IRF8 configures enhancer landscape in postnatal microglia and directs microglia specific transcriptional programs

Keita Saeki [1], Richard Pan[1], Eunju Lee[1], Daisuke Kurotaki[2], Keiko Ozato[1]

1: Division of Developmental Biology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, 20892

2: International Research Center for Medical Sciences, Kumamoto University, 860-011 Kumamoto City, Japan

Running title; IRF8 organizes enhancer modules in microglia

Keywords: IRF8; microglia maturation; enhancer; Alzheimer's disease

Address correspondence to Keiko Ozato (ozatok@nih.gov) or Keita Saeki (saekik2@nih.gov). Section on Molecular Genetics of Immunity, Division of Developmental Biology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, 20892

Abstract

Microglia are innate immune cells in the brain. Transcription factor IRF8 is highly expressed in microglia. However, its role in postnatal microglia development and the mechanism of action is unknown. We demonstrate here that IRF8 binds to enhancer regions of postnatal microglia in a stepwise fashion reaching a maximum after day 14, which coincided with the initiation of microglia function. Constitutive Irf8 deletion led to the loss of microglia identity gene expression and aberrant induction of Alzheimer's disease and neurodegeneration associated genes. Conditional Irf8 deletion in adult microglia showed similar transcriptome profiles, revealing the requirement of continuous IRF8 expression. Additional genome-wide analyses showed IRF8 is critical for setting microglia-specific chromatin accessibility and DNA methylation patterns. Lastly, in the 5xFAD mouse AD model, Irf8 deletion lessened the formation and spread of amyloidβ plaques, thereby reducing neuronal loss. Together, IRF8 sets the microglia epigenome landscape, required for eliciting microglia identity and function.

Abbreviations: AD, Alzheimer's disease; DAM/NDG, Alzheimer's disease associated and neurodegeneration associated genes; DEG, differentially expressed gene; FACS, fluorescence-activated cell sorting; FC, fold change; FDR, false discovery rate; GFP, green fluorescent protein; GO, gene ontology; IRF8, interferon regulatory factor 8; TF, transcription factor; TES, transcription end sites; TSS, transcription start sites; YFP, yellow fluorescent protein; WT, wild type.

Introduction

Microglia are a major cell type responsible for innate immunity in the central nervous system[1, 2]. Besides their immune functions, microglia also integrate into shaping the brain's architecture through the regulation of synaptic pruning, neuronal survival, apoptotic cell clearance, and vascular formation during murine development[3, 4, 5, 6]. In the adult brain, microglia continuously surveil and respond to changes in their environment[7, 8, 9]. Accordingly, dysregulation and dysfunction of microglia have been strongly linked to the pathogenesis of diseases affecting the nervous system, such as Alzheimer's disease (AD)[10, 11].

Microglia arise from the yolk sac-derived erythro-myeloid progenitors. Then they migrate to the embryonic brain at E9.5 and expand in SPI1 (PU.1) and IRF8-dependent manner[12, 13]. In addition, transcription factors, such as Sall1, Smad3, and Mafb, contribute to microglial signatures in a stepwise manner, in combination with environmental factors[14, 15, 16].

IRF8 is a DNA binding transcription factor that binds to nucleotide motifs, ISRE, and composite ETS/ISRE, the latter by interacting with SPI1 (PU.1)[17]. IRF8 is expressed in developing myeloid cells in bone marrow, Ly6C+ macrophages, and CD8+ or plasmacytoid dendritic cells in the periphery[18, 19]. IRF8 is essential for innate resistance against a variety of pathogens[20]. The constitutive Irf8 knockout (IRF8KO) mice are susceptible to infection while having excess neutrophils[21]. In the brain, IRF8 is expressed only in microglia, except for a few macrophages, although it is expressed in astrocytes under a stressed condition [22, 23]. IRF8, along with PU.1, plays a role in early microglia progenitor development[13]. Besides, IRF8 is shown to control microglia's

