Nucleic Acids Research, 2026, **54**, gkag824
https://doi.org/10.1093/nar/gkag824
Gene regulation, Chromatin and Epigenetics

OXFORD

CpG content contributes to tissue-resident macrophage enhancer selection

Dóra Bojcsuk[1], Zsófia Varga[2], Zsuzsanna Kolostyák[3], Petros Tzerpos[1], Noemí Caballero-Sánchez[1,4], Tímea Cseh[1], Andreas Patsalos[5], Zsolt Czimmerer[2], László Nagy[1,5,*], Gergely Nagy[1,*]

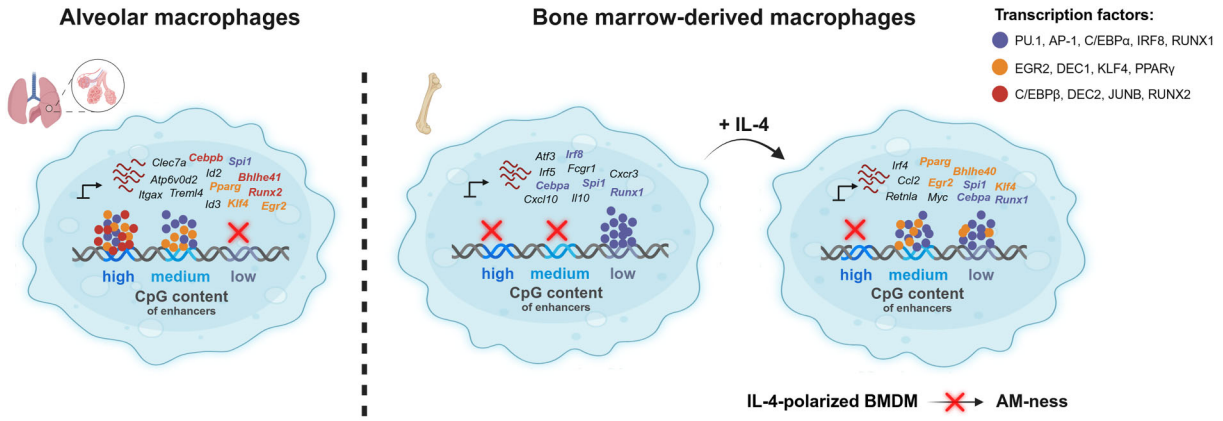
[1]Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary
[2]HUN-REN Biological Research Centre, Institute of Genetics, Macrophage Polarization Group, Szeged, Hungary
[3]Department of Surgery, Transplantation and Gastroenterology, Semmelweis University, Budapest, Hungary
[4]Doctoral School of Molecular Cell and Immune Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary
[5]Departments of Medicine and Biomedical Engineering, Johns Hopkins University School of Medicine, Institute for Fundamental Biomedical Research, Johns Hopkins All Children's Hospital, St. Petersburg, FL, United States

*To whom correspondence should be addressed. nagygergely@med.unideb.hu
Correspondence may also be addressed to László Nagy. Email: lnagy@jhmi.edu

Abstract

Tissue-resident macrophages acquire specialized identities through transcription factor networks operating in distinct chromatin environments. Early growth response 2 (EGR2) contributes both to alveolar macrophage identity and to interleukin (IL)-4-driven polarization of bone marrow-derived macrophages, yet the regulatory principles enabling its function across these contexts remain unclear. By integrating RNA-seq and ATAC-seq data from wild-type and *Egr2*-deficient macrophages, we show that macrophage identity is constrained by DNA methylation-dependent *cis*-regulatory landscapes. Alveolar macrophage-specific enhancers are predominantly CpG rich and hypomethylated, a feature shared with additional tissue-resident macrophage populations, whereas bone marrow-derived macrophage-accessible regulatory regions are largely CpG poor. Although IL-4 induces EGR2 together with KLF4 and DEC1 in bone marrow-derived macrophages, this transcription factor module fails to engage CpG-rich enhancers, indicating that the bone marrow-derived macrophage regulatory network lacks additional alveolar macrophage-specific components required for their activation. These findings identify CpG content as a determinant of enhancer competence and provide a mechanistic explanation for the context-dependent activity of EGR2 in macrophages, placing macrophage plasticity under the control of sequence-encoded epigenetic constraints.

Graphical abstract



Introduction

Mouse bone marrow-derived macrophages (BMDMs) are a widely used *in vitro* cellular model to examine macrophage plasticity upon different stimuli [1–6]. It can be considered an artificial model as it is differentiated *in vitro* in the presence of macrophage colony-stimulating factor (M-CSF), but several immune-related processes, such as endocytic, antimicrobial, inflammatory, or anti-inflammatory, have been shown,

Received: February 10, 2026. **Revised:** July 28, 2026. **Accepted:** July 30, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

confirmed, or identified in these naïve macrophages [7, 8]. It is unknown whether BMDM-like cells exist *in vivo* as naïve or ground-state macrophages, but—based on their origin—monocyte-derived naïve macrophages can be the most closely related types since monocytes develop from the myeloid progenitor cells of the bone marrow. In the body, several types of tissue-resident macrophages (TRMs) are replaced by bone marrow-derived cells upon monocyte infiltration and colonization, when monocytes differentiate into specific macrophages during accommodating to their new tissue environment [9]. Independent of the given tissue, damage or infection attracts circulating monocytes from the bloodstream, which contribute first to inflammation, then to tissue repair; this happens, for example, during skeletal muscle regeneration or upon lung infections. In the former case, the phenotypic transition of macrophages takes a few days, while in the latter case, monocytes first become protective macrophages, then—after 2 months—lose this profile and turn into regular alveolar macrophages (AMs) [10, 11]. This slow transition is probably due to the presence of granulocyte–macrophage colony-stimulating factor (GM-CSF), instead of M-CSF, produced by the alveolar epithelial cells and required for the differentiation and maintenance of AMs [12].

Given the distinct growth factor requirements and environmental conditions of AMs and BMDMs, it remains unclear to what extent macrophage polarization signals can induce overlapping transcriptional programs in these cell types. The alternative (non-inflammatory or tissue-repairing) polarization of BMDMs induced by interleukin 4 (IL-4) activates a transcription factor (TF) cascade that includes respectively early growth response 2 (EGR2), Krüppel-like factor 4 (KLF4), differentiated embryo-chondrocyte expressed gene 1 (DEC1; *Bhlhe40*), and peroxisome proliferator-activated receptor gamma (PPARγ), which are already highly expressed in AMs [13–19, 20]. This could suggest that BMDMs can be brought closer to AMs by IL-4 treatment. However, given that IL-4 does not induce *Egr2* in AMs, it remains highly questionable whether alternative polarization or the activity of these TFs—in the presence of M-CSF and in the absence of GM-CSF—are sufficient to obtain AM-like cells *in vitro*, and whether these factors can mimic the alveolar space microenvironment [20]. In any case, transdifferentiation can be achieved by introducing a few key TFs; for instance, the overexpression of CCAAT/enhancer-binding protein alpha (C/EBPα) and C/EBPβ is sufficient to generate macrophages from B cells [21, 22]. Besides the previously mentioned IL-4-inducible TFs, there are additional ones, such as KLF7, C/EBPβ, and DEC2 (*Bhlhe41*) that also appear to be important in AMs, but the identity of the principal regulator(s) is unknown [17, 20]. Among AM-specific TFs, C/EBP (basic leucine zipper; bZIP) and DEC (basic helix–loop–helix; bHLH) proteins must form homo- or heterodimers to bind DNA efficiently and regulate gene expression in a context-dependent manner, so the presence and relative abundance of dimerization partners can critically influence cellular function [23, 24].

However, TF availability alone does not determine which *cis*-regulatory elements can be bound and which *cis*-regulatory regions can be activated, as TF binding is also shaped by the local DNA sequence and the epigenetic state of the underlying chromatin. Both G/C content and the occurrence of CpG dinucleotides have been documented to contribute to this process [25]. G/C-rich sequences are frequently associated with accessible and transcriptionally active regions, whereas CpG dinucleotides provide sites for DNA methylation. When CpGs are located within or near TF-binding sites, their methylation state can alter TF recognition and thereby influence the accessibility and activity of the corresponding regulatory region [26, 27].

Our aim was to determine how shared and macrophage subtype-specific TFs interact with distinct chromatin environments to establish macrophage identity. To address this question, we performed RNA-seq and ATAC-seq experiments in *Egr2*$^{+/+}$ (wild-type, WT) and *Egr2*$^{fl/fl}$ (*Egr2* knockout, *Egr2* KO) mouse AMs and BMDMs in the presence or absence of IL-4. These models were selected because EGR2 contributes to the transcriptional regulatory networks of both cell types, although its role and the magnitude of the IL-4 response differ between AMs and BMDMs [1]. We compared differentially expressed genes (DEGs) and differentially open chromatin regions (DORs) across cell types, genotypes, and treatment conditions. TF motif enrichment within these regions was then integrated with TF expression data to identify candidate regulators of AM identity. To put our findings into a broader context, we incorporated publicly available RNA-seq, ATAC-seq, Bisulfite-seq, and ChIP-seq datasets from AMs, BMDMs, microglia, and other TRMs.

Our results show that AMs and BMDMs are regulated by different TF networks independent of the presence or absence of IL-4, although IL-4 modulates gene expression in the same direction in both cell types. EGR2 has distinct roles in these cells: it contributes to AM maintenance and a more moderate IL-4 response, whereas it is an indispensable factor for the alternative polarization of BMDMs. These findings are likely to be the result of an epigenetic barrier between BMDMs and TRMs: this is the CpG content of DORs, which is higher in TRMs relative to BMDMs. As the DNA methylation-mediated blockade at these CpGs cannot be removed by the BMDM-specific TF–co-regulator complexes even in the presence of EGR2, other TFs with pioneer functions, e.g. C/EBPβ in AMs, may be required for the activity of the CpG-enriched cell-type-specific gene regulatory regions. These findings provide a framework for interpreting macrophage plasticity in which the subtype-specific combination of TFs unlocks the appropriate epigenetically pre-determined regulatory sites.

Materials and methods

Isolation of alveolar macrophages and treatment conditions

AMs were isolated by bronchoalveolar lavage (BAL) from *Egr2*$^{+/+}$ and *Egr2*$^{fl/fl}$, 8- to 12-week-old female mice (n = 3 per condition). To collect BAL samples, animals were euthanized by isoflurane inhalation, and their tracheas were cannulated. Lavage was performed with two aliquots of 0.7 ml of ice-cold phosphate-buffered saline (PBS; pH 7.3). The BAL cells were centrifuged at 400 *g* for 10 min at 4°C, and the supernatants were removed. BAL cells were suspended in ACK buffer (4.012 g NH_4Cl, 0.5 g $KHCO_3$, 18.61 mg $Na_2EDTA{\cdot}2H_2O$, and 0.5 l of milliQ; pH 7.2–7.4) to lyse red blood cells for 2 min at room temperature, and then 1 ml of MACS buffer [2 mM EDTA and 0.5% bovine serum albumin (BSA) in 1× PBS, pH 7.2) was added. The alveolar macrophages were purified with CD45 magnetic bead selection according to the manufacturer's protocol (CD45 bead isolation kit; order no. 130-052-301; Miltenyi). We

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

cultured isolated cells in 24-well plates in 0.5 ml of RPMI medium containing 1% penicillin/streptomycin/amphotericin B (A5955; Sigma) and 10% foetal bovine serum (FBS; F7524, Sigma). After attachment, we added 20 ng ml^{-1} IL-4 (Peprotech; 214-14). For RNA-seq experiments, total RNA was isolated using Trizol reagent (Zymo Research; R2050-1-200) after 24 h of incubation.

RNA sequencing experiments

RNA-seq was completed as described earlier [19, 28]. Briefly, from total RNA, the Ultra II RNA Sample Preparation Kit (New England BioLabs) was used for library preparation following the manufacturer's protocol. The poly(A)-tailed RNA selection was performed by magnetic beads attached with poly(T) oligonucleotide. After purification, the mRNA samples were eluted and fragmented at 94°C. Random priming reverse transcription was applied for cDNA generation. After end repair, single adenine base addition, and barcode–adaptor ligation, the libraries were generated by adapter-specific PCRs. The sequencing run was performed on an Illumina NextSeq 500 instrument using single-end 75 cycle sequencing.

ATAC-seq experiments

ATAC-seq was performed as previously described [19]. Briefly, *Egr2*$^{+/+}$ and *Egr2*$^{fl/fl}$ AMs were isolated from 3-month-old male and female mice ($n = 2$ per condition), followed by CD45-based cell isolation as described in the corresponding section. After treatment, the cells were collected and counted, and an equal number of cells from each sample was used for ATAC-seq. The nuclei were isolated by ATAC-LB (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL). Cell nuclei were subjected for tagmentation using the Nextera DNA Library Preparation Kit (Illumina). DNA then was purified with the MinElute PCR Purification Kit (Qiagen) and amplified using the Kapa HiFi HotStart Kit (Kapa Biosystems). Libraries were purified again with the MinElute PCR Purification Kit (Qiagen), and fragment distribution was analysed with an Agilent Bioanalyzer. Sequencing was performed on the NextSeq 500 platform.

Cell isolation and differentiation of bone marrow-derived macrophages

Cell isolation and differentiation were performed as described earlier [4]. Briefly, isolated BMDMs were differentiated for 6 days in Dulbecco's modified Eagle's medium (DMEM; Sigma; D5671) containing L929 supernatant (high glucose DMEM with 1% penicillin/streptomycin/amphotericin B, 15% FBS, and 20% L929 supernatant). On day 5, differentiated BMDMs were treated with a final concentration of 20 ng ml^{-1} IL-4 (Peprotech; 214-14) for 24 hs.

Chromatin immunoprecipitation-seq experiment

Chromatin immunoprecipitation (ChIP) was performed with minor modifications of the previously described protocol [29]. Briefly, 1×10^7 BMDMs were double cross-linked with disuccinimidyl glutarate (DSG; 45 min) and formaldehyde (10 min). Nuclei were isolated using a lysis buffer containing 0.15 M NaCl, 5 mM EDTA pH 7.5, 0.05 M Tris–HCl pH 7.5, and 0.5% NP-40, supplemented with protease inhibitor (Roche) prior to use. Chromatin was sheared with sonication and immunoprecipitated overnight with the DEC1 antibody (NB100-1800, Novus Biologicals). Chromatin–antibody complexes were extensively washed and eluted after pull-down by paramagnetic beads. Eluted complexes were decross-linked overnight and purified by the MinElute PCR Purification Kit. Immunoprecipitated DNA was quantified by a Qubit fluorometer. Libraries were prepared from 10 ng of DNA with TruSeq ChIP Library Systems (Illumina) according to the manufacturer's instructions. Libraries were sequenced on NextSeq 2000.

