to any enhancer. **c** KEGG pathway maps significantly enriched for genes associated to a single enhancer, FDR < 0.05. **d** Top 15 KEGG pathway maps with the lowest FDR enriched for genes associated to more than 8 enhancers. In **c** and **d** next to the bars are the numbers of genes in the KEGG pathway covered by our gene list; dashed lines indicate FDR = 0.05.

We further investigated associations between enhancers and target protein-coding genes (Additional file 2). Of all 10,767 protein-coding genes with CAGE expression (see Methods), 4,149 genes (38.5%) were not associated to any enhancer (Fig. 1b, upper panel). Given previous evidence of additive action of enhancers [17, 49], we asked whether genes regulated by different numbers of enhancers have different gene expression levels. Genes without associated enhancers were overall lower expressed than genes associated to one (two-sided Wilcoxon signed-rank test p-value $< 2.2*10^{-16}$) or more enhancers. A steady increase in gene expression concomitant with higher numbers of associated enhancers (Fig. 1b, Kruskal-Wallis rank sum test p-value $< 2.2*10^{-16}$) was observed, supporting the model of additive enhancer action.

We further asked whether genes associated to different numbers of enhancers within the enhancer-gene interactome show functional differences. Gene set enrichment analysis (GSEA) was performed for gene sets of similar size to avoid a size-related bias (see Methods). The 1,416 genes associated to a single enhancer were enriched for general cellular pathways including "Aminoacyl-tRNA biosynthesis" and "Ubiquitin mediated proteolysis", as well as a few inflammation-related pathways (Fig. 1c). However, in contrast, the 1,358 genes associated to more than eight enhancers showed stronger enrichment for macrophage-related terms, such as "Jak-STAT signaling pathway" and "Chemokine signaling pathway" (Fig. 1d). GSEA for 1,306 genes associated to three or four enhancers showed enrichment for a combination of general cellular and macrophage-related pathways (Additional file 3). Finally, the larger set of 4,149 genes not associated to any enhancer

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

showed the strongest enrichment for general cellular pathways (Additional file 4). Hence, a shift towards stronger enrichment for macrophage-related pathways was a concomitant of higher numbers of associated enhancers.

Macrophage-specific expression

We opted for a similar strategy as Yao et al. [50] (see Methods) to uncover eRNAs and genes with higher expression in macrophages as compared to other FANTOM5 mouse tissues (further referred to as macrophage-specific). We identified 1,844 macrophage-specific and 8,923 non-macrophage-specific genes (Additional file 1: Figure S2). These two sets showed differences in numbers of associated enhancers, with 65.6% of macrophage-specific genes being associated to more than one enhancer, whereas this proportion dropped to 44.7% for non-macrophage-specific genes (Fig. 2a). These results were in agreement with our observation of stronger enrichment for macrophage-related functions in genes associated to many enhancers. Similar to the trend observed above, both macrophage-specific and non-macrophage-specific genes showed higher gene expression concomitant with higher numbers of associated enhancers, with non-macrophage-specific genes showing lower expression levels than macrophage-specific ones (Additional file 1: Figure S3).

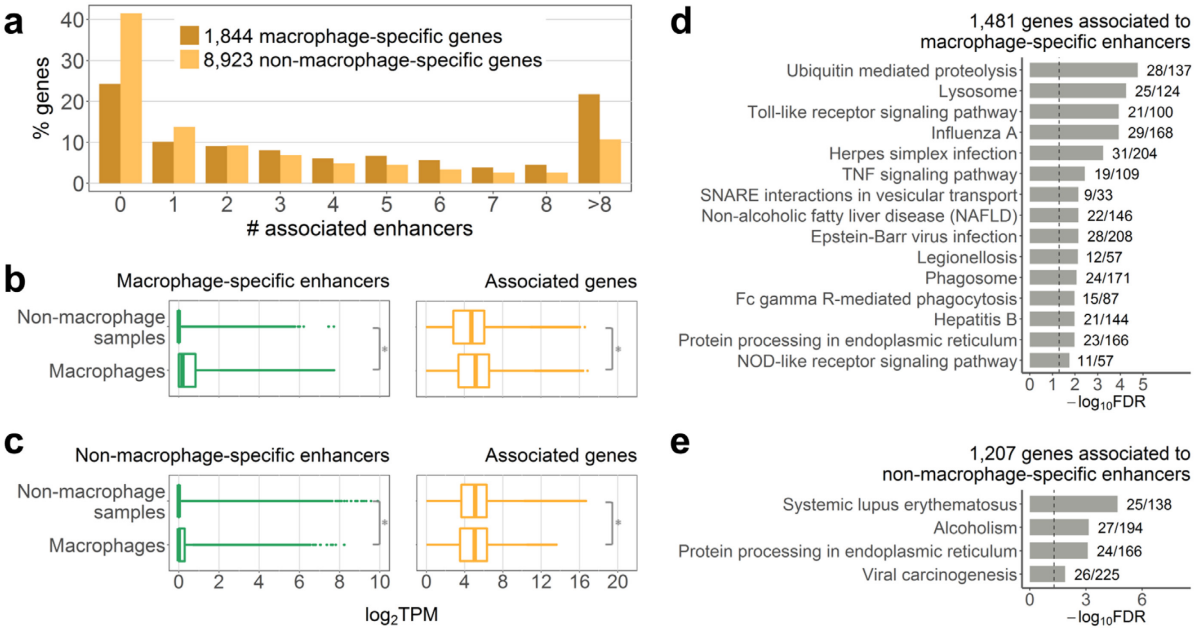

Fig. 2 Macrophage-specific enhancer and gene expression. **a** Percentage of genes associated to different number of enhancers. **b** Expression of 4,739 macrophage-specific enhancer eRNAs and 1,481 associated genes. **c** Expression of 3,928 non-macrophage-specific enhancer eRNAs and 1,207 associated genes. In **b** and **c** expression is shown in 184 macrophage and 774 non-macrophage samples, asterisks denote Wilcoxon rank sum test p-value < $2.2*10^{-16}$. **d** Top 15 KEGG pathway maps significantly enriched for genes associated exclusively to macrophage-specific enhancers. **e** KEGG pathway maps enriched for genes associated exclusively to non-macrophage-specific enhancers with FDR < 0.05. In **d** and **e** next to the bars are the numbers of genes in the KEGG pathway covered by our gene list; dashed lines indicate FDR = 0.05.

Among 8,667 active enhancers, 54.7% were deemed macrophage-specific (see Methods), in agreement with known tissue-specificity of enhancers [22, 41, 42]. Interestingly, non-macrophage-specific enhancers still showed higher eRNA expression in macrophages as compared to the non-macrophage samples (Fig. 2c, left

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

panel). This may be explained by the fact that for this analysis we excluded all enhancers that showed zero eRNA expression in the majority of our macrophage samples (see Methods).