responses to cytokines and pain injury. Importantly, microglia do not elicit full function in early postnatal stages and become active only after day 14 (P14)[24]. However, the mechanism of this postnatal microglia activation is largely unknown[14, 24]. Our study here shows that IRF8 plays a decisive role in the process. CUT&RUN assay revealed that IRF8 binds to microglia specific enhancer regions in a stepwise fashion during postnatal development reaching maximum after day 14. IRF8 binding was a prerequisite for establishing chromatin accessible sites and DNA methylation pattern, unique to microglia. Bulk and single-cell RNA-seq analyses of IRF8KO microglia, both constitutive and conditional, revealed that (a) IRF8 is essential for the expression of microglia identity and lysosomal CLEAR genes (b) IRF8 needs to be expressed continuously in adults to maintain microglia transcriptome profiles and (c) IRF8 forges transcription cascades to activate downstream factors, BATF3 and SALL1. Further, we document here that in the absence of IRF8, many Alzheimer's disease-associated genes and neurodegeneration genes (hereafter DAM/NDG) and interferon related genes are upregulated. Nevertheless, IRF8KO mice showed reduced AD pathology in a mouse model 5xFAD, indicating that IRF8 plays a complex role in AD pathogenesis. Together, IRF8 defines overall transcriptional programs and epigenetic landscape in postnatal microglia.

Materials and Methods

Mice

B6(Cg)-*Irf8*$^{tm1.2Hm}$/J (IRF8KO), B6.Cg-*Irf8*$^{tm2.1Hm}$/J (Irf8-GFP), B6.Cg-*Irf8*$^{tm1.1Hm}$/J (Irf8flox) and control wild-type C57BL/6J mice were maintained in the NICHD animal facility[25]. B6.129S(C)-*Batf3*^{tm1Kmm}/J (Batf3KO), B6.129P2(Cg)-*Cx3cr1*^{tm1Litt}/J (Cx3cr1^{GFP}), B6.129P2(C)-Cx3cr1$^{tm2.1(cre/ERT2)Jung}$/J (Cx3cr1cre^{ERT2}), and B6SJL-Tg$^{(APPSwFlLon, PSEN1*M146L*L286V)6799Vas}$/Mmjax (5xFAD) mice were purchased from Jackson Laboratory. Cx3cr1^{GFP} mice were crossed with IRF8KO mice. 5xFAD mice were backcrossed to C57BL/6J strain for at least seven generations and maintained hemizygote. To achieve microglia specific Irf8 depletion (IRF8cKO), we crossed Irf8flox mice with Cx3cr1cre^{ERT2} mice. Mice were given 300µg of Tamoxifen (Sigma) for five consecutive days at P12 or 2mg x 5 days at P28, and the Cre recombinase expression in microglia was confirmed using Rosa26$^{LSL-EYFP}$ reporter mice. Female mice were used in this study unless specified, and all animal experiments were performed according to the animal study (ASP#17-044 and 20-044) approved by the Animal Care and Use Committees of NICHD, NIH.

Microglia preparation

Mice were euthanized by asphyxiating with CO_2 and transcardially perfused with 10 mL of ice-cold PBS. The brain was harvested, chopped with a single-edge razor blade, and ground in ice-cold Hanks-balanced salt solution using a 7ml Dounce homogenizer, then filtered through a 70µm cell strainer. The cells were resuspended in 30% isotonic Percoll and spun down 800xg for 30 min. The floating myelin layer was removed, and the cell pellet was further

sorted by FACS AriaIIIμ (BD bioscience) after antibody staining. All reagents and instruments were kept in our 4°C cold room, and all procedures were performed at less than 4°C to avoid extra activation. For FACS cell sorting, the following antibodies were used: CD11b-Brilliant Violet 421 (CD11b-BV421, clone: M1/70, Biolegend), CD45-APC (clone: 30-F11, Invitrogen). The dead cells were gated out with staining of 7-Aminoactinomycin D (7-AAD; BD pharmingen, 1:100). We used CX3CR1-PE, Ly6C-PECy7, and F4/80-APCCy7 antibodies for the expression analysis and reanalyzed the data with FlowJo v.7.6.5 software.