RNA sequencing data analysis

Raw sequence reads were aligned to the mm10 mouse reference genome assembly with HISAT2 (v2.1.0) using default parameters [30]. Transcripts were assembled with StringTie (v1.3.4d) [31]. BAM files were indexed with SAMtools (v1.7), then coverage (bigwig) files were generated by the bamCoverage program of deepTools (v3.0.2) [32, 33]. Expression levels were determined in fragments per kilobase of transcripts per million mapped reads (FPKM). Considering all RNA-seq samples ($n = 24$), genes with expression levels >1 FPKM in at least two samples were considered expressed (12 334 in total). Then, FPKM values were normalized to the upper decile per sample and were included in comparative analyses, with cut-offs of $P \leq 0.05$ and fold change (FC) ≥ 1.5 applied. The DEGs were determined with one-way analysis of variance (ANOVA), supplemented with post-hoc Tukey HSD statistical test by using the stats, tidyverse, and sqldf packages in R (v4.3.3). A sample-to-sample (correlation) plot was generated in R using Pearson correlation coefficients calculated from decile-normalized expression values of all the 12 334 genes. For reference-based principal component analysis (PCA) projection, log-normalized FPKM values were analysed in R after excluding mitochondrial, ribosomal, predicted (Gm), and microRNA-associated transcripts. Sample-level technical variation related to total expression and expression distribution was reduced using linear modelling. PCA was then generated using public monocyte and tissue-resident macrophage samples, and in-house AM and BMDM samples were projected into the resulting reference space [34]. Standard PCAs were also performed in R using log-normalized FPKM values after filtering out non-expressed genes and genes with zero variance, followed by centring and scaling of the data using the prcomp function. Upstream regulators were predicted with the Ingenuity Pathway Analysis (IPA) software [35]. Regulators were ranked based on the statistical significance of enrichment [$-\log(P\text{-value})$], reflecting the overlap between the observed DEGs and known downstream targets of each regulator. The top 20 regulators in each group were visualized. Enriched biological pathways were predicted with Metascape (v3.5.20230501) with DEGs as input [36]. Results from thePathway and Process Enrichment Analysis module were used, and pathways were ranked by *P*-value. The most significantly enriched terms were selected and visualized as networks, and functionally related groups were coloured similarly.

ATAC-seq data analysis

Raw sequence reads of *Egr2*$^{+/+}$ and *Egr2*$^{fl/fl}$ AM and BMDM samples were aligned to the mouse mm10 reference genome assembly using BWA (v07.17) with default parameters, then BAM files were generated with SAMtools (v1.7) [32, 37]. Peaks were predicted with the callpeak function of MACS2, then artefacts were removed using the blacklisted genomic re-

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

gions of ENCODE [38, 39]. Genome coverage (bedgraph) files were generated by HOMER's makeUCSCfile [21]. Before differential analysis, peak widths were fixed to 100 bp centred on their summit positions as defined by MACS2. Considering the 100 bp wide peaks across all ATAC-seq samples (n = 16), an initial consensus peak set was generated from all predicted peaks (n = 136 934). Then, coverage [reads per kilobase per million mapped reads (RPKM)] values were calculated on the initial consensus set from all samples and were normalized with the upper decile values per sample. After decile normalization, peaks present in at least two samples with ≥ 0.5 RPKM coverage were considered as part of the final consensus peak set (n = 73 602). DORs were then identified using one-way ANOVA followed by a post-hoc Tukey HSD test, implemented with the stats, tidyverse, and sqldf packages in R.

The separation of the transcription start site (TSS)-proximal (promoter) and TSS-distal (enhancer) regions was performed by annotating the final consensus peak set using HOMER's annotatePeaks with a minor modification: instead of the default HOMER promoter definition (TSS −1000 bp/+100 bp), peaks located within 1000 bp of a gene TSS were considered TSS proximal. Then genomic distribution of the TSS-distal peaks [in introns, intergenic regions, exons, 3′ untranslated regions (UTRs), and TTSs] was also determined by HOMER's annotatePeaks, where intergenic regions include both intergenic and non-coding regions. Decile-normalized read distribution (tag density) heat maps of the different peak sets were visualized by HOMER's annotatePeaks in 20 bp windows within 2 kb frames relative to the peak summits.

The AM- and BMDM-specific DORs and common regions were subjected to a distance-dependent merging and intersection analysis. Peak positions were concatenated, sorted, and merged using sortBed and mergeBed (bedtools) with increasing maximum merging distances [d=sqrt(10)^i, where i = 6,7,…,12] [40]. This generated merged peak sets representing progressively more inclusive genomic regions (1 kb to 1 Mb). Overlaps with the original three peak sets were quantified for each merged set using intersectBed (bedtools) with the -c option to count the number of overlapping intervals per merged region. The output was reformatted to produce a summary table containing, for each merged interval, the number of overlaps with the original three peak sets. This approach allowed us to examine how the number and composition of merged regulatory regions change as the maximum allowed merging distance increases.

Data analysis of publicly available ATAC-seq samples

Publicly available ATAC-seq datasets were collected from the Gene Expression Omnibus (GEO) or Sequence Read Archive (SRA) databases, and the raw sequences were downloaded from the Nucleotide Archive (ENA) database of the EMBL-EBI. A detailed list of publicly available next-generation sequencing data used in the study can be found in Supplementary Table S19. Control and *Egr2* KO, unstimulated, and IL-4-treated ATAC-seq samples from BMDMs were derived from our previously published datasets (GSM456315-322) [4]. In the case of TRM (GSE63338; alveolar macrophage, microglia, Kupffer, spleen, peritoneal, small intestine, large intestine, and monocyte) ATAC-seq samples, replicates were merged, and samples were analysed similarly to as described in the previous section [34].

PCA plots were generated in R using decile-normalized RPKM values from publicly available and our unstimulated AM samples. For secondary analyses, to represent AMs, the samples generated for this study were used. The peak widths were fixed to 100 bp centred on their summit positions as defined by MACS2. The TSS-distal regions (introns, intergenic regions, exons, 3′ UTRs, and TTSs) were defined by HOMER's annotatePeaks with the same modification as in the case of other analyses: peaks located within 1000 bp of a gene TSS were considered promoter proximal and thus were excluded from further analyses.

Considering the filtered, 100 bp wide peaks of the TRM, monocyte, and our AM and BMDM samples, similarly to the AM and BMDM ATAC-seq samples, an initial consensus peak set was defined (n = 173 596). Then, coverage (RPKM) values were calculated on the initial consensus set from all samples and were normalized with the upper decile values per sample. After decile normalization, TSS-proximal (promoter) regions (TSS ± 1000 bp) were excluded, and peaks present in at least two samples with ≥ 0.5 RPKM coverage were considered as part of the final consensus peak set (n = 64 490). For the differential analysis, decile-normalized RPKM values were used. We defined cell-type-specific regions (peaks enriched (≥1.5-fold) in a given cell type compared with all others), shared BMDM–TRM regions (peaks enriched in BMDM together with a TRM type compared with the remaining samples), and shared AM–TRM regions (peaks enriched in AM together with a TRM type compared with the remaining samples). Finally, peaks were clustered accordingly.

Bisulfite-seq analysis

Processed data files of publicly available Bisulfite-seq experiments from AM (GSM3940198), BMDM (GSM2974655), and microglia (GSM7273559) samples were downloaded, and then, based on the genomic coordinates and scores, bed and bedgraph files were generated in the command line [41, 42, 43].

CpGs were mapped using a HOMER motif file generated with seq2profile.pl (HOMER) by HOMER's annotatePeaks (using the -mbed option) within 1 kb frames relative to the peak centres or TSSs. Then, CpGs and methyl-CpGs were visualized as density heat maps on the different peak sets by HOMER's annotatePeaks (using the -ghist and -p options) in 10 bp windows within 1 kb frames relative to the peak centres or TSSs. Based on the total methyl-CpG number of the samples, the following scales were set to the methyl-CpG plots during heat map visualization by TreeView: AM, 0.1; BMDM, 0.25; microglia, 0.25. CpG density and G/C content were visualized as histograms on the different peak sets by HOMER's annotatePeaks (using the -hist 10 and -p options) in 10 bp windows within 1.5 kb frames relative to the peak centres or TSSs.

Polymerase II (Ser5P) ChIP-seq analysis

Raw sequence reads of the publicly available microglia polymerase (Pol) II ChIP-seq sample (GSM6928507) were downloaded and analysed together with our previously published Pol II samples from BMDMs (GSM2845626-27) using the same pipeline as for ATAC-seq, with the HOMER peak-calling parameter set to '-style histone' [28, 44]. Before the ba-

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

sic analysis, fastq files of the samples generated from BMDMs were merged.

ChIP-seq analysis

Raw sequence reads of C/EBPα ChIP-seq data from unstimulated BMDMs (SRR4302495), C/EBPβ ChIP-seq data from unstimulated AMs (GSM6138980) and BMDMs (GSM2974658), DEC1 ChIP-seq data from unstimulated AMs (GSM3983822), and DEC1 (GSM9465547-48) and EGR2 ChIP-seq data from unstimulated and 24 h IL-4-treated BMDMs (GSM4564263, GSM4564267) were analysed using the same pipeline as for ATAC-seq and Pol II ChIP-seq [14, 17, 28, 45].

CpG content analysis

CpGs were mapped by HOMER's annotatePeaks (using the -mbed option) within 300 bp frames relative to the peak centres of different peak sets, then the number of CpG hits + 1 was visualized as violin plots. The select set of enriched motifs was mapped by HOMER's annotatePeaks (using the -mbed option) within 200 bp frames relative to the peak centres of AM- and BMDM-specific DORs. The mapped motifs were extended, and their sequence was obtained by the homerTools extract command. Nucleotide frequencies were transformed to Homer motif files and turned into motif logos by motif2Logo.pl (HOMER), and CpG frequencies were determined. The number of mapped motifs with or without CpGs and methyl-CpGs was determined by intersectBed (HOMER). The observed-to-expected (O/E) ratio of CpG enrichment was quantified within 300 bp windows centred on peak midpoints, and empirical cumulative distribution functions (ECDFs) were plotted for the regulatory regions. Statistical comparisons were performed using two-sided unpaired Wilcoxon rank-sum tests (Mann–Whitney) with false discovery rate (FDR) correction.

Motif enrichment analysis

De novo motif enrichment analysis was carried out by HOMER's findMotifsGenome.pl and was performed on the summit ± 100 bp regions of the peaks. The targeted motif lengths were 8, 10, and 14 bp. *P*-values were calculated by comparing the enrichment within the target regions and that of a random set of regions (background) generated by HOMER. For motif distribution plots, matrices of the enriched motifs were mapped following the order of regions on the read density heat maps by using the -mbed parameter of HOMER's annotatePeaks.pl in 20 bp windows within 2 kb frames relative to the peak centres.

Data visualization

Coverage (bigwig) files were visualized using the Integrative Genomics Viewer (IGV, v2.8.0) [46]. Scatter, violin, line, and dot plots, as well as pie and (stacked) bar charts were plotted in GraphPad Prism (v9.4.1). The area-proportional Chow–Ruskey plot was generated with the Intervene web tool [47], while area-proportional Venn diagrams were created using BioVenn [47, 48]. Row-normalized heat maps were plotted with the pheatmap package in R, and the read distribution heat maps and motif enrichment plots were visualized by Java TreeView (v1.1.6r4) [49]. The schematic model and the Graphical abstract were created with BioRender.

Results

Alveolar macrophages and bone marrow-derived macrophages have diverse transcriptomes

In this study, WT and *Egr2*-deficient BMDMs and AMs—differentiated *in vitro* in the presence of M-CSF or *in vivo* in a GM-CSF-containing environment, respectively—were subjected to transcriptomic and chromatin accessibility analyses under unstimulated or IL-4-treated conditions (Fig. 1A). First, focusing on the unstimulated, steady-state condition, we sought to place BMDMs and AMs into a broader macrophage context by projecting our samples onto a monocyte and TRM reference space comprising microglia and Kupffer cells as well as splenic, peritoneal, intestinal, and alveolar macrophages (Fig. 1B). As expected, AMs positioned in proximity to the public AM reference, whereas BMDMs occupied a distinct position within the macrophage landscape and localized closer to monocytes than to TRM populations. We next compared AM and BMDM transcriptomes in detail and found that a high proportion (>30%) of the expressed genes showed a significantly different mRNA level between the two cell types (Fig. 1C; Supplementary Tables S1 and S2). These include a large number of AM signature genes, most of which show a higher gene expression level in AMs than in BMDMs as expected (Fig. 1D; Supplementary Fig. S1A; Supplementary Table S3) [50]. Several AM-specific genes are related to lipid metabolism, while the BMDM-specific genes have functions in cell proliferation and immune responses (Supplementary Fig. S1B). These large differences between the two cell types may be due to the significant number of differentially expressed TFs (Fig. 1E; Supplementary Table S4), but most TFs—including the macrophage lineage-determining TFs [LDTFs; *Spi1* (encoding PU.1), *Cebpa*, *Irf8*, *Jund*, *Mef2a*, *Mef2d*, and *Runx1*]—exhibit similar gene expression levels (Fig. 1F) [21, 45, 51, 52]. In addition, most core macrophage-specific genes—even if the difference is statistically significant—also display similar gene expression levels (Supplementary Fig. S1C; Supplementary Table S5), indicating that despite the differences, both cell types share a similar core macrophage programme [50]. Among the predicted upstream regulators that may be responsible for the established gene expression patterns, EGR2, DEC1 (*Bhlhe40*), and SREBF1 show a higher gene expression level in AMs, while E2F1 and FOXM1 might be major TFs in the establishment of the differential BMDM mRNA levels (Supplementary Fig. S1D). Nevertheless, the mRNA abundance alone was not enough to discern the activity and effects of the resulting TFs governing the distinct gene expression patterns (e.g. the BMDM-specific IRF7 is a viral signal-dependent TF) [53].