Next, we asked whether these two enhancer sets could regulate genes with different functions. Genes associated exclusively to macrophage-specific enhancers, as well as genes associated exclusively to non-macrophage-specific enhancers were sub-selected. As expected, genes in the former set showed overall higher expression in macrophage samples as compared to the non-macrophage samples (Fig. 2b, right panel). In contrast, expression of genes associated exclusively to non-macrophage-specific enhancers was lower in macrophage samples. Interestingly, the opposite was observed for non-macrophage-specific enhancers (Fig. 2c). Genes associated to macrophage-specific enhancers were enriched for both general and macrophage-related processes (Fig. 2d). This observation reflects the fact that production of macrophage-specific factors and activation of housekeeping processes that facilitate it might be both regulated by the same set of enhancers. Genes associated to non-macrophage-specific enhancers were enriched for only four KEGG pathway maps with FDR < 0.05 (Fig. 2e), none of which can be considered a typical macrophage pathway. We obtained consistent results when we repeated the analysis for a subset of 500 genes with the highest expression in macrophages (Additional files 5-6). Taken together, these findings demonstrate that most of the identified active enhancers in macrophages show macrophage-specific eRNA expression and regulate genes with macrophage-specific as well as general cellular functions.

Stimuli-induced transcriptional changes

We set out to determine transcriptional changes that were dynamically induced by IFNγ (M1) and IL-4/IL-13 (M2) in mouse macrophages, and to infer enhancers important in these processes (Fig. 3a). M1- and M2-responsive enhancers and genes were identified as those up-regulated upon stimulation; regulatory associations were retained for pairs with a positive correlation of expression in the corresponding polarization state (see Methods). In this manner, we discovered 115 M1-responsive enhancers regulating 105 M1-responsive genes (further referred to as sets E1 and G1), as well as 131 M2-responsive enhancers regulating 98 M2-responsive genes (sets E2 and G2) (Fig. 3b and Additional files 8-9). Notably, 77% of E1 and 71% of E2 enhancers were deemed macrophage-specific in our settings. GSEA of G1 and G2 gene sets showed significant enrichment for GO and KEGG terms relevant to immune system and macrophage functions (Fig. 3c and Additional file 1: Figure S4). These results highlight the importance of enhancer regulatory control during macrophage polarization and suggest a striking influence of cytokine stimulation on activation of enhancers, which, in turn, drive some of the transcriptional responses seen during M1 and M2 polarization.

M1 and M2 macrophages are known to possess different phenotypes and functions [2]. As expected, G1 and G2 sets had only 19 genes in common. Similarly, a small overlap of only 14 enhancers was observed for E1 and E2 sets. Moreover, enhancers and genes selected as stimuli-responsive for a single activation state showed significant differences in time-course expression in M1 and M2 macrophages (Fig. 3d). These data indicate that M1 and M2 macrophages not only differ in their gene expression profiles, but also differ in their active enhancer repertoire that likely drives observed gene expression changes.

Previous studies reported and exploited positive expression correlation of eRNA and target genes [33, 37, 50]. Hence, we compared expression correlation of E1-G1 and E2-G2 pairs in M1 and M2 macrophages (Fig. 3e) to

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

determine how correlations differ between conditions. E1-G1 pairs showed higher correlation in M1 macrophages as compared to M2 (two-sided Wilcoxon signed-rank test p-value = $1.633*10^{-6}$). Similarly, correlation for E2-G2 pairs was higher in M2 macrophages (two-sided Wilcoxon signed-rank test p-value < $2.2*10^{-16}$). Such stimuli-specific expression correlation suggests stimuli-specificity of enhancer-gene regulatory associations in macrophages.