Immunohistochemistry

Mice were anesthetized with 125mg/kg Ketamine and 10mg/kg Xylazine cocktail intraperitoneal injection. The brain was fixed by the transcardial perfusion of 4% PFA/PBS, followed by overnight incubation in an additional 4% PFA/PBS. After removing the fixative, the brain was cryoprotected with a 10-30% sucrose solution. The section was made with 30μm slices. Blocking was achieved with 4% BSA in PBS and 0.3% Triton X-100. Primary antibodies and working dilution are the following: Mouse anti-β-Amyloid antibody (1:500, clone: 6E10, Biolegend), Rabbit anti-NeuN monoclonal antibody (clone: A60, Millipore). The tissue was then incubated with the following fluorescence-labeled secondary antibodies; Alexa Fluor 488 Goat anti-mouse IgG (H+L) (Invitrogen), Alexa Fluor 633 Goat anti-mouse IgG (H+L) (Invitrogen), Alexa Fluor 633 Goat anti-rabbit IgG (H+L) (Invitrogen). The core-dense amyloid plaques were stained with 0.5% Thioflavin S (Sigma, T1892). The nuclei were stained with Hoechst33342 dye. The images were taken on a Zeiss AiryScan 880 confocal microscope and further analyzed

with ImageJ/Fiji[26]. Five fields per section were randomly chosen in the cortical layer V and analyzed for quantification. The average value of 5 fields was presented for each mouse.

Bulk RNA-Seq

Fifteen thousand cells were sorted out into TRIzol and stored at -70°C until library preparation. Total RNA was purified by phase extraction and processed to cDNA using SMART-Seq v4 Ultra Low Input RNA Kit (Takara bio USA, CA) with 15-cycle amplification. Adaptors were added to 1ng cDNA using the Nextera XT library prep kit (Illumina, CA). The libraries were sequenced on Nextseq500 with 55bp paired end read. For sequence data analysis, adapter sequences were trimmed with Trimmomatic v0.33 after the initial quality check, and then data was aligned with STAR v2.5.2b. Sequence duplicates were removed using Picard v2.17.11. Aligned tags were counted with Subread featureCounts, and differentially expressed genes were estimated using DESeq2[27] with FDR<0.01. Batch effects were normalized at this point if needed. Heatmaps were depicted with Morpheus (https://software.broadinstitute.org/morpheus).

Single-cell RNA-seq

Ten thousand cells were sorted out from 3 mice pools and immediately emulsified using a Chromium kit and 10x Genomics as the manufacturer's protocol. The libraries were made using SMART-Seq v4 Ultra Low Input RNA Kit followed by Nextera XT library prep kit and sequenced 26bp for R1 and 98bp for R2 on HiSeq2000. The fastq files were first processed with Cellranger v3.1

pipeline to make gene expression matrices, and the following analyses were performed with Scanpy[28]. Cells with less than 500 or greater than 6,000 detected genes were filtered from analysis. Genes detected in less than three cells were also removed from the data. The feature counts were normalized to 10,000 reads per cell and logarithmized. The data were regressed out based on the highly variable genes marked with *min_mean=0.0125, max_mean=3, min_disp=0.5* options, and the cells exhibiting high standard deviation (SD>15) were excluded from the analysis. UMAP dimension reduction and Leiden clustering were employed to visualize the data with *n_neighbors=80, n_pcs=15,* and *min_dist=1.0* parameters. Each cluster's connectivity was computed using PAGA[29] with the default parameters.

CUT&RUN

CUT&RUN methodology[30] was used with some modifications. Protein A-bound micrococcal nuclease (pA-MNase) was kindly provided by Steven Henikoff. For histone marks, live microglial cells were sorted out and washed once with HEPES-NaCl buffer (HEN buffer; 20mM HEPES pH7.5, 150mM NaCl, 500μM Spermidine, Protease inhibitors cocktail), then counted. Ten thousand cells were immobilized to activated BioMag®Plus Concanavalin A (ConA)-beads (Bangs Laboratories, IN) in the presence of Mn^{2+} and Ca^{2+} ions (0.1mM $MnCl_2$, 0.1mM $CaCl_2$) at room temperature. The beads complex was then permeabilized using 0.005% Digitonin (Merck)-containing HEN buffer (Digitonin buffer) and incubated for 12hrs at 4°C with 0.25μg of the following antibodies in the presence of 2mM EDTA: rabbit anti-H3K27ac (ab4729, Abcam), anti-rabbit H3K4me1 (ab8895, Abcam). After antibody binding, beads