DNA methylation-dependent *cis*-regulatory regions are enriched in alveolar macrophages

ATAC-seq is a next-generation sequencing (NGS) method that allows the detection of the epigenetic imprint of DNA–TF interactions in an unbiased way; therefore, we carried out ATAC-seq experiments for both AMs and BMDMs [54]. As expected, many thousands of regions showed significantly different chromatin openness between the two cell types—this difference is balanced in the TSS-distal regions (enhancers), but the AM-specific TSS-proximal regions (promoters) are over-represented (Fig. 2A; Supplementary Fig. S2A, B; Supplementary Tables S6 and S7). AM- and BMDM-specific DORs are non-randomly organized along the genome, form-

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

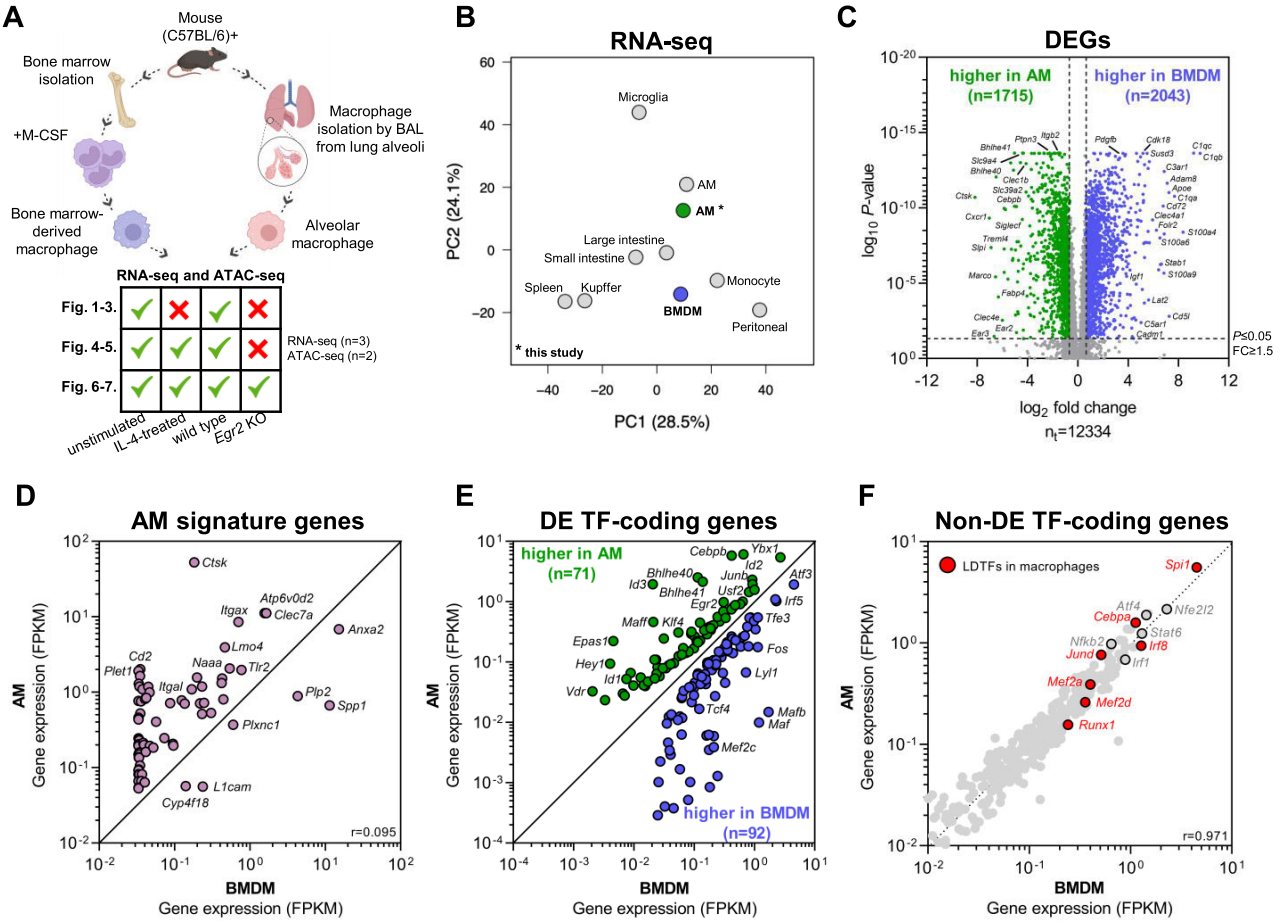


Figure 1. Alveolar and bone marrow-derived macrophages have distinct gene expression patterns. (**A**) Schematic overview of the experimental workflow. Biological replicates per condition were used for RNA-seq ($n = 3$) and ATAC-seq ($n = 2$). (**B**) PCA projection of in-house AM and BMDM samples onto a public monocyte and tissue macrophage reference space comprising microglia, spleen, Kupffer, peritoneal, intestinal, and alveolar macrophages. (**C**) The volcano plot represents DEGs between the unstimulated AMs and BMDMs. (**D–F**) Scatter plots depict the expression of the AM signature genes (D), the differentially expressed (E), and the non-differentially expressed (F) TF-coding genes in unstimulated AMs and BMDMs. Statistical significance was defined as $P \leq 0.05$ with an FC ≥ 1.5. (A) was partially created in BioRender. Barath, B.R. (2026) https://BioRender.com/g12f005.

ing spatially distinct clusters whose intermixing becomes apparent only beyond an ~30 kb scale (Supplementary Fig. S2C). In order to explain the observed ATAC-seq patterns and to identify the putative TF-binding sites of DORs, we applied *de novo* motif enrichment analyses (Supplementary Fig. S2D). In contrast to the BMDM-specific promoters, the AM-specific promoters show a strong GC-box motif enrichment, which can be responsible for the high chromatin openness in broader ranges. The differentially open TSS-distal regions are also well separated by the enriched motifs: within the AM-specific regions, the top differential motifs are those of EGR, CEBP, KLF, and the MITF/TFE (M)-box, the latter being a subtype of the bHLH-bound Ephrussi/enhancer (E)-boxes [55]. In contrast, within the BMDM-specific regions, they are the ETS:IRF composite element (EICE), the MAF response element (MARE), and the PU-box (Fig. 2A, B; Supplementary Fig. S2D) [56].

Notably, based on the HOMER motif collection, all the above-mentioned AM-specific motifs can contain a CpG dinucleotide (EGR, *TG*CG*TGGG*CG; CEBP, *TTG*CG*CAA*; KLF, *DGGG*CG*KGKY*; M-box, *TCA*CG*TGA*), which is the basis and target site of vertebrate DNA methylation, so we examined the CpG and G/C content of the DORs and common regions (Fig. 2C–E; Supplementary Fig. S2E, F) [21, 27, 57]. In addition, the extent of CpG methylation was examined using previously reported Bisulfite-seq data from AMs and BMDMs (Fig. 2C, D, right) [41, 43]. In contrast to the BMDM-specific promoters, almost all AM-specific promoters are CpG island promoters, which show a low degree of methylation in BMDMs and are largely demethylated and mostly active in AMs (Fig. 2C; Supplementary Fig. S2G). Similarly, the AM-specific TSS-distal regions are also enriched for CpGs relative to the BMDM-specific regions and, again, the extent of demethylation correlates with chromatin openness (Fig. 2C). In general, the open chromatin regions of the examined macrophages show a bimodal CpG distribution, where promoters are relatively CpG rich, and the TSS-distal regions are relatively CpG poor. In contrast, the BMDM-specific promoters themselves also display a bimodal distribution with a relatively low CpG content, and the CpG content of BMDM-specific TSS-distal regions is below the overall median (Fig. 2E). CpG content of the common regions is between the two extremes at both TSS-proximal and distal regions (Fig. 2D, F; Supplementary Fig. S2H).

Regarding the composition of putative PU-boxes and TPA response elements (TREs) and the mapped AM-specific TSS-distal motifs (EGR, CEBP, KLF, and M-box) together with

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

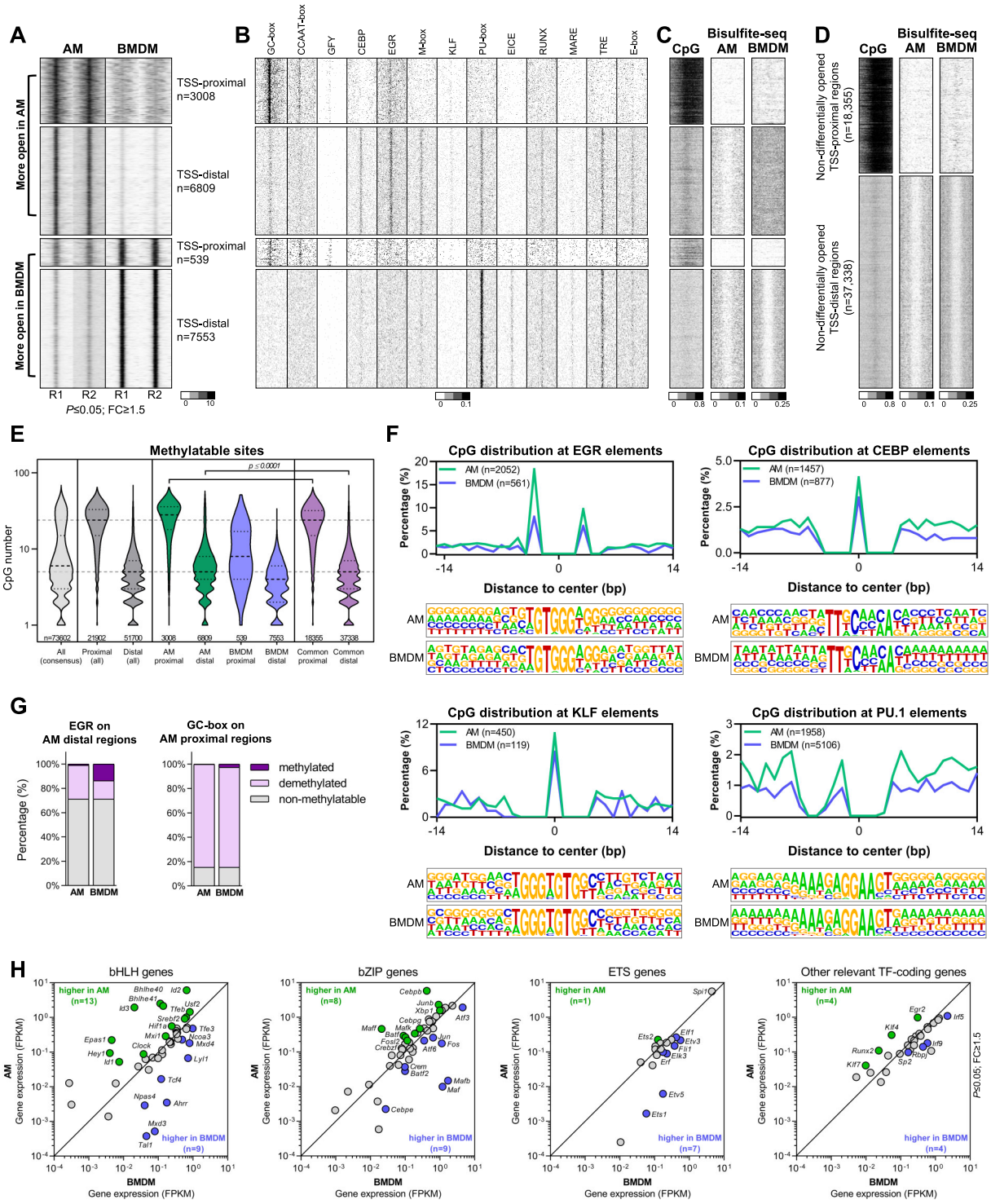


Figure 2. DNA methylation-dependent *cis*-regulatory regions are enriched in alveolar macrophages. (**A**) Read distribution heat map represents the chromatin openness of the differentially accessible TSS-proximal (promoter) and TSS-distal (enhancer) regions between the unstimulated AMs and BMDMs. Biological replicates were performed ($n = 2$); R: replicate. (**B**) Motif distribution heat map shows the presence of the *de novo* enriched motifs in 20 bp windows within 2 kb frames relative to the peak centres. (**C** and **D**) Heat maps show the CpG and methyl-CpG (Bisulfite-seq) density (C) across the differentially and (D) the non-differentially opened TSS-proximal and TSS-distal regions calculated from AMs and BMDMs. (**E**) Violin plots show the CpG number in the consensus peak set (all), TSS-proximal and TSS-distal regions, AM- or BMDM-specific, and non-differentially accessible (common) TSS-proximal and TSS-distal regions. The middle, dashed line represents the median, and the upper and lower dashed lines indicate the deciles. Applied statistics: two-sided unpaired Wilcoxon rank-sum test. (**F**) Line plots represent the CpG distribution calculated on the predicted EGR, CEBP, KLF, and PU.1 elements in AMs (green) and BMDMs (blue). The proportional motif logos represent the base frequency at each position. (**G**) Stacked bar charts show the proportions of methylated, demethylated, and non-methylatable putative EGR elements in AM-specific distal regions (left) and GC-boxes in AM-specific proximal regions (right) in AMs and BMDMs.(**H**) Scatter plots depict the gene expression levels of the bHLH, bZIP, ETS, and a set of other relevant TF-coding genes in unstimulated AMs and BMDMs. The DEGs are highlighted in colour. Statistical significance was defined as $P \leq 0.05$ with an FC $\geq$ 1.5.

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

their immediate vicinity, those within the AM-specific DORs have a higher CpG and G/C content relative to those of the BMDM-specific DORs (Fig. 2F; Supplementary Fig. S2I) [56]. Notably, putative EGR elements can be methylated at their edges, while the centre of putative CEBP and KLF elements and the 5′ region of putative PU-boxes can contain a CpG, although the latter three are largely CpG free. Overall, close to 30% of the AM-specific putative EGR elements are methylatable, out of which a small fraction is methylated in AMs, while close to half of them are methylated in BMDMs (Fig. 2G, left). As a control, we also checked the AM-specific promoters, where most GC-boxes contain a CpG, and the vast majority of these are demethylated in both cell types (Fig. 2G, right).

In the next step, we sought to associate the enriched motifs with their corresponding TFs using mRNA levels (Fig. 2H; Supplementary Table S8). Although there is a shift towards AMs, M/E-boxes are enriched in both macrophage types, and several putative bHLH proteins may bind to them. In BMDMs, *Tfe3*, *Mxd4*, and *Lyl1* are the most highly expressed bHLH genes, while in AMs, *Bhlhe40* and *Bhlhe41* are the top hits, as heterodimers containing ID2 or ID3 are not able to bind to DNA [58]. In the case of the bZIP TFs, the high level of C/EBPβ can be responsible for the AM-specific CEBP motif enrichment, whereas the BMDM-specific MAREs may be bound by large MAFs (*Maf*, *Mafb*). The activity of large MAFs is consistent with the differentiated state of BMDMs [59]. ETS genes typically show a higher expression in BMDMs than in AMs, which correlates with the PU-box enrichment in BMDMs. While the enrichment of RUNX motifs is similar in the AM- and BMDM-specific regions, according to the mRNA levels, in AMs, RUNX2 may also function as well as RUNX1. In addition, the AM-specific motif enrichments also correlate with the mRNA levels of *Egr2*, *Klf4*, and *Klf7*, and *Rbpj* and *Irf8* can be coupled with the GFY and EICE motif enrichments in BMDMs (Fig. 2H) [60, 61]. Notably, in the latter case, *Irf5* and *Irf9* show a significant difference, but EICE is only bound by IRF4 or IRF8, and, among their coding genes, *Irf8* shows a significantly higher expression.