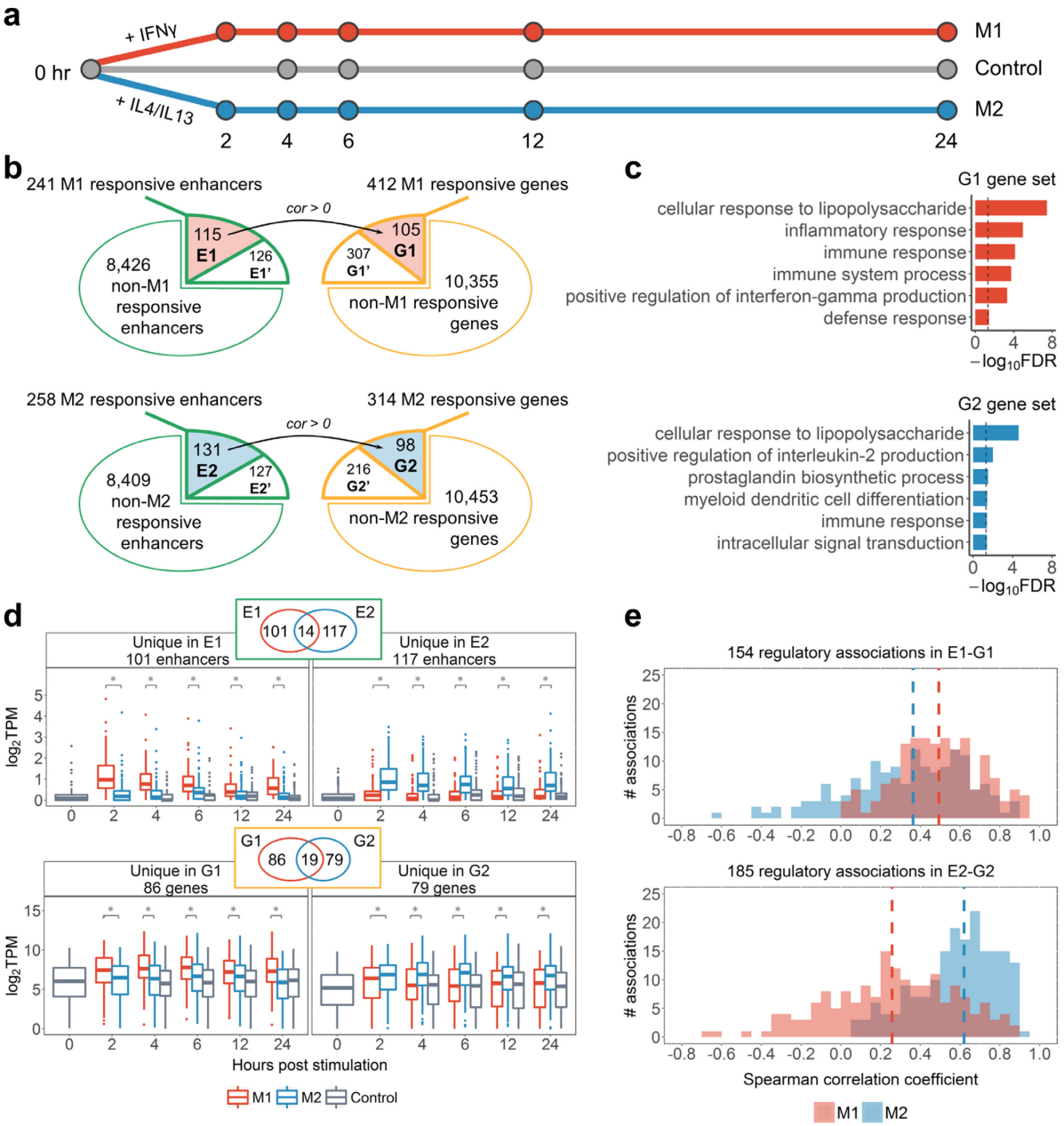

Fig. 3 Stimuli-responsive genes and enhancers. **a** Time-course data used in this study. **b** Enhancer and gene sets. E1 and E2: M1- and M2-responsive enhancers regulating M1- and M2-responsive genes (G1 and G2), respectively; E1' and E2': M1- and M2-responsive enhancers regulating non-stimuli-responsive genes; G1' and G2': M1- and M2-responsive genes not regulated by stimuli-responsive enhancers. Black arrows denote regulatory associations between stimuli-responsive enhancers and genes. **c** GO biological process terms enriched for G1 and G2 genes (all terms with FDR < 0.05 for G1; six terms with the lowest FDR for G2 are shown); dashed lines indicate FDR = 0.05. **d** Expression of stimuli-responsive enhancer eRNAs (upper panel) and genes (lower panel) unique to M1 and M2. Statistical significance was determined by Wilcoxon signed-rank test, asterisks indicate p-value < 10^{-5}. **e** Correlation of time-course expression of M1-responsive (upper panel) and M2-responsive (lower panel) enhancers and genes. Vertical dashed lines show median values.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Marker genes of macrophage activation are regulated by stimuli-responsive enhancers

We further asked which known marker genes of macrophage activation [1-3, 6, 51] were identified in our setting (Table 1). Out of 20 M1 and 26 M2 marker genes, we found eight M1 marker genes in the G1 set and eight M2 marker genes in the G2 set, that were associated to responsive enhancers in the respective stimulation (significant overlap with hypergeometric test p-value < 10^{-10}) (Table 1). The G1' set contained an additional four M1 marker genes (Gpr18, Il12b, Il6, Inhba) and the G2' set an additional three M2 marker genes (Il27ra, Klf4, Myc), which, although stimuli-responsive themselves, were not associated to stimuli-responsive enhancers.

Table 1 M1 and M2 macrophage polarization markers.

M1 markers			M2 markers		
Gene (G1)	***# enh. in E1***	***M1-specific enhancers in E1***	***Gene (G2)***	***# enh. in E2***	***M2-specific enhancers in E2***
Cd38	1		Arg1	1	chr10:25119065..25119466
Cxcl9	3	chr5:92368373..92368774, chr5:92369052..92369453, chr5:92374704..92375105	Ccl24	1	
Cxcl10	4	chr5:92353639..92354040, chr5:92368373..92368774, chr5:92369052..92369453, chr5:92374704..92375105	Egr2	9	chr10:67595184..67595585, chr10:67598488..67598889, chr10:67628888..67629289, chr10:67636538..67636939, chr10:67694800..67695201, chr10:67695848..67696249, chr10:67712611..67713012, chr10:67713071..67713472, chr10:67715029..67715430
Cxcl11	5	chr5:92353639..92354040, chr5:92368373..92368774, chr5:92369052..92369453, chr5:92374704..92375105, chr5:92375350..92375751	Fn1	3	chr1:71938511..71938912,
Nos2	1	chr11:78916390..78916791	Igf1	7	chr10:87731929..87732330, chr10:87753519..87753920, chr10:87805812..87806213, chr10:87830718..87831119, chr10:87832100..87832501, chr10:87839444..87839845
Ptgs2	1		Irf4	2	chr13:30714614..30715015
Socs3	3		Mrc1	2	chr2:14185406..14185807, chr2:14206798..14207199
Tnf	1		Socs2	2	chr10:95232562..95232963, chr10:95236240..95236641

The columns list marker genes in G1 and G2, number of associated enhancers in the corresponding activation state, and potential polarization marker enhancers.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Next we inferred potential marker enhancers that regulate marker genes specifically during M1 or M2 polarization. Each of the 16 M1 and M2 marker genes in G1 and G2 was associated to a minimum of one and maximum of nine enhancers in the E1 and E2 stimuli-responsive sets, respectively (Table 1). Of those, we identified enhancers that were selectively responsive in a single polarization state and showed higher expression in this state as compared to the other polarization (see Methods). A total of 13 M1 and 22 M2 enhancers were inferred as potential polarization markers (Table 1).

Interestingly, three of the M1 marker genes, Cxcl9, Cxcl10, and Cxcl11 are located within one TAD and are co-regulated by a group of three marker enhancers (Fig. 4a-c). These enhancers, along with the two marker enhancers regulating Cxcl10 or Cxcl11 but not Cxcl9 (Table 1) are located in close proximity, in the intronic regions of the Art3 gene (Fig. 4c). These enhancer regions were previously reported to show induced RNA polymerase II binding in macrophages upon stimulation with LPS, one of the known M1 activators [34]. In addition, these marker enhancer regions were shown to carry H3K4me1 enhancer histone marks in untreated macrophages [47]. Moreover, H3K27ac modification, associated with active enhancers, is stronger enriched in these regions in M1 as compared to M2 and untreated macrophages [47] (Fig. 4c), providing further evidence of their functionality in macrophage M1 polarization.

Among M2 marker genes, Arg1 as expected, is substantially expressed in M2 polarized macrophages but has extremely low expression in M1 and untreated macrophages (Additional file 1: Figure S5a). We found a single M2-responsive enhancer that might drive expression of Arg1 in M2 macrophages and might serve as a marker enhancer (Table 1 and Additional file 1: Figure S5b). On the contrary, M2 marker gene Egr2, a TF that activates macrophage-specific genes [52], is associated to as many as nine M2-specific enhancers (Table 1). Egr2 showed immediate up-regulation in response to both IFNγ and IL-4/IL-13 stimulation, however, in M2 macrophages the up-regulation sustained for up to 24 hours, whereas in M1 macrophages expression dropped rapidly after 2 hours (Fig. 4d, upper panel). Time-course eRNA expression for two Egr2 marker enhancers with the highest expression at 2 and 4 hours is shown in Fig. 4d. The distribution of all nine Egr2 marker enhancers within a TAD (Fig. 4e) may suggest that the regions identified as nine individual enhancers potentially demarcate fewer regions of stretch enhancers [53, 54]. We observed a similar distribution for enhancers of M2 marker gene Igf1, which is known to shape the M2 macrophage phenotype and regulate immune metabolism [55] (Additional file 1: Figure S6). Importantly, in both Egr2 and Igf1, marker enhancer regions carried H3K4me1 in untreated macrophages and showed the strongest enrichment with H3K27ac in M2 as compared to M1 and untreated macrophages [47] (Fig. 4e and Additional file 1: Figure S6c).