were washed twice with Digitonin buffer and bound pA-MNase. After the washout of extra pA-MNase, beads-immunocomplex was activated in the presence of 3mM $CaCl_2$ at 0°C for 30 min. The digestion was stopped by adding 10x Stop buffer (1.70M NaCl, 100mM EDTA, 20mM EGTA, 0.25mg/ml Glycogen, 0,25μg/μl RNaseA) and 15pg MNase-digested yeast DNA spike-in control. The nuclei were digested at 56°C for five hours after adding 40μg of Proteinase K and SDS at the final concentration of 0.1%. The immunoprecipitated DNA was purified by phase separation and removed undigested DNA with a half volume of AMPure beads. For library preparation, purified DNA was processed using SMARTer® ThruPLEX DNA-Seq Kit (TAKARA Bio USA, CA) as the manufacturer's protocol except for PCR cycles to avoid amplifying long fragments (13 cycles, 10s for extension). For transcription factors, the antibody-labeled brain homogenate was fixed with 1% Paraformaldehyde-PBS at room temperature for 10min right before cell sorting. Sorted cells were immediately lysed with nuclear extraction buffer (20mM HEPES pH7.5, 10mM KCl, 500μM Spermidine, 0.1% Triton X-100, 20% Glycerol, Protease Inhibitor cocktail) and counted. Thirty-thousand nuclei were immobilized with ConA-beads in the same nuclear extraction buffer and blocked using Wash buffer (20mM HEPES pH7.5, 150mM NaCl, 500μM Spermidine, 0.1% Bovine Serum Albumin, 1% Triton X-100, 0.05% SDS, Protease inhibitor cocktail) containing 4mM EDTA. The nuclei were incubated with a rabbit anti-GFP antibody (0.5μg/reaction, ab290, Abcam) or a rabbit anti-mouse PU.1 (EPR22624-20, Abcam) for 12hrs at 4°C. After binding to pA-MNase in Wash buffer, DNA was cleaved by adding 3mM $CaCl_2$ for 30min at 0°C. MNase was inactivated by adding chelators mixture at the final concentration of 170mM

NaCl, 10mM EDTA, 2mM EGTA, 0.005% NP-40, 25μg/ml Glycogen, and 15pg yeast spike-in control. Total DNA was extracted, and de-crosslinked by incubating overnight at 65°C in the high-salt buffer containing 0.1% SDS, 0.5mg/ml Proteinase K and 0.1mg/mL RNase A. The undigested DNA was removed as above, and the small fragments for sequencing libraries were harvested. The final libraries were purified by AMPure beads and sequenced with a 30bp paired end read. For data analysis, the adapter and low-quality sequences were removed from raw fastq data using TrimGalore. Then the data was aligned on mm10 mouse genome using Bowtie2 with *--local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700 --no-overlap --no-dovetail* options. Peak calling was performed by MACS2 with *--broad --keep-dup all -p 0.01 --min-length 1000* option for a broad histone mark, *--keep-dup all -q 0.01 –max-gap 640* option for a narrow histone mark and IRF8[31]. For PU.1, we employed *-q 0.05* instead because it is more consistent with the previous dataset (GSM1533906)[16]. The peak co-occurrence, genomic position, and motif analysis were performed using HOMER[32] with *-d given* option. To determine highly concordant peaks across biological replicates, we employed “the Majority rule,” which picks up only peaks found within more than 50% of replicates[33]. The differential peaks were obtained using Diffbind-DESeq2[34] with FDR<0.05 cut-off. Then all bam files were merged and used for IGV visualization and depicting average plots and heatmaps with Deeptools[35]. H3K27ac^{high} large regions were identified as “super-enhancer regions” of the ROSE’s readouts with default parameters[36].

Bulk ATAC-Seq

ATAC-Seq libraries were constructed as described before[37]. Fifty thousand isolated microglia were lysed in lysis buffer (10mM Tris-HCl pH7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% NP-40). Following lysis, nuclei were tagmented using the Nextera DNA library prep kit (Illumina) for 30 minutes. Then DNA was purified and amplified for 12 cycles. The libraries were sequenced with a 50bp paired end read. After removing the adapter and low-quality sequences, the data were aligned to the mm10 genome with Bowtie2. Peaks were called on aligned sequences using MACS2 with *-p 0.01 --nomodel --shift -37 --extsize 73 -B --SPMR --keep-dup* all options. Peaks were then evaluated with less than 0.01 irreversible differential rates. Differentially accessible regions were identified using Diffbind-EdgeR with FDR<0.05[34]. Deeptools and HOMER were employed for further analyses, including motif enrichment analysis with the indicated options.