In summary, the AM-specific DORs are CpG rich relative to the BMDM-specific DORs, and this also directly correlates with the putative TF-binding sites themselves. While AM-specific putative elements contain more and rather demethylated CpGs, BMDMs are enriched for the typically CpG-free PU-box, EICE, CEBP, TRE, MARE, RUNX, and M/E-box elements (Fig. 2B). These data suggest that CpG content and methylation state fundamentally shape macrophage-specific regulatory landscapes.

Tissue-resident macrophage identity is determined by DNA methylation-dependent enhancers

In order to examine whether the CpG content of TSS-distal *cis*-regulatory regions (enhancers) is a determining factor in various macrophages, we analysed publicly available ATAC-seq data from different TRMs and monocytes [34]. PCA confirmed that our samples are consistent with the public data, and BMDMs also fit in with the TRMs (Fig. 3A). Then, we generated clusters of those TSS-distal DORs that are specific for exactly one TRM, monocyte, or BMDM (Fig. 3B; Supplementary Table S9). Importantly, AMs differentiated *in vivo* in a GM-CSF-containing environment have the most specific regions relative to the other cell types differentiated in the presence of M-CSF [62]. The *de novo* motif enrichments of the clusters appeared specific as relevant motifs enriched included those of LDTFs and, for example, the AM-specific EGR motif, the canonical BACH1-bound MARE motif associated with splenic macrophages, or the Kupffer cell-specific LXRE in the respective cluster (Fig. 3C; Supplementary Fig. S3A) [34, 63]. We then assessed the CpG content of the clusters, and it turned out that each TRM- and the monocyte-specific cluster showed CpG enrichment relative to the BMDM-specific one (Fig. 3D left; Supplementary Fig. S3B). In addition, CpG demethylation and Pol II enrichment followed the cluster-specific patterns: the BMDM-specific Pol II signal correlated with demethylation throughout all clusters, while in microglia, it was largely restricted to the specific cluster; however, CpG demethylation was also detectable in multiple clusters beyond the specific one (Fig. 3D right) [44]. As a control, DORs specific for exactly two cell types were also selected for further testing (Fig. 3E; Supplementary Fig. S3C). In order to examine whether BMDMs utilize any CpG-rich regions, DORs open in BMDMs and one of the other cell types were chosen (Fig. 3E; Supplementary Table S10). These showed a generally low CpG content with no site-specific enrichment, a lower resolution but specific demethylation pattern, and specific Pol II signals. In contrast, DORs specific for AMs and one of the other cell types were also selected (Supplementary Fig. S3C; Supplementary Table S11). As expected, the CpG content was high in each cluster, and the demethylation and Pol II patterns showed cell type specificity. Interestingly, we observed that the three resulting microglia-specific clusters exhibited a correlation between CpG content and Pol II enrichment, suggesting a potential additional activating role associated with CpG-rich regions (Supplementary Fig. S3D). Taken together, unlike those of BMDMs, *cis*-regulatory regions associated with TRMs and monocytes consistently display varying levels of CpG enrichment.

Alternative polarization is a divergent process in AMs and BMDMs

While BMDMs are largely responsive to IL-4 mostly via EGR2, AMs show a restricted IL-4 response [1, 13, 28]. In order to examine whether, despite this, the alternative polarization of these macrophages shows similarities, we treated our cells with IL-4. First, DEGs were determined in the two cell types and, unsurprisingly, BMDMs have significantly more DEGs upon IL-4 treatment (Supplementary Fig. S4A, B; Supplementary Table S12). This is also clearly visible on the overall picture (PCA) and, importantly, the polarized AMs and BMDMs did not converge but rather diverged from each other (Fig. 4A). Overall, the gene expression changes upon IL-4 treatment were broadly similar in the two cell types and largely dependent on the initial levels, but the overlap between the DEGs of these macrophages was limited (Fig. 4B, C; Supplementary Fig. S4C).

Most of the genes affected by IL-4 were associated with immune functions, lipid metabolism, the cell cycle, or protein synthesis (Supplementary Fig. S4D). Out of these, the last process could be tested by our data. To explore how transcriptional changes affect the protein synthesis system, we used summed mRNA levels, specifically the amounts gained or lost upon IL-4 treatment. Overall, the number of induced and repressed genes was approximately equal in AMs. However, the summed mRNA levels showed that the induction

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

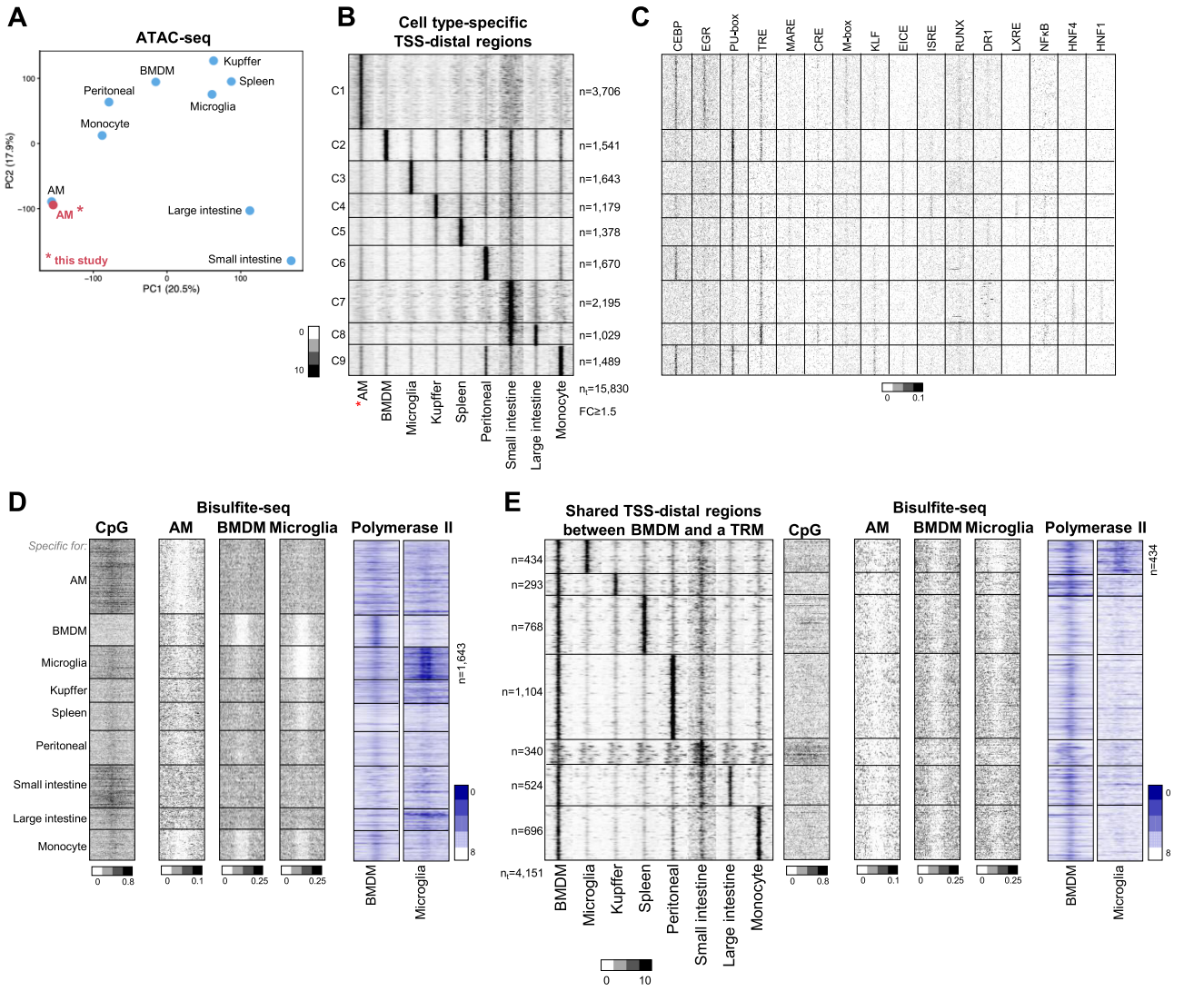


Figure 3. Tissue-resident macrophage identity is determined by DNA methylation-dependent enhancers. (**A**) PCA plot based on ATAC-seq coverage represents the relationships among the predicted chromatin accessibility profiles of BMDMs, monocytes, microglia, spleen, Kupffer, peritoneal, small intestine, large intestine, and alveolar macrophages (AMs). The AM sample marked with a red asterisk represents our own dataset, which was used for the subsequent analyses. (**B**) Clustered read distribution heat map (C1–C9) represents the cell type-specific TSS-distal regions ($n = 15\,830$ in total). (**C**) Motif distribution heat map shows the presence of the *de novo* enriched motifs in 20 bp windows within 2 kb frames relative to the peak centres. (**D** and **E**) Heat maps show the CpG and methyl-CpG density (calculated from AMs, BMDMs, and microglia), and the Pol II (pS5) activity (calculated from BMDMs and microglia) across (D) the C1–C9 clusters of the cell-type-specific TSS-distal regions and (E) the shared TSS-distal regions between BMDMs and a TRM ($n = 4151$ in total).

produced more than twice as much mRNA as was reduced by repression. In contrast, in BMDMs, the repressed genes were over-represented, and there was a minimal difference in the gained/lost aggregate mRNA levels (Fig. 4B, D). Although there are technical limitations of this approach, it appears that in AMs, IL-4 generates a significant mRNA surplus, which burdens the protein synthesis system, leading to elevated levels of endoplasmic reticulum (ER) stress genes (and presumably ER stress proteins).

Next, we investigated the overlap between the AM-specific genes and the IL4-induced genes in BMDMs, which was relatively small and mainly involved in cholesterol biosynthesis. In addition, five TF-coding genes were in this group: *Id2*, *Bhlhe40*, *Klf4*, *Srebf2*, and *Fosl2* (Fig. 4E). The other TF-coding genes differentially expressed upon IL-4 treatment showed various changes in the two cell types: most were repressed in BMDMs, and there was a set of 17 commonly induced TF-coding genes (Fig. 4F, G). Out of these, 12 showed a significant induction in both cell types, and there were nine TF-coding genes that were significantly induced only in BMDMs (Supplementary Fig. S4E). Despite the induction of these TF-coding genes, only a few AM signature genes showed a significant induction upon IL-4 treatment in BMDMs (Fig. 4H).

In summary, IL-4 affected gene expression in both AMs and BMDMs but elicited a substantially broader response in BMDMs. Despite partial overlap in the direction of transcriptional changes, the two cell types did not converge, indicating that alternative polarization is shaped by pre-existing cell identity and is layered onto, rather than replacing, the tissue-specific transcriptional programme of AMs.