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

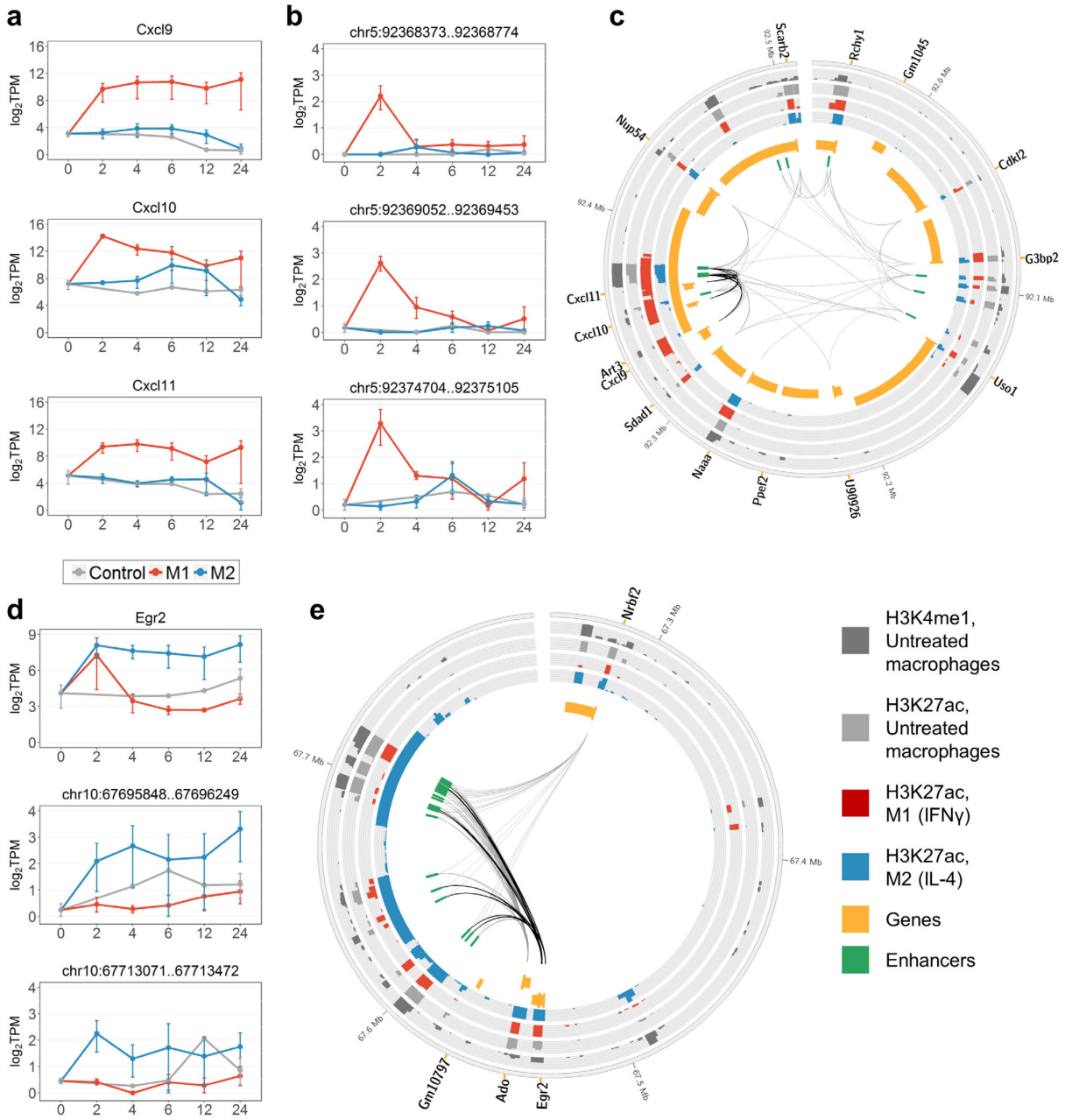

Fig. 4 Examples of M1 and M2 marker genes and enhancers. **a** Expression of M1 marker genes Cxcl9, Cxcl10 and Cxcl11. **b** eRNA expression of three potential marker enhancers that co-regulate Cxcl9, Cxcl10 and Cxcl11. **c** Genomic region of a TAD containing Cxcl9, Cxcl10, Cxcl11, and associated enhancers. Black links connect the marker genes with the three potential marker enhancers. Grey links denote other enhancer-gene interactions that we identified in macrophages. **d** Expression of M2 marker gene Egr2 and two of M2 marker enhancers associated to Egr2. **e** Genomic region of a TAD containing Egr2 and associated enhancers. Black links connect Egr2 with the nine M2 marker enhancers. Grey links denote other enhancer-gene interactions that we identified in macrophages. In **a**, **b** and **d** data were averaged over replicates and log-transformed. Error bars are the SEM. In **c** and **e** genes are split into two tracks based on the strand, wide orange marks denote gene promoters; histone mark tracks show ChIP-seq peaks with the height of -10*log10(p-value) (data from [47]).

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Transcription factor binding sites are enriched in enhancer regions

To investigate whether our enhancer sets are enriched for known TFBS, we performed an over-representation analysis of experimentally determined protein DNA binding sites established through ChIP-seq [56, 57] (see Methods). The sets of macrophage-specific and non-macrophage-specific enhancers are both enriched for binding sites of general factors (p300, Tbp), as well as a range of TFs with well-established roles in macrophages, such as macrophage lineage-determining factor Spi1 (PU.1) [40, 58], Cebpb, required for macrophage polarization [59], and Rela, regulating inflammatory genes [60] (Additional file 7). Interestingly, transcription binding sites (TFBS) for Spi1 overlap 54.1% of macrophage-specific enhancers, but only 38% of non-macrophage-specific enhancers (overlap ratio of 1.4 for macrophage-specific/non-macrophage-specific enhancers). We observed similar and higher overlap ratios for other functionally important TFs in macrophages, including Stat1, Rela, Irf1, Junb, and Cebpb [3, 59-61] (Table 2).

Table 2 TFs regulating more macrophage-specific than non-macrophage-specific enhancers.

TF	mean TPM in macrophages	% Macrophage-specific enhancers	% Non-macrophage-specific enhancers	Ratio
Stat1	284.6	15.8	8.5	1.9
Rela	130.7	26.6	14.5	1.8
Atf4	329.4	25.5	14.3	1.8
Irf1	508.7	21.7	12.2	1.8
Junb	176.3	11.2	7.4	1.5
Spi1	599.9	54.1	38	1.4
Cebpb	647.9	48	34.8	1.4

Columns show TF name, average expression in macrophage samples (in TPM, see Methods), percentage of enhancers overlapping corresponding binding sites, and a macrophage-specific/non-macrophage-specific percentage ratio. TFBS are statistically significantly enriched in both enhancer sets.

Similarly, the E1 and E2 stimuli-responsive enhancer sets are enriched for TFBS of known macrophage TFs including Spi1, Cebpb, Rela, and Irf and Stat families [3, 62, 63] (Table 3, Additional file 10). Interestingly, TFBS of Stat1, Rela and Irf1, involved in M1 polarization [3, 64], overlap a higher percentage of E1 enhancers as compared to E2 and macrophage-specific enhancers (Tables 2-3, and Additional files 7 and 10). For instance, Irf1 TFBS overlap 21.7% of macrophage-specific enhancers, 26.7% of E2 but 44.3% of E1 enhancers. In addition, the expression of genes encoding these TFs is higher in M1 as compared to M2 macrophages. Taken together, these results provide an additional layer of support for our regions as functionally important macrophage enhancers and implicate key macrophage TFs in modulating their activity. These findings further reflect that enhancers are selectively activated depending on the transcriptional machinery involved in the cellular response.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3 TFs with binding sites enriched in both E1 and E2 enhancer sets.