Whole-genome bisulfite sequence

Genomic DNA was extracted from 40,000 male microglial cells using Quick-DNA Microprep plus kit (Zymo Research) as the manufacturer's protocol. Then 50ng DNA was bisulfite-converted and processed to the library using Pico Methyl-Seq Library Prep Kit (Zymo Research), with 8-cycle PCR amplification. The libraries were sequenced with 75bp paired end read on HiSeq3000 in the NHLBI sequencing core (9.2x coverage on average). Based on the Bismark M-bias plot, the adapters, low-quality sequence, and an additional 10bp of 5'-end and 2bp of 3'-end were removed from the sequencing data using TrimGalore. The data were first aligned using Bismark[38] with *-q --score-min L,0,-0.4 --ignore-quals --no-mixed --no-discordant --dovetail --maxins* options, and then

unmapped sequences were aligned additionally with single-end mode. After deduplication and methylation calls, both files were merged into a single coverage file. Differential analyses were performed using R/DSS package[39] using FDR<0.05 for the differentially methylated loci (DMLs) and p<0.001 for the regions (DMRs).

Results

IRF8 binds to the microglia genome during postnatal development in a stepwise fashion

IRF8 is expressed in microglia throughout development from the embryonic stage to adulthood. IRF8, along with PU.1, is shown to play an essential role in embryonic microglia development[13]. However, little is known as to the role during postnatal development and in adults. To gain insight into this question, we sought to identify genome wide IRF8 binding sites in postnatally developing and adult microglia.

For this purpose, it was necessary to verify IRF8 expression in postnatal microglia. To this end, we examined Irf8-GFP knock-in mice that express IRF8 fused to GFP at the C-terminus[40]. FACS analyses of GFP-positive cells in the brain showed that IRF8 is specifically expressed in microglia at early (P9), late (P14), and adult (P60) stages. IRF8 levels were similar at all stages and higher than peritoneal macrophages (Figure S1A, S1B).

CUT&RUN experiments were carried out using Irf8-GFP mice to identify genome-wide IRF8 binding sites in microglia and peritoneal macrophages, as a reference. We identified 16,935 IRF8 peaks in adult microglia and 12,062 in peripheral peritoneal macrophages (Figure 1A). About half of the IRF8 peaks were in the intergenic regions with >5 kb distal from the genes, while about 20% were at the promoter regions in both cell types (Figure 1A). *De novo* motif analysis revealed enrichment of (a) PU.1/IRF composite motif, TTCC…G/CTTT, as shown in the row of "Sfpi1," and (b) classical ISRE (GAAAG) (Figure 1B). DEseq2-based statistics showed a differential IRF8 binding profile between microglia and peritoneal macrophages (Figure 1C). Further, GREAT GO

annotation indicated microglia specific IRF8 peaks (n=9,679) to be related to cell/leukocyte activation and differentiation (Figure 1D; upper panel). In contrast, macrophage specific IRF8 peaks (n=2,251) were related to immune responses and inflammatory responses against pathogens (Figure 1D; lower panel). In addition, IRF8 binding was found near microglia identity genes, such as Sall1 and Cx3cr1[11, 41], but not in peritoneal macrophages (Figure 1G).

IRF8 binding peaks on P9, P14, and adult microglia are shown in Figure 1E. The IRF8 binding peaks on P9 microglia were very sparse. They increased at P14 microglia, although still less than on adult microglia, indicating a stepwise increase in IRF8 binding (Figure 1E). Furthermore, cumulative distribution frequency graphs plotting the distance between two neighboring IRF8 binding sites (Figure 1F, right) showed that the distance in the adult is the shortest, while it is longer in P14 microglia and peritoneal macrophages, and the longest in P9 microglia, which indicates that IRF8 on microglia could form arrays as illustrated schematically (Figure 1F, left). IGV tracks corroborated the stepwise increase in IRF8 binding, observed over the enhancer region of Sall1 and Cx3cr1 genes in the postnatally developing microglia (Figure 1G, left). These regions were not occupied by IRF8 in peritoneal macrophages (Figure 1G, left). In contrast, in peritoneal macrophages, IRF8 is bound to the regulatory regions of Clec4n, Lyz1, and Lyz2, where these genes are active (Figure 1G, right)[11]. These results reveal that IRF8 binding to the microglial genome is developmentally controlled and reaches the maximum level when microglia gain full functional activity, presumably coinciding with the initiation of the adult microglia transcriptional program. It is interesting to note that despite a stepwise increase in DNA binding, the IRF8 levels were virtually the same during

postnatal stages (Figure S1B), suggesting a mechanism that directs IRF8's ability to gain access to the microglial genome.