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

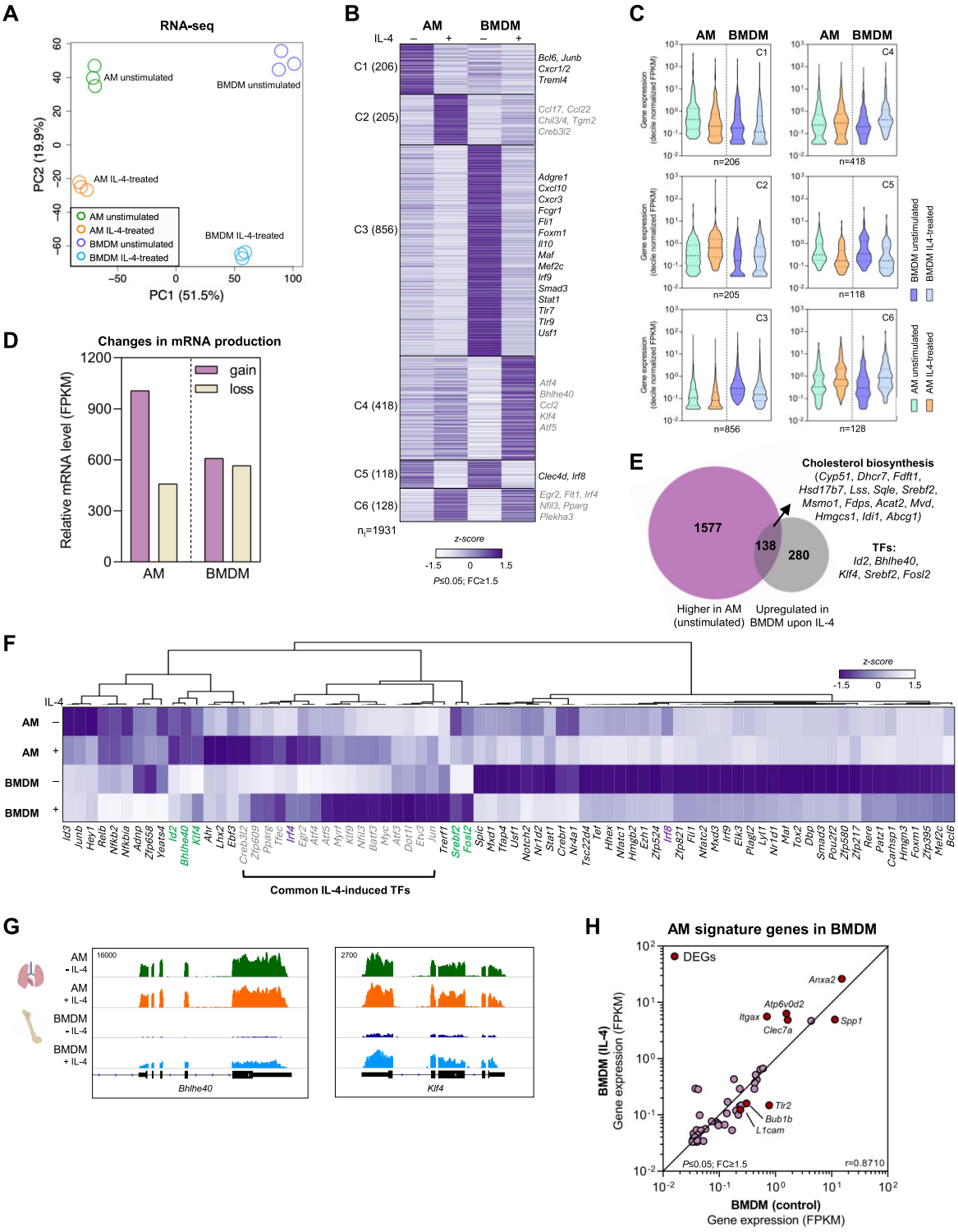


Figure 4. Alternative polarization is a diverse process in AMs and BMDMs. (**A**) PCA plot represents relationships among the unstimulated and IL-4-stimulated AM and BMDM RNA-seq samples ($n = 3$ biological replicates per condition). (**B**) The row-normalized clustered heat map (C1–C6) depicts the average expression of the cell-type-specific and common DEGs upon IL-4 treatment. Representative genes from each cluster are indicated. (**C**) Violin plots quantify the gene expression levels of the C1–C6 clusters. The middle, dashed line represents the median, and the upper and lower dashed lines indicate the deciles. (**D**) Bar charts represent the summed mRNA levels of genes in AMs and BMDMs upon IL-4 treatment. (**E**) Venn diagram depicts the overlap between the genes of AM-specific ($n = 1715$) and IL-4-treated BMDM-specific ($n = 418$) genes. The top biological pathway of the common genes, the genes involved in this pathway, and the shared TFs are highlighted. (**F**) The row-normalized heat map shows TF-coding genes ($n = 76$) that were significantly induced or repressed by IL-4 treatment. The following genes are highlighted: common IL-4-induced genes (grey); genes with significantly higher expression in unstimulated AMs and IL-4-treated BMDMs (green); and *Irf4* and *Irf8* (purple). (**G**) IGV representation of the RNA-seq coverages of *Bhlhe40* and *Klf4* loci in unstimulated and IL-4-treated AMs and BMDMs. The interval scale is highlighted. (**H**) Scatter plot shows the gene expression of AM signature genes in unstimulated and IL-4-treated BMDMs. DEGs are shown in red. Statistical significance was defined as $P \leq 0.05$ with an FC ≥ 1.5.

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

Alternative polarization cannot alter the CpG requirements of open chromatin regions of AMs and BMDMs

In order to examine how alternative polarization affects chromatin openness, ATAC-seq experiments were also carried out for IL-4-treated AMs and BMDMs. The two cell types showed minimal overlap in the changes in chromatin openness; in general, ∼4.6 times more BMDM-specific sites were affected, and >5 times more regions were opened than closed (Fig. 5A; Supplementary Fig. S5A; Supplementary Table S13). The closing and/or AM-specific regions showed the highest ratio of promoters (Supplementary Fig. S5B), which were enriched for CpGs, while TSS-distal regions used by BMDMs were CpG poor independent of the direction of the change (Fig. 5B, C). During the *de novo* motif enrichment analyses, it turned out that the EGR, EICE, and STAT6 motifs were specific for the opening regions in both cell types, especially for their commonly used regions. While the TRE, CRE, and MARE motifs were more enriched in the opening BMDM-specific regions, the CEBP and RUNX motifs were more enriched in AMs and in the closing BMDM-specific regions. A PU-box was present in each cluster, but its highest density was detected at the closing BMDM-specific regions. In addition, the NF-κB motif appeared in the opening AM-specific regions, and the OCT motif—probably bound by OCT2 (*Pou2f2*)—showed an enrichment in the closing BMDM-specific regions (Fig. 5D; Supplementary Fig. S5C, D).

The opening regions showed characteristic motif enrichments, so we examined the gene expression levels of the induced TFs to find the possible DNA–TF interactions corresponding to these positive changes. Naturally, in the case of the IL-4 signal-dependent STAT6, we did not expect or detect gene expression changes as *Stat6* is already highly expressed, and the EGR motif was easy to connect with *Egr2* [6]. In the case of EICE, there was a gene expression shift, during which the level of *Irf8* is getting lower and *Irf4* is induced. As the overall amount of *Irf4/8* did not change significantly, this suggests that the PU.1/IRF4 heterodimer is more efficient in chromatin opening than PU.1/IRF8 (Supplementary Fig. S5E). Out of the bZIP motifs, the BMDM-specific TRE enrichment was the most significant, which can be coupled with several AP-1 heterodimers; however, due to its extremely high expression, ATF3 appears to be the major partner of cJUN (and other JUNs) in BMDMs (Fig. 5E) [6, 56]. For CRE binding, there are also several candidates, including cJUN and ATF4/5. Although the mRNA levels of the bZIP genes are similar and change similarly in the two cell types, in AMs the top bZIP gene is *Cebpb*, so C/EBPβ may play a role also in opening the AM-specific regions (Fig. 5E).

Altogether, the AM-specific regions opened upon IL-4 treatment showed similar characteristics to the initial AM-specific open chromatin regions with EGR, CEBP, PU-box, RUNX, and TRE motifs, supplemented with STAT6, EICE, and NF-κB motifs, with a similar CpG enrichment. In contrast, the newly opened BMDM-specific regions were also CpG poor despite the presence of the EGR motifs. In accordance with this, approximately a tenth of the opening BMDM-specific regions overlapped with the initial AM-specific open chromatin regions (Fig. 5F). Out of these, 22 regions fall on promoters, and only a few targeted genes showed detectable induction upon IL-4 treatment. These included *Atp6v0d2* and *Bhlhe40*, relevant for AM-ness, as well as three members of the *Bcl2a1* family represented by the identified promoters (Fig. 5F; Supplementary Fig. S5F) [19].

EGR2 has distinct roles in AMs and BMDMs

In the next step, we tested how AMs and BMDMs function in the absence of EGR2. For this, we used myeloid-specific conditional *Egr2* exon 2 KO mice, and carried out the same experiments as for the WT macrophages (Fig. 6A; Supplementary Fig. S6A). Considering all expressed genes, the unstimulated BMDMs and the IL-4-treated AMs show the least difference between the two genotypes (Fig. 6B; Supplementary Fig. S6B). The former can be easily explained by the lower basal mRNA level of *Egr2*, while the latter suggests that there are TFs other than EGR2 responsible for alternative polarization in AMs. As expected, the biggest difference was observed between the IL-4-treated BMDMs. The numbers of DEGs between the genotypes also showed that, out of the four conditions, the unstimulated BMDMs were the least affected, and the IL-4-stimulated BMDMs were the most affected by the lack of EGR2 (Fig. 6C; Supplementary Fig. S6C; Supplementary Table S14). Indeed, while in AMs, ∼41% and ∼33% of the genes up- or down-regulated by IL-4 are dependent on EGR2, respectively, in BMDMs these ratios are ∼75% and ∼56%, respectively, with almost four times more affected genes (Fig. 6D). It is important to note that 675 KO-dependent DEGs are identical in the unstimulated and IL-4-stimulated AMs, representing a high ratio, ∼65% and ∼49% of the DEGs, respectively. In contrast, only ∼26% of the IL-4-treated BMDM-specific DEGs overlapped with the AM-specific DEGs (Fig. 6E).

The contribution of EGR2 to the gene expression patterns varied considerably (Fig. 6F, G). First, we analysed those genes responding to IL-4 in WT macrophages (Fig. 6F). In the AM-specific clusters (C1–C3), primarily the baseline was altered by the lack of EGR2 (with C2 including *Itgax* as a representative example highlighted in Supplementary Fig. S6D), while in BMDMs, in general, rather the IL-4 effects changed: the IL-4 response was weakened or eliminated (C4, C11–C13; e.g. *Il4i1*), enhanced (C5; e.g. *Ccl24* and *Mrc1*), or even reversed (C14; e.g. *Adgre1* and *Il15*). This pattern was also evident in the common clusters C6, C7, and C9, where AMs were not significantly affected, while the IL-4 effect in BMDMs was inhibited by the KO (e.g. *Tgm2*, *Pparg*, and *Myc*). The only cluster, where—similarly to BMDMs—the IL-4 effect itself was inhibited in AMs was C8, and in C10, rather the basal expression levels were altered in both cell types (Fig. 6F; Supplementary Fig. S6D; Supplementary Table S15). Out of the 50 KO-dependent TF genes, most were located in C4, C8, and C9, indicating that their IL-4-dependent induction can be eliminated by the KO. In other words, in BMDMs, 45 TF genes were direct or indirect targets of EGR2 upon IL-4 treatment (Supplementary Fig. S6E). In contrast, in AMs, 12 TF genes of clusters II and IV showed an IL-4- and EGR2-dependent induction. The remaining *Egr2* KO-dependent genes did not show statistically significant changes upon IL-4 treatment, except when the changes occurred in the KO samples (Fig. 6G; Supplementary Table S16). Clusters 4–7 represented those where the lack of EGR2 enabled an IL-4 response in AMs, although this was already present for BMDMs in C5; and C10 contained genes that become highly induced, instead of repressed, by IL-4 in the KO BMDMs (Fig. 6G). The two clusters where cell-type-specific IL-4 responsiveness emerged in

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

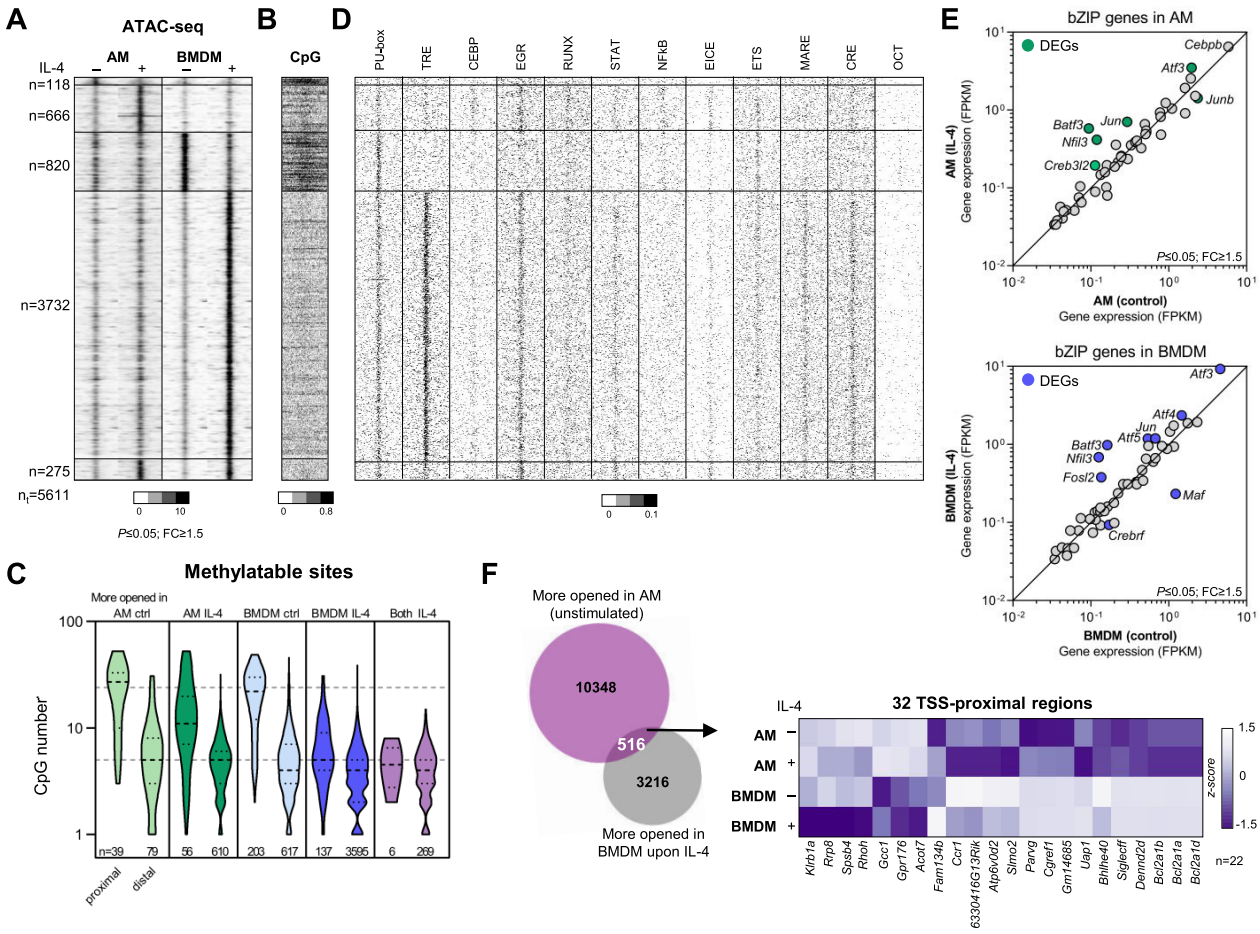


Figure 5. Alternative polarization cannot alter the CpG requirements of open chromatin regions of AMs and BMDMs. (**A**) Read distribution heat map represents the chromatin openness of the differentially accessible regions between the unstimulated and IL-4-stimulated AMs and BMDMs. TSS-proximal and TSS-distal regions are not separated, and clusters containing <10 regions are not shown. Differential analysis was performed using two biological replicates per condition, while a representative sample per condition is shown. (**B**) Heat map shows the CpG density across the regions represented on (A). (**C**) Violin plots show the CpG number in regions more open in unstimulated or IL-4-treated AMs, in unstimulated or IL-4-treated BMDMs, or in both IL-4-treated AMs and BMDMs. Regions are separated into TSS-proximal and TSS-distal groups. The middle, dashed line represents the median, and the upper and lower dashed lines indicate the deciles. (**D**) Motif distribution heat map shows the presence of the *de novo* enriched motifs in 20 bp windows within 2 kb frames relative to the peak centres. (**E**) Scatter plot represents gene expression levels of bZIP genes in unstimulated and IL-4-stimulated AMs and BMDMs. The DEGs are highlighted in colour. (**F**) Venn diagram represents the overlap between the more accessible chromatin regions of unstimulated AMs (compared with unstimulated BMDMs; $n = 10\,384$) and more accessible chromatin regions of BMDM upon IL-4 treatment (compared with unstimulated BMDMs; $n = 3732$). Both TSS-proximal and TSS-distal regions are considered. Genes of the overlapping TSS-proximal regions ($n = 22$) are depicted on a clustered, row-normalized heat map. Statistical significance was defined as $P \leq 0.05$ with an FC ≥ 1.5.

the KO were enriched for collagen synthesis (C6, AM) and immune functions (C10, BMDM) (Supplementary Fig. S6F).