TF	mean TPM in M1	% enhancers in M1	mean TPM in M2	% enhancers in M2	Ratio
Stat1	726.3	40.9	218.6	22.9	1.8
Irf1	1345.4	44.3	426.7	26.7	1.7
Ets1	1.4	39.1	7	25.2	1.6
Jun	74.3	14.8	70	9.9	1.5
Rela	131.6	42.6	102.8	29.8	1.4
Atf4	277.3	33.9	208.3	27.5	1.2
Junb	122	13	99.7	13	1.0
Irf4	2.3	13	18.7	13.7	0.9
Spi1	766.3	52.2	681.7	56.5	0.9
Cebpb	424.6	51.3	262.9	59.5	0.9
Atf3	233.9	9.6	242.9	12.2	0.8

Columns show TF name, average expression in M1 and M2 macrophages (in TPM, see Methods), percentage of enhancers overlapping corresponding binding sites, and a M1/M2 percentage ratio.

Discussion

In this study we investigated the enhancer landscape in mouse macrophages and its dynamic changes during M1 and M2 polarization. Using CAGE data combined with ChIP-seq, we identified 8,667 active enhancers forming 64,891 regulatory associations with protein-coding gene promoters in mouse macrophages. We highlighted tissue- and stimuli-specificity of both enhancers and their regulatory interactions. The enhancer-gene interactome established here supports a model of additive action of enhancers [17, 49], with higher gene expression concomitant with higher numbers of associated enhancers. Moreover, we observed a shift towards stronger enrichment for macrophage-specific biological processes in genes associated to many enhancers. Cytokine stimulation in macrophages had a striking influence on enhancer activation, which highlights the importance of enhancers in polarization towards M1 and M2 phenotypes. In addition, we inferred potential stimuli-specific marker enhancers. Finally, we find that binding sites of inflammatory TFs are enriched in enhancer regions, proposing a link between the response to stimuli and enhancer transcriptional activation.

We used a combination of two complementary data types, transcriptomic and epigenomic data, to infer more reliable active enhancer regions in mouse macrophages. Instead of a widely used linear proximity-based approach [34, 37, 44], we employed TAD data for the inference of enhancer-gene associations. Accumulating evidence suggests that linear proximity might not be an accurate predictor of enhancer-gene interactions, as many enhancers regulate distal genes, bypassing the nearest promoter [65, 66]. At the same time, TADs have emerged as units of chromatin organisation that favour internal DNA contacts [48], and the majority of characterized interactions between enhancers and target promoters occur within the same TAD [48, 66, 67].

TADs are well conserved across cell types, whereas active enhancers and enhancer-promoter regulatory interactions within TADs vary between tissues and conditions [42, 65]. Different populations of tissue macrophages were shown to be highly heterogeneous and to possess distinct sets of active enhancers, as defined

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by ChIP-seq profiling of histone modifications [27, 28]. Kaikkonen et al. used H3K4me2 ChIP-seq to identify a set of enhancers that is established in macrophages *de novo* in response to the stimulation with TLR4 agonist, highlighting stimuli-specific action of enhancers in macrophages [39]. These and other studies have stressed the complexity of enhancer action in macrophages and other cell types, which yet remains to be fully characterized. In our study, we inferred 8,667 active enhancers in mouse macrophages. Of these, 70% overlap RNA polymerase II ChIP-seq peaks in untreated mouse macrophages [47]. Our enhancer regions show significant enrichment for binding sites of histone acetyltransferase p300, an enhancer-associated marker [25], and known inflammatory TFs. Hence, the regions identified here show a range of known enhancer properties, generally supporting our approach. Most of the active enhancers show macrophage-specific eRNA expression, in line with known tissue-specificity of enhancers [22, 41, 42].

Comparing our enhancer set to previous studies of mouse enhancers, we found for example that 39.8% of our enhancers overlap a set of *cis*-regulatory elements from 19 non-macrophage mouse tissues identified by Shen et al. [41]. In another recent study, Schoenfelder et al. employed a Capture Hi-C approach to identify enhancers and their target promoters in mouse fetal liver cells and embryonic stem cells [32]. These data support only 24.8% of our 64,891 E-P pairs, highlighting again tissue-specificity of enhancers and enhancer-promoter interactions.

Recent reports suggested that genes regulated by multiple enhancers were higher expressed than those regulated by a single enhancer, proposing that enhancers might contribute additively to the expression of their target genes [17, 49]. In support of this, we observed a steady increase of gene expression concomitant with increasing numbers of associated enhancers, with the genes not associated to any enhancers showing the lowest overall expression. A study of 12 mouse tissues has reported the enrichment for tissue-specific functions in genes associated to enhancers that transcribe eRNAs as compared to genes associated to non-transcribed enhancers [68]. Jin et al. recently showed that genes that did not interact with distal enhancers were enriched for housekeeping genes and also suggested that cell-specific genes were extensively controlled by *cis*-regulators [69]. We showed in this study that genes associated to many enhancers were more enriched for macrophage-related functions as compared to genes associated to only few or no enhancers. This finding might reflect a more fundamental principle of genome organisation and evolution, such as the importance of multiple enhancers for fine-tuned and redundant control of cell specialization and cell-specific responses.

We focused on enhancers and genes that responded to the polarization stimuli with increased expression in order to study transcriptional changes during macrophage activation. Notably, many stimuli-responsive genes were associated to stimuli-responsive enhancers, highlighting the importance of enhancer regulation in macrophage polarization. As expected for such a cell-type-specific process as macrophage polarization, most enhancers showed macrophage-specific eRNA expression, and genes were enriched for macrophage-specific functions. As an important example, we assessed 20 M1 and 26 M2 marker genes to demonstrate enhancer regulation for known macrophage polarization markers [1-3, 6, 51]. Three M1 markers (Ccl20, Fpr2, Ido1) and four M2 markers (Chi3l3, Chi3l4, Alox12e, Chia) were not expressed in our data. For a total of 16 M1 or M2 macrophage marker genes, we identified associated enhancers. Moreover, for 11 of these marker genes, we found enhancers that might regulate these genes specifically in M1 or M2 stimulation (Table 1). Hence, these enhancers present new potential markers for a particular macrophage activation status. Seven additional marker

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genes, identified as stimuli-responsive, were not associated to any stimuli-responsive enhancer (Gpr18, Il12b, Il6, Inhba in the G1' set; Il27ra, Klf4, Myc in the G2' set). The remaining marker genes were not deemed stimuli-responsive. Of those, M1 markers Il1b, Cd86, Marco, and Il23a, and M2 markers Mmp12, Tgm2, Clec4a2, Stab1, F13a1 were associated to at least one enhancer in macrophages. M1 marker Ccr7 and M2 markers Retnla, Ccl17, Ccl22, Chi3l1, Cxcl13, and Ccl12, were not associated to any enhancers in macrophages.