IRF8 directs transcription cascades to activate microglia identity genes: bulk RNA-seq of wild type and IRF8KO microglia

To find out how IRF8 binding affects microglia transcriptional programs, we carried out bulk RNA-seq for microglia from wild-type (WT) and IRF8KO mice. To identify IRF8KO microglia among brain cells, we performed FACS analyses, which revealed two major $CD45^+$ populations in the IRF8KO brain residing even after a sufficient perfusion procedure, one $CD11b^{high}CD45^+$ corresponding to WT microglia, and another $CD11b^{low}CD45^+$ cell subset, termed P3 (Figure S2A). The $CD11b^{high}CD45^{low}$ subset from WT (i.e., microglia) and the $CD11b^{high}CD45^+$ subset from IRF8KO mice expressed CX3CR1, a typical microglial marker, although the amount on IRF8KO cells was lower. In contrast, the P3 subset from IRF8KO mice did not express this marker but instead expressed Ly6C, unraveling the presence of an extra cell population in the IRF8KO brain (Figure S2B). Thus, we isolated these cells from WT and IRF8KO brains and processed them for microglial transcriptome analyses. The volcano plot presenting a gene expression profile of WT and IRF8KO microglia exhibited a large number of genes that were up or downregulated in the absence of IRF8 (Figure 2A). A set of microglia identity genes, e.g., P2ry12, Siglech, and Ccr5, encoding surface proteins and a transcription factor, Sall1, were drastically reduced in IRF8KO microglia (Figure 2B left). In addition, we found that many lysosomal genes represented by CLEAR genes were also reduced in IRF8KO cells, e.g., Ctsb, Ctsc, Ctsh, Ctss, Ctsz, and Lamp1 (Figure 2B right)[42].

Consistent with the data, Tfeb, a transcription factor that drives CLEAR lysosomal gene expression and autophagy[42], was reduced in IRF8KO microglia (Figure 2C). On the other hand, a set of interferon related genes were aberrantly upregulated in IRF8KO cells, ranging from antiviral genes such as Mx1, Oas2, and Oas3 to transcription factors critical for interferon signaling, Stat2 and Irf7, suggesting dysregulated host defense (Figure 2B middle). In addition, many DAM/NDG genes were upregulated in IRF8KO microglia, including Apoe, a major AD risk factor (Figure 7A, B)[10, 11].

Given that IRF8 promotes the differentiation of monocytes by activating downstream transcription factors such as KLF4, we examined levels of transcription factors expressed in microglia[14, 19, 43]. Among a series of transcription factors shown to be expressed in postnatal and adult microglia, Batf3 and Sall1 were markedly downregulated in IRF8KO cells (Figure 2C, blue square)[14]. BATF3, a basic leucine zipper motif (bZIP) family of a transcription factor, is known to be critical for generating a dendritic cell subset[44]. Because the role of BATF3 in microglial transcriptome remained unelucidated, we next analyzed RNA-Seq in Batf3KO and WT microglia (Figure 2D; DEGs up in KO:68, down in KO:147 with FDR<0.01). As might be anticipated, genes downregulated in Batf3KO microglia, such as P2ry13 and Ctsc were also downregulated in IRF8KO cells, indicating that IRF8 and BATF3 control an overlapping set of target genes. Consistent with this, BATF3-dependent genes showed a significant correlation with IRF8-dependent genes (Figure 2E; $r=0.19$, $p=2.4\times10^{-9}$). These results indicate that IRF8 promotes the expression of these genes by enhancing Batf3 expression. Sall1 is a transcription factor of the Spalt family, expressed selectively in microglia. Buttgereit et al. reported that

SALL1 is required for the expression of microglia signature[41]. The correlation biplot also showed a significant correlation between SALL1 dependent genes and IRF8 dependent genes, including microglia identity and interferon related gene sets (Figure 2F; r=0.23, $p=5.6x10^{-39}$). These data reveal that IRF8 activates Sall1 expression, which in turn activates or represses genes that shape microglia signature. Thus, IRF8 creates transcription cascades and activates Sall1 and Batf3 expression, thereby inducing downstream target genes necessary for microglia identity and function (Figure 2G).