As a conclusion, the loss of EGR2 had a much stronger impact on IL-4-stimulated BMDMs than on AMs, indicating that EGR2 is a key of alternative polarization in BMDMs, whereas AMs rely more on other TFs. In AMs, EGR2 deficiency mainly affected basal gene expression, while in BMDMs, it largely altered or abolished IL-4-induced transcriptional responses. Overall, these findings highlight distinct, cell-type-specific roles of EGR2 in shaping macrophage gene expression and IL-4 responsiveness.

EGR2 is a key contributor to chromatin openness in both AMs and BMDMs

Finally, we extended our loss-of-function study of EGR2 to chromatin accessibility, and a significant number of genomic regions were altered by the KO, even in the TSS-proximal or distal regions (Fig. 7A; Supplementary Fig. S7A; Supplementary Tables S17 and S18). TSS-distal regions were grouped into seven clusters, two of which were specific for AMs, unaffected by IL-4, and repressed by the *Egr2* KO (Fig. 7A). While cluster 1 could not be opened upon IL-4 treatment in BMDMs, cluster 2 showed IL-4- and EGR2-dependent chromatin opening, which—based on our previous findings—correlated with the CpG content of the regions (Fig. 7B; Supplementary Fig. S7C left). Specifically, cluster 1, which contained more CpGs, could not be opened in BMDMs, whereas cluster 2, with fewer CpGs, was opened. Clusters 3 and 4 were also specific for AMs; both become open in the *Egr2* KO, but the latter contained a large number of CpGs. The last three clusters were BMDM specific with EGR2-dependent changes and low CpG levels: clusters 5 and 6 opened upon IL-4 treatment, and cluster 7 closed (Fig. 7A, B).

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

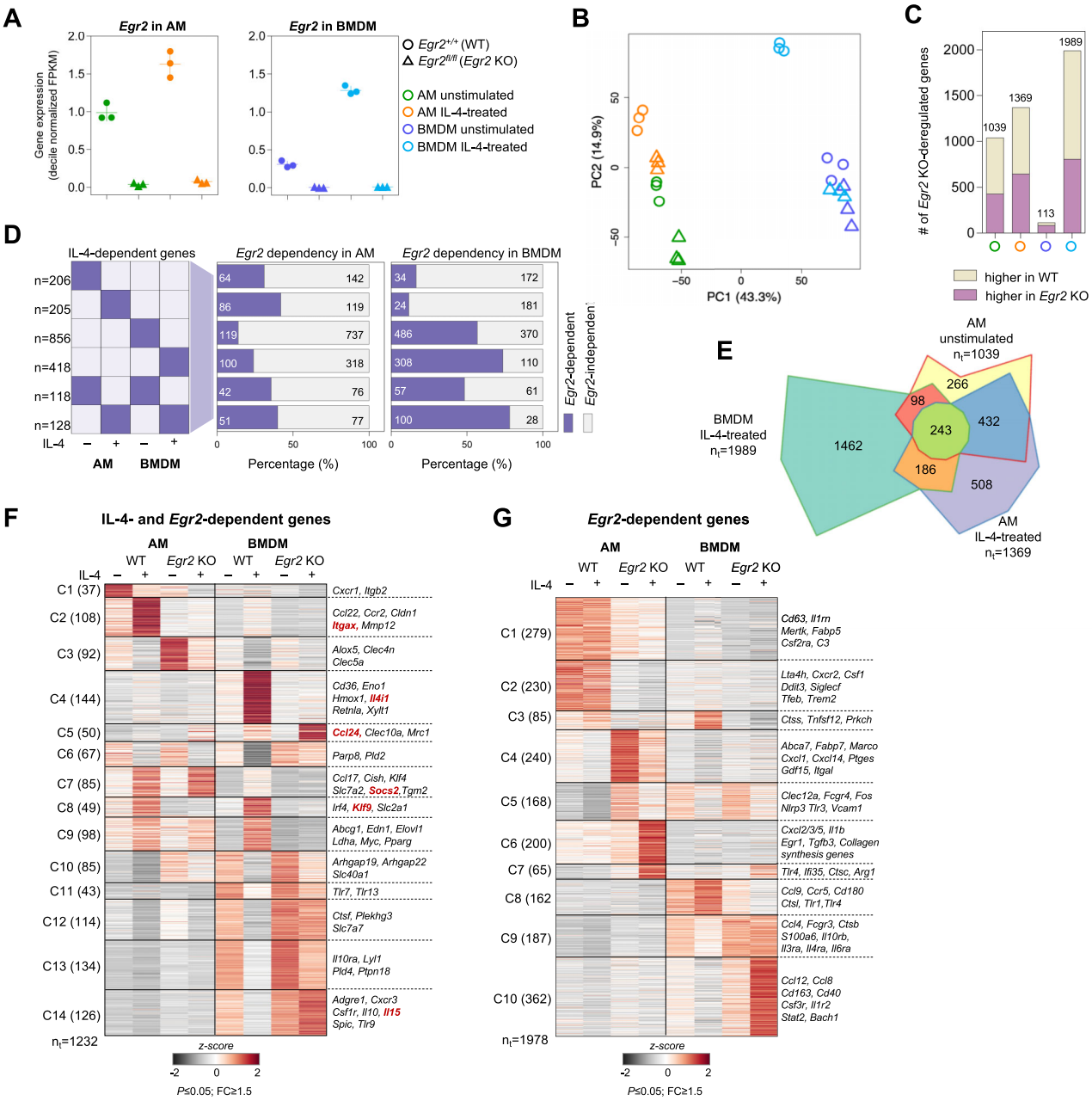


Figure 6. EGR2 has distinct roles in AMs and BMDMs. (**A**) Dot-and-whisker plots depict the gene expression level of *Egr2* in *Egr2*$^{+/+}$ (WT) and *Egr2*$^{fl/fl}$ (*Egr2* KO) AMs and BMDMs in unstimulated and IL-4-stimulated conditions ($n = 3$ biological replicates per condition). (**B**) PCA plot shows the relationships between the gene expression pattern of unstimulated and IL-4-stimulated WT and *Egr2* KO AMs and BMDMs. (**C**) Bar charts show the number of the *Egr2* KO-deregulated genes in all conditions. (**D**) Stacked bar charts (next to the schematic representation of IL-4-deregulated gene clusters of Fig. 4) visualized the ratio of *Egr2*-dependent and *Egr2*-independent genes in AMs and BMDMs separately for each cluster. (**E**) Chow–Ruskey diagram highlights the overlap between the DEGs of the IL-4-treated BMDMs, IL-4-treated AMs, and the unstimulated AMs. (**F** and **G**) Row-normalized, clustered heat maps show the (F) IL-4- and EGR2-dependent (C1–C14) and the (G) *Egr2*-dependent (C1–C10) genes in WT and *Egr2* KO AMs and BMDMs before and after IL-4 treatment. Representative genes from each cluster are highlighted. Genome browser representations of the genes highlighted in red in (F) are shown in the Supplementary figures. Statistical significance was defined as $P \leq 0.05$ with an FC $\geq$ 1.5.

Promoters were divided into four clusters (Supplementary Fig. S7A): the first showed an EGR2-dependent weak openness in AMs; the second was more open in the KO AMs and showed reduced openness upon IL-4 treatment in BMDMs; the third exhibited an EGR2- and IL-4-dependent opening in BMDMs; and the last behaved similarly in both cell types, showing higher openness in the KO (Supplementary Fig. S7A). The CpG content correlated with chromatin openness: the second promoter cluster was extremely CpG rich, while the others contained fewer CpGs than the average (Supplementary Fig. S7B, C right).

The TSS-distal motif enrichments showed clearly distinct patterns in each cluster (Fig. 7C; Supplementary Fig. S7D). Clusters 1 and 2 expectedly resulted in similar motif enrichments including the EGR, CEBP, PU-box, and TRE motifs, but also exhibited characteristic differences in accordance with the previous results. While the CEBP and M-box motifs were more specific for cluster 1—which could not be opened in

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

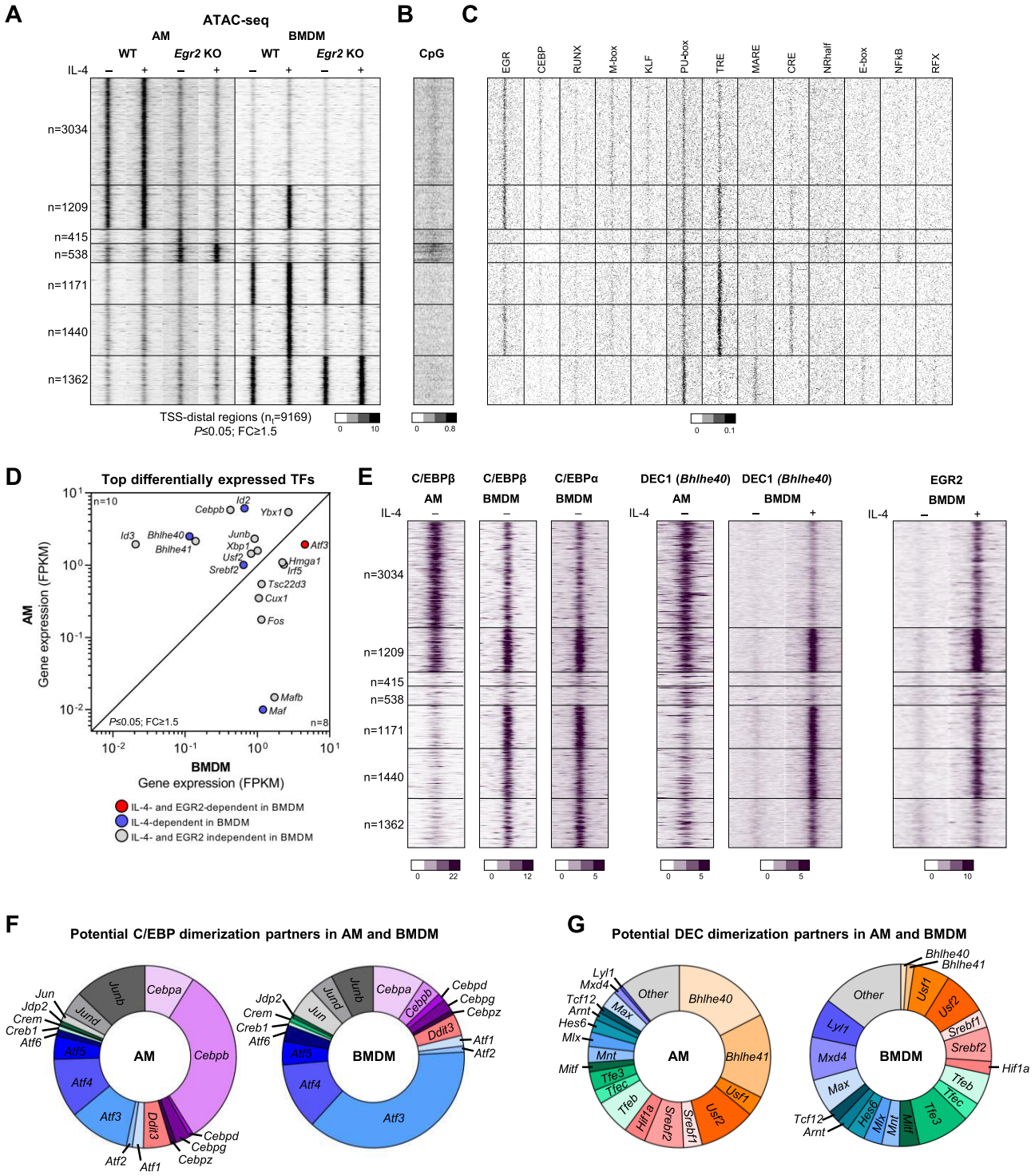


Figure 7. Identification of the keys of AM-ness revealed the necessity of C/EBPβ and DEC dimers. (**A**) Read distribution heat map represents the chromatin openness of the differentially accessible TSS-distal regions ($n = 9169$) between the unstimulated and IL-4-stimulated WT and *Egr2* KO AMs and BMDMs. Differential analysis was performed using two biological replicates per condition, while a representative sample per condition is shown. (**B**) Heat map depicts the CpG density across the regions represented in (A). (**C**) Motif distribution heat map shows the presence of the *de novo* enriched motifs in 20 bp windows within 2 kb frames relative to the peak centres. (**D**) Scatter plot shows the top differentially expressed TFs in unstimulated AMs and BMDMs. Genes that are IL-4 and EGR2 dependent in BMDMs or IL-4 dependent in BMDMs are highlighted in colour. (**E**) Read distribution heat maps depict C/EBPβ binding in unstimulated AMs and BMDMs, C/EBPα binding in unstimulated BMDMs, DEC1 binding in unstimulated AMs, and DEC1 and EGR2 binding in unstimulated and IL-4-stimulated BMDMs. (**F** and **G**) Pie charts show the potential C/EBP (F) and DEC (G) dimerization partners based on their gene expression levels in AMs and BMDMs.

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

BMDMs—the EGR, TRE, and PU-box motifs were more enriched in cluster 2. The AM-specific clusters 3 and 4—with similar chromatin changing tendencies—showed enrichment for the general PU-box and TRE motifs. The BMDM-specific clusters 5 and 6 were also similar to each other, except for the basal signals and the extent of the changes. While the already open regions showed strong enrichment of PU-box, TRE, and CRE, the largely EGR2- and IL-4-dependent *de novo* opening regions displayed TRE, EGR, PU-box, and CRE motifs as the top hits. The last BMDM-specific cluster, which was slightly repressed by IL-4 in an EGR2-dependent manner, was enriched for the PU-box, MARE, RUNX, and E-box motifs (Fig. 7C; Supplementary Fig. S7D). In the case of the promoters, the EGR2-dependent regions were indeed enriched for the EGR motif; the second cluster was enriched for CpG-rich and CCAAT-box motifs; and the last two clusters contained respectively TRE, PU-box, EGR, and KLF, or the typical promoter elements (Supplementary Fig. S7A, E, F).

Unsurprisingly, independent of the IL-4 treatment of AMs, the direct DNA binding by EGR2 contributed decisively to chromatin openness, whereas in BMDMs, EGR2 acted as a positive effector of IL-4, regarding chromatin opening. In the former case, the main differential partners were C/EBP and bHLH dimers, while in the latter case, they were AP-1 dimers.