Moreover, we observed a particular genomic distribution of potential marker enhancers associated to Egr2 and Igf1 M2 marker genes, which suggested that these regulatory DNA regions might represent stretch enhancers. Parker et al., in a recent study, investigated stretch enhancers in human cells and proposed that such extended regions could serve as molecular runways to attract tissue-specific TFs and focus their activity [54]. Similarly to Parker et al., potential stretch enhancer regions identified here were associated to cell-type specific genes and were demarcated by broad H3K27ac signals, specifically higher enriched in M2 as compared to M1 and untreated macrophages (Fig. 4e, Additional file 1: Figure S6c). Therefore, we propose that stretch enhancers might be involved in the regulation of macrophage polarization. However, further studies are required to investigate this phenomenon in more detail.

Our approach inferred M1- and M2-responsive enhancers that were strongly enriched for TFBS of known inflammatory TFs. These results are in line with previous reports in mouse macrophages. For example, Spi1 (PU.1) has been extensively studied as a crucial TF involved in macrophage differentiation and transcriptional regulation [40]. Moreover, Spi1 was deemed a pioneering or lineage-determining TF in macrophages, which defines enhancer regions and occupies many enhancers in macrophages [27, 34, 40, 58]. Furthermore, Heinz et al. suggested that collaborative action of Spi1 with Cebpb was required for the deposition of enhancer-associated chromatin marks [58]. Ghisletti et al. reported enrichment for NF-kB (Rel) and Irf TFs in enhancers induced by LPS in mouse macrophages [40]. Likewise, transcribed enhancers induced by LPS and IFNγ stimulation showed enrichment for NF-kB/Rel, Irf, and Stat1 binding motifs [34]. In addition, we previously showed that TFs including Rela and Irf1 drive expression of protein-coding and lncRNA genes during macrophage polarization [9]. Taken together, our results link enhancer activation to the transcriptional program induced by M1 and M2 stimuli.

Conclusions

In this study, we have established a genome-wide catalogue of enhancers and enhancer-promoter regulatory interactions in mouse macrophages. We identified 8,667 active enhancers forming 64,891 regulatory associations with protein-coding gene promoters and highlight tissue- and stimuli-specificity. We observed increased gene expression concomitant with increasing numbers of associated enhancers, which is in support of a model of additive enhancer action. In addition, we noticed a shift towards stronger enrichment for macrophage-specific biological processes in genes associated to many enhancers. We found that cytokine stimulation in macrophages had a striking influence on enhancer activation and we used our enhancer-gene interactome to infer potential stimuli-specific marker enhancers. Our study represents the most comprehensive analysis of transcribed enhancer activities in mouse macrophages to date and our results extend current

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

knowledge of transcriptional regulation and highlight the importance of enhancers in macrophages in general and during polarization in particular.

Methods

CAGE data and processing

Mouse genomic coordinates (mm10) and tag counts of CAGE transcription start sites (TSSs) were obtained from the FANTOM5 project [16] data repository [70]. Data for 969 mouse samples classified as "primary_cell", "timecourse", "tissue", and "cell_line" were used. The set included 184 macrophage samples profiled by us as described elsewhere [9], which we used here to construct a macrophage enhancer-promoter interactome (Additional file 11 for the list of macrophage samples).

The DPI program [71] was used as described in Forrest et al. [16] to cluster CAGE TSSs into CAGE peaks. Briefly, the algorithm uses independent component analysis to decompose regions with continuous CAGE signals into separate peaks based on their profile across different samples and tissues. With the default parameters, similarly to Forrest et al. [16], we obtained a list of all CAGE peaks and a subset of CAGE peaks enriched for promoter-associated signals. The latter file represents a subset of peaks meeting the FANTOM5 'robust' criteria, with a single TSS supported by 11 or more observations and one or more tag per million (TPM) in at least one experiment [16]. These two peak sets were used for identification of enhancers and annotation of protein-coding gene promoters, respectively. Tag counts of all TSSs composing a CAGE peak were summed up to derive a total tag count for that CAGE peak.

Annotation of protein-coding gene promoters

The set of 'robust' CAGE peaks derived by DPI (see above) was used to annotate promoters of protein-coding genes. Ensembl gene models version 75 [72] downloaded from the UCSC Table Browser [73] on 11 Aug 2016 was used to obtain coordinates of protein-coding transcripts and genes. We assigned a CAGE peak to an Ensembl protein-coding transcript if its 5' end was mapped within 500bp of the 5' end of the transcript on the same strand. The transcript annotation was extended to gene annotation by combining the CAGE peaks associated to all of the gene's transcripts.

Calculation of gene and promoter expression

TMM-normalization [74] of promoter tag counts was performed to derive normalized expression values in a form of tags per million (TPM). We excluded lowly expressed promoters from the analysis and retained only those with expression of at least one TPM in 10% of the macrophage samples. Expression of each gene was derived as a sum of expression of the gene's promoters. The resulting set included 24,449 promoters of 10,767 protein-coding genes.

Identification of mouse enhancers with CAGE data

The full set of 3,188,801 DPI-derived CAGE peaks was used for identification of mouse enhancers. CAGE peaks located within 500bp of protein-coding transcript start sites or within 200bp of exons were excluded (based on the Ensembl gene models version 75 [72]). This filtering resulted in 1,890,465 CAGE peaks. Next, we used a strategy similar to Andersson et al. to infer enhancer regions as clusters of closely located bi-directional divergent CAGE peaks and to derive the corresponding tag counts [37]. The resulting 42,470 regions were

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

deemed mouse enhancer regions. The counts were normalized to tags per million (TPM) using TMM-normalization procedure [74]. Enhancers with non-zero expression in at least 10% of our macrophage samples were deemed transcribed in our macrophage samples.

Selection of enhancers regulating protein-coding genes in macrophages

Enhancer-specific chromatin signatures were based on ChIP-seq profiling of H3K4me1 and H3K27ac histone marks and were obtained from a study by Ostuni et al. [47]. Transcribed enhancers with at least 1bp overlap with the regions inferred by Ostuni et al. [47] were retained. Genomic coordinates of TADs in mouse embryonic stem cells were obtained from a study by Dixon et al. [46]. We selected pairs of enhancers and promoters where both features were entirely located within the same TAD. For each of these pairs, we calculated Spearman's correlation coefficient between expression levels of enhancer eRNA and promoter across our macrophage samples and selected only pairs with positive correlation coefficient and FDR < 10^{-4} (Benjamini-Hochberg procedure [75]). We considered an enhancer to regulate a gene if it was associated to at least one of the gene's promoters. All mm9 genomic coordinates were converted to mm10 using the liftOver program [76].

Gene set enrichment analysis

KEGG pathway maps [77] or GO biological process ontology [78] were used as sets of biological terms for GSEA. GO terms and associated genes were retrieved using the R package GO.db [79]. We used hypergeometric distribution to calculate the probability of obtaining the same or larger overlap between a gene set of interest and each biological term [80]. Derived p-values were corrected for multiple testing using Benjamini-Hochberg procedure [75]. As a background, a set of 22,543 Ensembl protein-coding genes (version 75) was used [72].

Identification of macrophage-specific features

Normalized TPM expression data were used to calculate a z-score for each of our 184 macrophage samples for each enhancer and gene by subtracting the mean and dividing by the standard deviation of expression values of the same feature in 744 FANTOM5 non-macrophage mouse samples (Additional file 12), similarly to Yao et al. [50]. Enhancers and genes with z-score > 3 (i.e. expressed more than 3 standard deviations above the mean of the non-macrophage samples) in at least 10% of macrophage samples were deemed macrophage-specific.