IRF8 configures microglia specific enhancers

The above results established that genome wide IRF8 binding represents a critical basis of the microglia transcriptional programs. To further characterize the chromatin environment in which IRF8 binds, and the relation to target gene expression, we first examined histone marks linked to active transcription, particularly those denoted for enhancers, H3K4me1 and H3K27ac[32, 36]. Of 16,935 IRF8 binding sites, about 73% and 66% of IRF8 colocalized with H3K4me1 and H3K27ac, respectively, and 64% of IRF8 binding sites carried both enhancer marks (Figure 3A). Moreover, H3K4me1 and H3K27ac marks surrounding IRF8 peaks were greatly decreased in IRF8KO cells compared to WT cells, indicating that IRF8 helps to set an active chromatin environment in microglia (Figure 3B). On the other hand, differential deposition analysis for H3K27ac and H3K4me1 marks revealed that both WT and IRF8KO microglia exhibited cell type specific histone modification patterns (Figure S3A). Since the H3K27ac is a surrogate marker for super-enhancers and super-enhancers neighbor many genes critical for defining cell type

specificity and function[36, 45], we asked whether IRF8 localizes to microglial super-enhancers. To examine super-enhancers in microglia and the possible contribution of IRF8, we employed the ROSE program[36] and analyzed how many regions were occupied by IRF8. As expected, many IRF8-dependent microglial identity genes were neighbored by some of 837 super-enhancers, including Sall1, Siglech, and Cst3 (Figure 3C). Indeed 735 of 837 (87.8%) super-enhancers were bound by IRF8 (Figure 3D). While most of the H3K27ac^{high} regions were populated by IRF8 binding in adult microglia, this is not the case in early postnatal P9 microglia and peritoneal macrophages (Figure 3E). In addition to super enhancers, IRF8 was abundantly enriched in smaller enhancers found in intergenic and introns overlapping with H3K4me1 (Figure S3C-G). IGV tracks for the Sall1 and Batf3 genes show that clusters of IRF8 binding coincide with those of H3K27ac and H3K4me1 (Figure 3F). In addition to these genes, we show other IGV examples of smaller enhancers that carry IRF8 clusters for transcription factors, Nfatc1 and Fos (Figure S3G). These results indicate that IRF8 binding guides the formation of microglia-specific enhancers during postnatal development.

IRF8 confers chromatin accessibility on postnatal microglia and sets DNA methylation patterns

To investigate genomic sites accessible for transcription factors in microglia and how IRF8 influences the process, we performed transposase-accessible chromatin (ATAC) sequencing[37]. We found more than 200,000 ATAC-seq peaks in both WT and IRF8KO microglia. The differential accessibility analysis revealed that 10,514 peaks were lost in IRF8KO microglia while

gaining 5,926 new peaks (Figure 4A). As anticipated, gained and lost peaks were mostly localized to intergenic regions and introns (Figure 4B). *De novo* motif analysis showed that lost peaks were enriched with the PU.1-IRF composite motifs (Figure 4C, left). Whereas, ATAC-seq peaks gained in IRF8KO cells were enriched with motifs for CEBP, AP-1, and RUNX binding sites (Figure 4C, right). We found that ATAC peaks lost in IRF8KO cells were enriched with WT specific H3K4me1 and H3K27ac marks (Figure S3A, 4D). On the other hand, the IRF8KO specific H3K4me1 marks were found in gained ATAC peaks, indicating IRF8KO microglia acquire a new enhancer landscape. Indeed, PU.1 binding sites identified by CUT&RUN showed that IRF8KO microglia gained extra PU.1 peaks that were absent in WT microglia (Figure S3B). We then sought relationships between ATAC-seq peaks and RNA-seq profiles. We found that ATAC-seq peaks lost in IRF8KO cells were enriched in genes downregulated in IRF8KO cells, such as microglia identity genes (Figure 4E, in red). Whereas ATAC-seq peaks gained in IRF8KO cells were enriched in genes upregulated in IRF8KO cells (Figure 4E, in purple). IGV tracks showed multiple ATAC peaks on and near the Sall1 and Batf3 genes in WT cells but few in IRF8KO cells (Figure 4I). Further, in WT cells, H3K27ac and H3K4me1 marks were abundantly present at these ATAC-seq sites but reduced in IRF8KO cells (Figure 4I). PU.1 binding sites also revealed a correlation with ATAC-seq peaks (Figure 4I). However, Apoe and the neighboring Apoc1 genes whose expression was negligible in WT cells, but greatly enhanced in IRF8KO cells, showed multiple ATAC-seq peaks in IRF8KO cells, but fewer in WT cells. H3K4me1 marks for Apoe and Apoc1 genes were few in WT cells but many in IRF8KO cells, consistent with the data in Figure 4E showing that this mark is associated