C/EBPβ and DEC dimers are IL-4-independent keys of AM identity

Although IL-4 induced the expression of several AM-specific TFs in BMDMs, none of these factors appears to account for the AM-specific transcriptional programme. Therefore, we focused on AM-specific TF genes whose expression could not be induced by IL-4 in BMDMs. Six of the 10 most AM-enriched TF genes belonged to the bHLH superfamily, and three of them could not be induced by IL-4 in BMDMs—these are *Id3*, *Bhlhe41*, and *Usf2*. The function of ID proteins is to repress other bHLH proteins, but DEC2 (*Bhlhe41*) and USF2 dimers could be responsible for the AM-specific M-box enrichment (Fig. 7D) [58]. Since both *Usf1* and *Usf2* showed high expression, and USF2 displayed an extended cistrome in BMDMs, DEC2 dimers are likely to be the main differentiating factors between the two cell types [6]. The next over-represented TF family was the bZIP group, including *Cebpb*, *Junb*, and *Xbp1*. C/EBPβ could account for the AM-specific CEBP motif enrichments; JUNB is a TRE and CRE binder, whose motifs were less enriched in AMs; and XBP1, an ER stress-dependent TF, is probably less relevant in the basic state [6, 56, 64]. The protein product of the last AM-specific gene, *Ybx1,* can theoretically function as a CCAAT (Y)-box binding TF, although its TF-ness had been questioned in a more recent study [65]. Half of the BMDM-specific TF genes belonged to the bZIP family, encoding the TRE-binding AP-1 (*Atf3* and *Fos*) and MARE-binding large MAF (*Maf* and *Mafb*) proteins. For the other TFs, no motif enrichment was identified.

In order to confirm the direct DNA–TF interactions, we mapped densities of the available C/EBPα/β partners and the newly generated DEC1 ChIP-seq data to the seven TSS-distal clusters (Fig. 7E) [14, 17, 43, 45, 66]. C/EBPβ binding showed a cell-type-specific pattern: in AMs, it was restricted to AM-specific DORs, while in BMDMs, it co-localized with C/EBPα. This means that in BMDMs—besides clusters 5–7 and cluster 1 to a certain extent—cluster 2 was already occupied by C/EBPα and C/EBPβ, poised for activation by specific stimuli. DEC1 followed a very similar pattern: it bound together with C/EBPβ in AMs, and co-localized with C/EBPα and C/EBPβ in BMDMs upon IL-4 treatment. As a control, we also examined the EGR2 enrichment in BMDMs, which corresponded to the presence of its specific motifs (Fig. 7E).

Finally, as C/EBPβ and DEC2 appeared to be the IL-4-independent key determinants of AM-ness, we checked the relative ratio of potential C/EBP dimerization partners, including CREB/ATF and JUN proteins, and bHLH proteins based on their mRNA levels (Fig. 7F, G) [23]. Out of the related bZIP genes, in AMs, *Cebpb* showed the highest mRNA level by far, and as the enriched CEBP motifs did not show CEBP:ATF response element (CARE) characteristics, C/EBPβ homodimers probably dominated the AM-specific DORs, followed by ATF and JUN dimers [56]. In contrast, in BMDMs, C/EBPα was predominant over the other C/EBPs, while ATFs, especially ATF3, and JUNs probably played a more prominent role (Fig. 7F; Supplementary Fig. S7G). In the other case, out of the expressed bHLH superfamily members, *Bhlhe40* and *Bhlhe41* are the top transcripts in AMs, while they show a low expression in BMDMs (Fig. 7G).

Taken together, several lines of evidence, including gene expression levels, motif enrichments, and DNA–TF interactions, indicate that C/EBPβ dimers and DEC1/DEC2 heterodimers are likely to be key regulators of the AM-specific CpG-rich enhancer landscape. However, JUNB and RUNX2 also emerge as candidates contributing to the establishment of AM-specific DORs (Fig. 2B, H).

Discussion

In this study, we compared AMs and BMDMs, and identified large differences in both their specific *cis*-regulatory regions and their gene expression patterns. We revealed an epigenetic barrier that distinguishes BMDMs from monocytes and TRMs—including AMs—and found it impenetrable by the TFs induced by IL-4. Specifically, BMDMs preferentially utilize rather CpG-poor enhancers, while the examined TRMs tend to use another kind of region with CpG enrichment. Furthermore, induction of the AM-specific TFs such as EGR2, KLF4, and DEC1 in BMDMs is not sufficient for the activation of AM-determining *cis*-regulatory regions and target genes. In certain cases, this can be due to the direct methylation of TF recognition sites, as EGR and KLF elements or even M-boxes may be directly inactivated through their CpGs. However, since most of these elements are CpG free or demethylated, they should remain available for TF binding in BMDMs. This suggests that some 'pioneer' factors, either absent, dormant, or not fully functional in BMDMs, are required to open up the AM-specific CpG-rich regions.

By integrating chromatin accessibility, motif enrichment, and TF expression analyses, we identified DEC2, C/EBPβ, and JUNB as the most likely regulators of AM identity. This conclusion was supported by several converging observations: all three factors are highly expressed in AMs, show substantially lower expression in BMDMs that is not restored by IL-4 treatment, and can be linked to motifs enriched in AM-specific DORs. DEC2 was supported by the strong AM-specific expression of *Bhlhe41* and the enrichment of M-box motifs. Although direct DEC2 binding could not be demonstrated, the DEC1 cistrome is consistent with the possible involvement of DEC1/DEC2 heterodimers [17]. The role of C/EBPβ was supported by its high AM-specific expression, the enrichment of

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

CEBP motifs, and its direct occupancy of AM-specific regulatory regions. JUNB was also considered a candidate based on its high expression in AMs and its potential role in TRE- and CRE-mediated transcriptional regulation. Collectively, these factors may represent an AM-enriched TF repertoire.

The putative elements of these TFs typically cannot be repressed by DNA methylation, so the pioneer function can be extended to the proximal CpG sequences. Importantly, although C/EBPα and C/EBPβ have not been definitively established as pioneer factors, they exhibit several pioneer-like properties, including interactions with the SWI/SNF chromatin-remodelling complex, transcriptional co-regulators, and the Mediator complex. Moreover, both are macrophage LDTFs [21, 67, 68, 69]. In contrast, there is currently no direct evidence that DEC2 functions as a pioneer factor, partly owing to the lack of a suitable antibody, but it is indispensable for Th2 differentiation, a process associated with JUNB activation, and—together with DEC1—contributes to AM identity. These converging lines of evidence raise the possibility that DEC2 acts as a lineage-restricted regulator with pioneer-like functions in Th2 cells and AMs [17, 70].

In general, G/C-rich regions, even outside canonical CpG islands and independent of cytosine methylation, tend to form a less bendable DNA double helix that disfavours nucleosome assembly, thereby correlating with increased chromatin accessibility and specific nucleosome positioning [25, 71]. According to our findings, G/C and CpG content indeed contributes to chromatin accessibility, especially at CpG islands, but comparable effects are also observed across diverse *cis*-regulatory elements. The advantage of G/C content around TF-binding sites is that it is a lower-risk evolutionary tool to assign and fine-tune the function of enhancers. This implies that not all the pioneer factor characteristics are required for the activation/activity of these regions: an elevated gene expression—a major characteristic of LDTFs—is sufficient for cell-type-specific enhancer selection. Moreover, a higher G/C content suggests roles beyond DNA structure and methylation control. Similar to DNA, high G/C content influences the structure of the produced enhancer RNAs (eRNAs), as distinct G/C-rich segments can form stem–loops or even functional eRNA regulatory motifs (FERMs), which can interact with and activate the locally assembling regulatory complexes including, for example, p300 [72, 73]. We assume that CpG-flanked *cis*-regulatory elements can be freely and cell type-specifically used by cell type-specific (LD)TFs, which initiate the formation of co-activator complexes. Thus, if a TRM expresses the specific set of TFs, they can access the pre-assigned enhancers to perform their function, with limited interconversion between cell types. Because of cell type specificity and low risk of side effects, CpG-rich enhancers represent promising gene therapy targets.

Based on our results, BMDMs do not have their own specific CpG-rich enhancers, although shared enhancers contain an elevated CpG level relative to random genomic regions. This phenomenon has been previously observed in pluripotent cells, while differentiated cells instead use enhancers with CpG enrichment [74]. Recently, it has been shown in epidermal stem cells that CpG-rich regulatory regions are able to maintain persistent inflammatory memory through sustained chromatin accessibility, whereas CpG-poor regions are generally associated with more transient and reversible inflammatory responses [75]. These observations suggest that CpG-enriched regulatory architectures recurrently associate with stable forms of cellular adaptation and commitment. In this context, the use of CpG-rich enhancers by TRMs may reflect an epigenomically engaged regulatory state, supporting long-term tissue adaptation, compared with BMDMs. At the same time, the high *Maf/Mafb* expression and MARE motif enrichment observed in BMDMs are consistent with the established role of large MAFs in terminal macrophage differentiation and restricted self-renewal [76]. This likewise stands in contrast to what was observed in AMs [59]. These may reflect the naïve and plastic state of BMDMs, having a potential to respond extensively to IL-4 or other signals. In BMDMs, >1500 genes show a significant change upon IL-4 treatment, and EGR2 transmits 75% of the positive and more than half of the negative IL-4/STAT6 signals by activating numerous downstream TFs. In contrast, AMs are burdened by IL-4, with an increased mRNA production leading to ER stress. In this case, EGR2 has a less significant role, and also the changes are milder.

Our findings indicate that BMDMs do not closely resemble TRMs and instead occupy a distinct transcriptional and epigenetic state, suggesting that the identity of *in vitro*-derived macrophages remains elusive. While BMDMs offer practical advantages, including relative homogeneity, genomic stability, reproducibility, and rapid responsiveness to perturbations, these same features also limit their ability to capture the heterogeneity, functional specialization, and environmental imprinting of TRMs *in vivo*. Nevertheless, selected response programmes, such as IL-4-induced activation, remained directionally conserved between BMDMs and AMs despite marked differences in baseline state and response magnitude. Together, these findings probably explain the broad utility of BMDMs for studying general polarization dynamics under controlled conditions, while also highlighting that they should not be considered direct models of TRMs. Overall, our results establish CpG-dependent enhancer usage as a key determinant separating *in vitro* differentiated macrophages from their tissue-resident counterparts.

Limitations of the study

Although our study integrates multiple datasets to uncover the CpG-dependent enhancer logic in macrophages, several limitations should be considered. First, the analyses were limited to mouse cells, and the extent to which CpG-dependent enhancer selection applies in human macrophages remains to be determined. Second, functional validation was not performed; for instance, C/EBPβ and/or DEC2 overexpression in BMDMs could help to determine whether they are sufficient to activate AM-specific CpG-rich enhancers. In addition, the patterns of DNA binding by DEC2, represented, for example, by ChIP-seq, could show the cooperation of DEC proteins in AMs. Finally, the IL-4 stimulation model captures only one polarization pathway, and additional cytokines or environmental stimuli may reveal distinct regulatory mechanisms of enhancer usage.

Acknowledgements

The authors would like to offer special thanks to Dr Szilárd Póliska for his expert technical assistance with sequencing (Genomic Medicine and Bioinformatic Core Facility, University of Debrecen, Debrecen, Hungary). The authors also thank Dr István Szatmári (Department of Biochemistry and Molecular Biology, University of Debrecen, Debrecen, Hungary) and

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

members of the Nagy laboratory for discussions and comments on the manuscript. The graphical abstract was created in BioRender. Barath, B.R. (2026) https://BioRender.com/nfq1skx.

Author contributions D.B., G.N., and L.N. planned the project. Z.V., Z.K., Z.C., P.T., and T.C. isolated the cells and carried out the ATAC-seq and RNA-seq experiments. D.B. and G.N. performed the computational analyses and designed the figures. N.C.S. and A.P. contributed to the interpretation of the data. G.N. and L.N. directed and supervised the work. D.B. and G.N. drafted the manuscript. All authors discussed the results and commented on the manuscript.

Supplementary data

Supplementary data are available at NAR online.

Conflict of interest

The authors declare no competing interests.

Funding

The Hungarian Scientific Research Fund [support for the Nuclear Hormone Receptor Laboratory, OTKA PD137902 to D.B., FK146945 to G.N., K147147, and ADVANCED-152422 to L.N.]; the Hungarian Academy of Sciences [KGYNK2026-43 to D.B.]; the National Institutes of Health [DK115924, DK124782, and HL170426 to L.N.]; the European Union's Horizon 2020 research and 730 innovation programme under the Marie Skłodowska-Curie Actions [860034 to N.C.-S.]; Johns Hopkins All Children's Foundation (JHACF)/All Children's Research Institute (ACRI) [fellowship 90105887 to A.P.]; the Muscular Dystrophy Association (MDA) [1064247 to A.P.]; the National Research, Development and Innovation Office [2023-1.1.1-PIACI_FÓKUSZ-2024-00036, K145994, EXCELLENCE_151396 to Z.C.]; and the Hungarian Academy of Sciences János Bolyai Research Scholarship [BO/00594/22/8 to Z.C.].

Data availability

The generated RNA-seq and ATAC-seq datasets are available in the Gene Expression Omnibus (GEO) repository under GSE317131. All public data used in this study and the corresponding access information are listed in Supplementary Table S19. This study did not generate new code. All analyses were performed using publicly available software, and all parameters and versions are described in the Materials and methods.

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Gergely Nagy (nagygergely@med.unideb.hu).