TFBS over-representation analysis

TFBS data for mouse were obtained from ENCODE [57] and HT-ChIP [56]. Raw sequencing data were mapped to the mm10 genome build for each tissue and cell type separately and peaks were called using MACS2 [81]. TFBS summits with FDR < 10^{-4} were retained. We used three different background sets: the whole set of identified mouse enhancers, the subset of enhancers not expressed in macrophages, and a set of random genomic regions located within TADs excluding gaps, repeated sequences, Ensembl coding regions, and mouse enhancers identified here. Significantly over-represented TFBS were selected based on empirical p-value < 0.01 from a Monte Carlo analysis of 1,000 trials [82]. We retained only TFBS which showed p-value < 0.01 using all three background sets and non-zero expression of the corresponding TF in our macrophages samples.

Identification of stimuli-responsive features

We calculated a z-score for each of 16 M1 and 16 M2 macrophage samples for each enhancer and gene by subtracting the mean and dividing by the standard deviation of expression values of the same feature in ten non-

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

stimulated macrophage samples, similarly to the approach for identification of macrophage-specific features. Genes and enhancers with z-score > 3 in more than 25% of the corresponding samples were deemed stimuli-responsive. Of associations between stimuli-responsive enhancers and genes, we sub-selected those with a positive Spearman's correlation of expression in the corresponding activation state.

Identification of stimuli-specific enhancers

To identify M1-specific enhancers, we first selected enhancers which were deemed M1-responsive, but not M2-responsive. Second, a z-score for each of 16 M1 samples was calculated using 16 M2 samples as a background. Enhancers with z-score > 3 in more than 25% of M1 samples were deemed M1-specific. Similar strategy was used to infer M2-specific enhancers.

All analyses made extensive use of the BEDTools utilities [83] and the R software [84].

Abbreviations

CAGE: cap analysis of gene expression; ChIP-seq: chromatin immunoprecipitation followed by sequencing; E-P pairs: enhancer-promoter pairs; FDR: false discovery rate; GSEA: gene set enrichment analysis; TADs: topologically associating domains; TFBS: transcription factor binding sites; TF: transcription factor; TPM: tags per million.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The dataset analysed in the study is available in the FANTOM5 repository, http://fantom.gsc.riken.jp/5/datafiles/reprocessed/mm10_v2/basic/. The datasets supporting the conclusions of this article are included within the article and its additional files.

Additional files

Additional file 1: Supplementary figures. Supplementary document (.pdf) containing all Supplementary figures. (PDF 3487 kb)

Additional file 2: Established regulatory associations between transcribed enhancers, promoters and protein-coding genes. (XLSX 3505 kb)

Additional file 3: Gene set enrichment analysis of 1,306 genes associated to three or four enhancers. (XLSX 11 kb)

Additional file 4: Gene set enrichment analysis of 4,149 genes not associated to any enhancer. (XLSX 10 kb)

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Additional file 5: Gene set enrichment analysis of top 500 genes with the highest expression in macrophages among 1,481 genes associated to macrophage-specific enhancers. (XLSX 11 kb)

Additional file 6: Gene set enrichment analysis of top 500 genes with the highest expression in macrophages among 1,207 genes associated to non-macrophage-specific enhancers. (XLSX 9 kb)

Additional file 7: Over-representation analysis of binding sites in macrophage-specific and non-macrophage-specific enhancers. (XLSX 13 kb)

Additional file 8: Regulatory associations between 115 M1-responsive enhancers and 105 M1-responsive genes. (XLSX 22 kb)

Additional file 9: Regulatory associations between 131 M2-responsive enhancers and 98 M2-responsive genes. (XLSX 19 kb)

Additional file 10: Over-representation analysis of binding sites in E1 and E2 enhancers. (XLSX 12 kb)

Additional file 11: 184 macrophage samples (including biological replicates) used in this study. (XLSX 16 kb)

Additional file 12: 744 FANTOM5 non-macrophage mouse samples that were used as a background set for the calculation of macrophage-specific expression. (XLSX 27 kb)

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the South African National Research Foundation (NRF) and from the Department of Science and Technology, South African Research Chair Initiative (SARCHi) and South Africa Medical Research Council (SAMRC) to FB, grant from the Japan Society for the Promotion of Science (JSPS) and National Research Foundation of South Africa to FB and HS, and grants from the South African National Research Foundation (NRF) Competitive Programme for Unrated Researchers (CSUR) to RG.

Authors' contributions

ED performed computational analyses. SS designed the study. SS and ED analysed data, interpreted results, and wrote the manuscript with input from all authors. RG, MM, HS, and FB helped interpret results and provided data. All authors read and approved the final manuscript.

References

1. Murray PJ, Wynn TA. Protective and pathogenic functions of macrophage subsets. Nat Rev Immunol. 2011;11:723-737.
2. Mosser DM, Edwards JP. Exploring the full spectrum of macrophage activation. Nat Rev Immunol. 2008;8:958-969.
3. Martinez FO, Gordon S. The M1 and M2 paradigm of macrophage activation: time for reassessment. F1000Prime Rep. 2014;6:13.
4. Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdt S, et al. Macrophage activation and polarization: nomenclature and experimental guidelines. Immunity. 2014;41:14-20.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5. Sica A, Mantovani A. Macrophage plasticity and polarization: in vivo veritas. J Clin Invest. 2012;122:787-795.
6. Gordon S, Martinez FO. Alternative activation of macrophages: mechanism and functions. Immunity. 2010;32:593-604.
7. Herbert DR, Holscher C, Mohrs M, Arendse B, Schwegmann A, Radwanska M, et al. Alternative macrophage activation is essential for survival during schistosomiasis and downmodulates T helper 1 responses and immunopathology. Immunity. 2004;20:623-635.
8. Jenkins SJ, Ruckerl D, Thomas GD, Hewitson JP, Duncan S, Brombacher F, et al. IL-4 directly signals tissue-resident macrophages to proliferate beyond homeostatic levels controlled by CSF-1. J Exp Med. 2013;210:2477-2491.
9. Roy S, Schmeier S, Arner E, Alam T, Parihar SP, Ozturk M, et al. Redefining the transcriptional regulatory dynamics of classically and alternatively activated macrophages by deepCAGE transcriptomics. Nucleic Acids Res. 2015;43:6969-6982.
10. Lugo-Villarino G, Verollet C, Maridonneau-Parini I, Neyrolles O. Macrophage polarization: convergence point targeted by mycobacterium tuberculosis and HIV. Front Immunol. 2011;2:43.
11. Guler R, Parihar SP, Savvi S, Logan E, Schwegmann A, Roy S, et al. IL-4Ralpha-dependent alternative activation of macrophages is not decisive for Mycobacterium tuberculosis pathology and bacterial burden in mice. PLoS One. 2015;10:e0121070.
12. Odegaard JI, Ricardo-Gonzalez RR, Goforth MH, Morel CR, Subramanian V, Mukundan L, et al. Macrophage-specific PPARgamma controls alternative activation and improves insulin resistance. Nature. 2007;447:1116-1120.
13. Komili S, Silver PA. Coupling and coordination in gene expression processes: a systems biology view. Nat Rev Genet. 2008;9:38-48.
14. Weake VM, Workman JL. Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet. 2010;11:426-437.
15. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet. 2012;13:233-245.
16. Forrest ARR, Kawaji H, Rehli M, Baillie JK, de Hoon MJL, Haberle V, et al. A promoter-level mammalian expression atlas. Nature. 2014;507:462-70.
17. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15:272-286.
18. Gorkin DU, Leung D, Ren B. The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell. 2014;14:762-775.
19. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107:21931-21936.
20. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39:311-318.
21. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13:613-626.
22. Romanoski CE, Link VM, Heinz S, Glass CK. Exploiting genomics and natural genetic variation to decode macrophage enhancers. Trends Immunol. 2015;36:507-518.
23. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009;459:108-112.
24. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013;49:825-837.
25. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009;457:854-858.
26. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129:823-837.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

27. Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. Cell. 2014;159:1327-1340.
28. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell. 2014;159:1312-1326.
29. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295:1306-1311.
30. van Steensel B, Dekker J. Genomics tools for unraveling chromosome architecture. Nat Biotechnol. 2010;28:1089-1095.
31. de Laat W, Klous P, Kooren J, Noordermeer D, Palstra RJ, Simonis M, et al. Three-dimensional organization of gene expression in erythroid cells. Curr Top Dev Biol. 2008;82:117-139.
32. Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre BM, et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. 2015;25:582-597.
33. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465:182-187.
34. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. PLoS Biol. 2010;8:e1000384.
35. Denisenko E, Ho D, Tamgue O, Ozturk M, Suzuki H, Brombacher F, et al. IRNdb: the database of immunologically relevant non-coding RNAs. Database. 2016;2016:baw138.
36. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. Nature. 2017;543:199-204.
37. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507:455-461.
38. Kanamori-Katayama M, Itoh M, Kawaji H, Lassmann T, Katayama S, Kojima M, et al. Unamplified cap analysis of gene expression on a single-molecule sequencer. Genome Res. 2011;21:1150-1159.
39. Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell. 2013;51:310-325.
40. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, et al. Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. Immunity. 2010;32:317-328.
41. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012;488:116-120.
42. Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell. 2013;155:1507-1520.
43. Thomas GD, Hanna RN, Vasudevan NT, Hamers AA, Romanoski CE, McArdle S, et al. Deleting an Nr4a1 super-enhancer subdomain ablates Ly6Clow monocytes while preserving macrophage gene function. Immunity. 2016;45:975-987.
44. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science. 2015;347:1010-1014.
45. Mukhopadhyay S, Ramadass AS, Akoulitchev A, Gordon S. Formation of distinct chromatin conformation signatures epigenetically regulate macrophage activation. Int Immunopharmacol. 2014;18:7-11.
46. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485:376-380.
47. Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, et al. Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell. 2013;152:157-171.
48. Rocha PP, Raviram R, Bonneau R, Skok JA. Breaking TADs: insights into hierarchical genome organization. Epigenomics. 2015;7:523-526.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

49. Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell Res. 2012;22:490-503.
50. Yao P, Lin P, Gokoolparsadh A, Assareh A, Thang MW, Voineagu I. Coexpression networks identify brain region-specific enhancer RNAs in the human brain. Nat Neurosci. 2015;18:1168-1174.
51. Jablonski KA, Amici SA, Webb LM, Ruiz-Rosado JdD, Popovich PG, Partida-Sanchez S, et al. Novel markers to delineate murine M1 and M2 macrophages. PLoS One. 2015;10:e0145342.
52. Laslo P, Spooner CJ, Warmflash A, Lancki DW, Lee HJ, Sciammas R, et al. Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell. 2006;126:755-766.
53. Hnisz D, Abraham Brian J, Lee Tong I, Lau A, Saint-André V, Sigova Alla A, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155:934-947.
54. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A. 2013;110:17921-17926.
55. Spadaro O, Camell CD, Bosurgi L, Nguyen KY, Youm YH, Rothlin CV, et al. IGF1 Shapes Macrophage Activation in Response to Immunometabolic Challenge. Cell Rep. 2017;19:225-234.
56. Garber M, Yosef N, Goren A, Raychowdhury R, Thielke A, Guttman M, et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol Cell. 2012;47:810-822.
57. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014;515:355-364.
58. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576-589.
59. Ruffell D, Mourkioti F, Gambardella A, Kirstetter P, Lopez RG, Rosenthal N, et al. A CREB-C/EBPbeta cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. Proc Natl Acad Sci U S A. 2009;106:17475-17480.
60. Saliba DG, Heger A, Eames HL, Oikonomopoulos S, Teixeira A, Blazek K, et al. IRF5:RelA interaction targets inflammatory genes in macrophages. Cell Rep. 2014;8:1308-1317.
61. Fontana MF, Baccarella A, Pancholi N, Pufall MA, Herbert DR, Kim CC. JUNB is a key transcriptional modulator of macrophage activation. J Immunol. 2015;194:177-186.
62. Tugal D, Liao X, Jain MK. Transcriptional control of macrophage polarization. Arterioscler Thromb Vasc Biol. 2013;33:1135-1144.
63. Taniguchi T, Ogasawara K, Takaoka A, Tanaka N. IRF family of transcription factors as regulators of host defense. Annu Rev Immunol. 2001;19:623-655.
64. Roy S, Guler R, Parihar SP, Schmeier S, Kaczkowski B, Nishimura H, et al. Batf2/Irf1 induces inflammatory responses in classically activated macrophages, lipopolysaccharides, and mycobacterial infection. J Immunol. 2015;194:6035-6044.
65. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012;148:84-98.
66. Lupianez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. Trends Genet. 2016;32:225-237.
67. Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 2014;24:390-400.
68. Cheng JH, Pan DZ, Tsai ZT, Tsai HK. Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues. Sci Rep. 2015;5:12648.
69. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013;503:290-294.

bioRxiv preprint doi: https://doi.org/10.1101/163519; this version posted July 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

70. FANTOM consortium: FANTOM5 Data. http://fantom.gsc.riken.jp/5/datafiles/reprocessed/mm10_v2/basic/ (2016). Accessed 07 Sept 2016.
71. Kawaji H. Decomposition-based peak identification (DPI): source code. https://github.com/hkawaji/dpi1/ (2016). Accessed 04 Sept 2016.
72. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, et al. Ensembl 2011. Nucleic Acids Res. 2011;39:D800-806.
73. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004;32:D493-496.
74. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139-140.
75. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57:289-300.
76. UCSC Genome Browser Utilities: Batch Coordinate Conversion (liftOver). https://genome.ucsc.edu/cgi-bin/hgLiftOver. Accessed 15 Sept 2016.
77. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27-30.
78. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25-29.
79. Carlson M. GO.db: A set of annotation maps describing the entire Gene Ontology. 2015.
80. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1-13.
81. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.
82. North BV, Curtis D, Sham PC. A note on the calculation of empirical P values from Monte Carlo procedures. Am J Hum Genet. 2002;71:439-441.
83. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841-842.
84. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015. http://www.R-project.org/.