References

1. Svedberg FR, Brown SL, Krauss MZ *et al*. The lung environment controls alveolar macrophage metabolism and responsiveness in type 2 inflammation. *Nat Immunol* 2019;20:571–80. https://doi.org/10.1038/s41590-019-0352-y
2. Barish GD, Yu RT, Karunasiri M *et al*. Bcl-6 and NF-kappaB cistromes mediate opposing regulation of the innate immune response. *Genes Dev* 2010;24:2760–5. https://doi.org/10.1101/gad.1998010
3. Ostuni R, Piccolo V, Barozzi I *et al*. Latent enhancers activated by stimulation in differentiated cells. *Cell* 2013;152:157–71. https://doi.org/10.1016/j.cell.2012.12.018
4. Daniel B, Nagy G, Hah N *et al*. The active enhancer network operated by liganded RXR supports angiogenic activity in macrophages. *Genes Dev* 2014;28:1562–77. https://doi.org/10.1101/gad.242685.114
5. Eichenfield DZ, Troutman TD, Link VM *et al*. Tissue damage drives co-localization of NF-κB, Smad3, and Nrf2 to direct Rev–erb sensitive wound repair in mouse macrophages. *eLife* 2016;5:e13024. https://doi.org/10.7554/eLife.13024
6. Nagy G, Bojcsuk D, Tzerpos P *et al*. Lineage-determining transcription factor-driven promoters regulate cell type-specific macrophage gene expression. *Nucleic Acids Res* 2024;52:4234–56. https://doi.org/10.1093/nar/gkae088
7. Cobo I, Murillo-Saich J, Alishala M *et al*. Particle uptake by macrophages triggers bifurcated transcriptional pathways that differentially regulate inflammation and lysosomal gene expression. *Immunity* 2025;58:826–42. https://doi.org/10.1016/j.immuni.2025.02.023
8. Fejer G, Sharma S, Gyory I. Self-renewing macrophages—a new line of enquiries in mononuclear phagocytes. *Immunobiology* 2015;220:169–74. https://doi.org/10.1016/j.imbio.2014.11.005
9. Ginhoux F, Jung S. Monocytes and macrophages: developmental pathways and tissue homeostasis. *Nat Rev Immunol* 2014;14:392–404. https://doi.org/10.1038/nri3671
10. Aegerter H, Kulikauskaite J, Crotta S *et al*. Influenza-induced monocyte-derived alveolar macrophages confer prolonged antibacterial protection. *Nat Immunol* 2020;21:145–57. https://doi.org/10.1038/s41590-019-0568-x
11. Patsalos A, Pap A, Varga T *et al*. In situ macrophage phenotypic transition is affected by altered cellular composition prior to acute sterile muscle injury. *J Physiol* 2017;595:5815–42. https://doi.org/10.1113/JP274361
12. van de Laar L, Saelens W, De Prijck S *et al*. Yolk sac macrophages, fetal liver, and adult monocytes can colonize an empty niche and develop into functional tissue-resident macrophages. *Immunity* 2016;44:755–68. https://doi.org/10.1016/j.immuni.2016.02.017
13. Daniel B, Nagy G, Czimmerer Z *et al*. The nuclear receptor PPARγ controls progressive macrophage polarization as a ligand-insensitive epigenomic ratchet of transcriptional memory. *Immunity* 2018;49:615–26. https://doi.org/10.1016/j.immuni.2018.09.005
14. Daniel B, Czimmerer Z, Halasz L *et al*. The transcription factor EGR2 is the molecular linchpin connecting STAT6 activation to the late, stable epigenomic program of alternative macrophage polarization. *Genes Dev.* 2020;34:1474–92. https://doi.org/10.1101/gad.343038.120
15. Daniel B, Nagy G, Horvath A *et al*. The IL-4/STAT6/PPARγ signaling axis is driving the expansion of the RXR heterodimer cistrome, providing complex ligand responsiveness in macrophages. *Nucleic Acids Res* 2018;46:4425–39. https://doi.org/10.1093/nar/gky157
16. Liao X, Sharma N, Kapadia F *et al*. Krüppel-like factor 4 regulates macrophage polarization. *J Clin Invest* 2011;121:2736–49. https://doi.org/10.1172/JCI45444
17. Rauschmeier R, Gustafsson C, Reinhardt A *et al*. Bhlhe40 and Bhlhe41 transcription factors regulate alveolar macrophage self-renewal and identity. *EMBO J* 2019;38:e101233. https://doi.org/10.15252/embj.2018101233
18. Pestal K, Slayden LC, Barton GM. KLF family members control expression of genes required for tissue macrophage identities. *J Exp Med* 2025;222 (5):e20240379. https://doi.org/10.1084/jem.20240379
19. Kolostyak Z, Bojcsuk D, Baksa V *et al*. EGR2 is an epigenomic regulator of phagocytosis and antifungal immunity in alveolar macrophages. *JCI Insight* 2024;9 (17):e164009. https://doi.org/10.1172/jci.insight.164009
20. McCowan J, Fercoq F, Kirkwood PM *et al*. The transcription factor EGR2 is indispensable for tissue-specific imprinting of

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

alveolar macrophages in health and tissue repair. *Sci Immunol* 2021;6:eabj2132. https://doi.org/10.1126/sciimmunol.abj2132
21. Heinz S, Benner C, Spann N *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89. https://doi.org/10.1016/j.molcel.2010.05.004
22. Xie H, Ye M, Feng R *et al.* Stepwise reprogramming of B cells into macrophages. *Cell* 2004;117:663–76. https://doi.org/10.1016/S0092-8674(04)00419-2
23. Tsukada J, Yoshida Y, Kominato Y *et al.* The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. *Cytokine* 2011;54:6–19. https://doi.org/10.1016/j.cyto.2010.12.019
24. Bret C, Desmots-Loyer F, Moreaux J *et al.* BHLHE41, a transcriptional repressor involved in physiological processes and tumor development. *Cell Oncol* 2025;48:43–66. https://doi.org/10.1007/s13402-024-00973-3
25. Fenouil R, Cauchy P, Koch F *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res* 2012;22:2399–408. https://doi.org/10.1101/gr.138776.112
26. Yin Y, Morgunova E, Jolma A *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 2017;356:eaaj2239. https://doi.org/10.1126/science.aaj2239
27. Héberlé É, Bardet AF. Sensitivity of transcription factors to DNA methylation. *Essays Biochem* 2019;63:727–41. https://doi.org/10.1042/EBC20190033
28. Czimmerer Z, Daniel B, Horvath A *et al.* The transcription factor STAT6 mediates direct repression of inflammatory enhancers and limits activation of alternatively polarized macrophages. *Immunity* 2018;48:75–90. https://doi.org/10.1016/j.immuni.2017.12.010
29. Daniel B, Balint BL, Nagy ZS *et al.* Mapping the genomic binding sites of the activated retinoid X receptor in murine bone marrow-derived macrophages using chromatin immunoprecipitation sequencing. *Methods Mol Biol* 2014;1204:15–24. https://doi.org/10.1007/978-1-4939-1346-6_2
30. Kim D, Paggi JM, Park C *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;37:907–15. https://doi.org/10.1038/s41587-019-0201-4
31. Pertea M, Pertea GM, Antonescu CM *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;33:290–5. https://doi.org/10.1038/nbt.3122
32. Li H, Handsaker B, Wysoker A *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352
33. Ramírez F, Dündar F, Diehl S *et al.* DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187–91. https://doi.org/10.1093/nar/gku365
34. Lavin Y, Winter D, Blecher-Gonen R *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 2014;159:1312–26. https://doi.org/10.1016/j.cell.2014.11.018
35. Krämer A, Green J, Pollard J *et al.* Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 2014;30:523–30. https://doi.org/10.1093/bioinformatics/btt703
36. Zhou Y, Zhou B, Pache L *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1–10. https://doi.org/10.1038/s41467-019-09234-6
37. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324
38. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol* 2008;9:R137. https://doi.org/10.1186/gb-2008-9-9-r137
39. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep* 2019;9:1–5. https://doi.org/10.1038/s41598-019-45839-z
40. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033
41. McQuattie-Pimentel AC, Ren Z, Joshi N *et al.* The lung microenvironment shapes a dysfunctional response of alveolar macrophages in aging. *J Clin Invest* 2021;131 (4):e140299. https://doi.org/10.1172/JCI140299
42. Saeki K, Pan R, Lee E *et al.* IRF8 defines the epigenetic landscape in postnatal microglia, thereby directing their transcriptome programs. *Nat Immunol* 2024;25:1928–42. https://doi.org/10.1038/s41590-024-01962-2
43. Link VM, Duttke SH, Chun HB *et al.* Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. *Cell* 2018;173:1796–809. https://doi.org/10.1016/j.cell.2018.04.018
44. Li Z, Mao K, Liu L *et al.* Nuclear microRNA-mediated transcriptional control determines adult microglial homeostasis and brain function. *Cell Rep* 2024; 43:113964. https://doi.org/10.1016/j.celrep.2024.113964
45. Horvath A, Daniel B, Szeles L *et al.* Labelled regulatory elements are pervasive features of the macrophage genome and are dynamically utilized by classical and alternative polarization signals. *Nucleic Acids Res* 2019;47:2778–92. https://doi.org/10.1093/nar/gkz118
46. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–92. https://doi.org/10.1093/bib/bbs017
47. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* 2017;18:1–8. https://doi.org/10.1186/s12859-017-1708-7
48. Hulsen T, de Vlieg J, Alkema W. BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 2008;9:488. https://doi.org/10.1186/1471-2164-9-488
49. Saldanha AJ. Java treeview—extensible visualization of microarray data. *Bioinformatics* 2004;20:3246–8. https://doi.org/10.1093/bioinformatics/bth349
50. Gautier EL, Shay T, Miller J *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol* 2012;13:1118–28. https://doi.org/10.1038/ni.2419
51. Heinz S, Romanoski CE, Benner C *et al.* The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* 2015;16:144–54. https://doi.org/10.1038/nrm3949
52. Lichtinger M, Ingram R, Hannah R *et al.* RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J* 2012;31:4318–33. https://doi.org/10.1038/emboj.2012.275
53. Ning S, Pagano JS, Barber GN. IRF7: activation, regulation, modification and function. *Genes Immun* 2011;12:399–414. https://doi.org/10.1038/gene.2011.21
54. Buenrostro JD, Giresi PG, Zaba LC *et al.* Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;10:1213–8. https://doi.org/10.1038/nmeth.2688
55. Pogenberg V, Ögmundsdóttir MH, Bergsteinsdóttir K *et al.* Restricted leucine zipper dimerization and specificity of DNA recognition of the melanocyte master regulator MITF. *Genes Dev* 2012;26:2647–58. https://doi.org/10.1101/gad.198192.112
56. Nagy G, Nagy L. Motif grammar: the basis of the language of gene expression. *Comput Struct Biotechnol J* 2020;18:2026–32. https://doi.org/10.1016/j.csbj.2020.07.007

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026

57. Schübeler D. Function and information content of DNA methylation. *Nature* 2015;517:321–6. https://doi.org/10.1038/nature14192
58. Ling F, Kang B, Sun XH. Id proteins: small molecules, mighty regulators. *Curr Top Dev Biol* 2014;110:189–216. https://doi.org/10.1016/B978-0-12-405943-6.00005-1
59. Soucie EL, Weng Z, Geirsdóttir L *et al.* Lineage-specific enhancers activate self-renewal genes in macrophages and embryonic stem cells. *Science* 2016;351:aad5510. https://doi.org/10.1126/science.aad5510
60. Ngondo-Mbongo RP, Myslinski E, Aster JC *et al.* Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11. *Nucleic Acids Res* 2013;41:4000–14. https://doi.org/10.1093/nar/gkt088
61. Meraro D, Gleit-Kielmanowicz M, Hauser H *et al.* IFN-stimulated gene 15 is synergistically activated through interactions between the myelocyte/lymphocyte-specific transcription factors, PU.1, IFN regulatory factor-8/IFN consensus sequence binding protein, and IFN regulatory factor-4: characterization of a new subtype of IFN-stimulated response element. *J Immunol* 2002;168:6224–31. https://doi.org/10.4049/jimmunol.168.12.6224
62. Hume DA, Irvine KM, Pridans C. The mononuclear phagocyte system: the relationship between monocytes and macrophages. *Trends Immunol* 2019;40:98–112. https://doi.org/10.1016/j.it.2018.11.007
63. Daassi D, Hamada M, Jeon H *et al.* Differential expression patterns of MafB and c-maf in macrophages in vivo and in vitro. *Biochem Biophys Res Commun* 2016;473:118–24. https://doi.org/10.1016/j.bbrc.2016.03.063
64. Chen X, Shi C, He M *et al.* Endoplasmic reticulum stress: molecular mechanism and therapeutic targets. *Signal Transduct Target Ther* 2023;8:352. https://doi.org/10.1038/s41392-023-01570-w
65. Dolfini D, Mantovani R. YB-1 (YBX1) does not bind to Y/CCAAT boxes in vivo. *Oncogene* 2013;32:4189–90. https://doi.org/10.1038/onc.2012.521
66. Dörr D, Obermayer B, Weiner JM *et al.* C/EBPβ regulates lipid metabolism and *Pparg* isoform 2 expression in alveolar macrophages. *Sci Immunol* 2022;7:eabj0140. https://doi.org/10.1126/sciimmunol.abj0140
67. Pedersen TÅ, Kowenz-Leutz E, Leutz A *et al.* Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. *Genes Dev* 2001;15:3208–16. https://doi.org/10.1101/gad.209901
68. Schwartz C, Beck K, Mink S *et al.* Recruitment of p300 by C/EBPbeta triggers phosphorylation of p300 and modulates coactivator activity. *EMBO J* 2003;22:882–92. https://doi.org/10.1093/emboj/cdg076
69. Li H, Gade P, Nallar SC *et al.* The Med1 subunit of transcriptional mediator plays a central role in regulating CCAAT/enhancer-binding protein-β-driven transcription in response to interferon-γ. *J Biol Chem* 2008;283:13077–86. https://doi.org/10.1074/jbc.M800604200
70. Yang XO, Angkasekwinai P, Zhu J *et al.* Requirement for the basic helix–loop–helix transcription factor Dec2 in initial TH2 lineage commitment. *Nat Immunol* 2009;10:1260–6. https://doi.org/10.1038/ni.1821
71. Marin-Gonzalez A, Vilhena JG, Perez R *et al.* A molecular view of DNA flexibility. *Q Rev Biophys* 2021;54:e8. https://doi.org/10.1017/S0033583521000068
72. Fallmann J, Will S, Engelhardt J *et al.* Recent advances in RNA folding. *J Biotechnol* 2017;261:97–104. https://doi.org/10.1016/j.jbiotec.2017.07.007
73. Hou TY, Kraus WL. Analysis of estrogen-regulated enhancer RNAs identifies a functional motif required for enhancer assembly and gene expression. *Cell Rep* 2022; 39:110944. https://doi.org/10.1016/j.celrep.2022.110944
74. Kreibich E, Kleinendorst R, Barzaghi G *et al.* Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. *Mol Cell* 2023;83:787–802. https://doi.org/10.1016/j.molcel.2023.01.017
75. Cowley CJ, Sajjath SM, Soto-Ugaldi LF *et al.* Distinctive DNA sequence features define epigenetic longevity of inflammatory memory. *Science* 2026;391:adz6830. https://doi.org/10.1126/science.adz6830
76. Chueh AC, Northrop EL, Brettingham-Moore KH *et al.* MafB/c-Maf deficiency enables self-renewal of differentiated functional macrophages. *Science* 2009;326:867–71. https://doi.org/10.1126/science.1176056

Received: February 10, 2026. **Revised:** July 28, 2026. **Accepted:** July 30, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from academic.oup.com/nar/article/54/16/gkag824/8769254 by School of Public Health user on 26 August 2026