with both gain and loss of ATAC-seq peaks. Together, these data indicate that IRF8 sets chromatin accessible sites in microglia, thereby establishing a microglia specific transcriptional program.

DNA methylation influences the epigenetic landscape in various somatic cells[46]. DNA methylation has been extensively described on CpG-rich sequences near the proximal promoters, where methylated CpG correlates with transcriptional repression, while demethylated CpG with active transcription[47]. Thus, we asked whether IRF8 has a role in DNA methylation in microglia. To this end, we performed whole genome bisulfite sequencing. We identified over 20,000 methylated loci (DML) that were differentially regulated in WT and IRF8KO microglia (Figure 4F)[38, 39]. In addition, 1,808 DNA methylated regions (DMR) and 1,581 demethylated regions were differentially regulated in WT and IRF8KO cells, indicating that IRF8 takes part in DNA methylation in microglia (Figure 4F). Both methylated and demethylated regions were in intergenic regions and introns, away from the proximal promoter and TSS (Figure 4G). We found that IRF8 binding sites were located mostly in demethylated regions in WT cells (Figure 4H, top). Further, methylated regions were enriched with the H3K4me1 and H3K27ac marks present in IRF8KO cells, while demethylated regions were enriched with those marks in WT cells (Figure 4H). Correlation biplot analysis found that demethylated regions in WT cells correlated well with ATAC-seq peaks specific for WT cells, while methylated regions correlated with ATAC-seq peaks specific for IRF8KO cells (Figure S4A). These data indicate that demethylated regions were accessible to transcription factors, while methylated sites were not. Consistent with this idea, GO analysis revealed that demethylated regions were enriched with terms such as regulation of immune

responses and leukocyte activity, while methylated regions were associated with terms unrelated to microglia function (Figure S4B). In IGV tracks, individual methylated and demethylated sites are shown for Sall1, Batf3, and Apoe (Figure 4I, bottom two lanes). Sall1 and Batf3 were enriched with demethylated sites in WT cells, but not IRF8KO cells. Conversely, Apoe and Apoc exhibited many methylated sites in WT cells, whereas in IRF8KO cells, these genes displayed many demethylated sites. The above data provide the first evidence that IRF8 suppresses DNA methylation in gene specific manner in microglia. Together our results indicate that the regulation of DNA methylation is an integral mechanism for setting epigenetic landscape in microglia, which IRF8 dictates.

Continuous IRF8 expression is required for the transcriptome in postnatal and adult microglia

Although IRF8 has been shown to play a role in embryonic microglia development, it has been unclear whether IRF8 is required for postnatal microglia[13, 14]. For this purpose, we constructed Irf8$^{flox/flox}$Cx3cr1cre^{ERT2} mice to delete Irf8 in adult microglia conditionally. In combination with a Rosa26-loxP-stop-loxP-EYFP reporter, we could identify and isolate Cre-expressing microglia[48] (Figure 5A for experimental diagram). To our surprise, the efficiency of Irf8 deletion was very poor (28% ± 18%SD, Figure S5A) when Tamoxifen was administered on 4 weeks old mice five consecutive times, where EYFP-positive microglia were isolated. To mitigate this problem, we resorted to deleting Irf8 in younger mice, between P12 and P16, The Irf8 deletion efficacy improved when tested at three months old (62% ± 5%SD, Figure S5A). The basis of poor

deletion efficiency is currently unclear, but it may be related to the developmentally imposed alteration of the epigenome during cell proliferation[14, 24]. We performed bulk RNA-seq for IRF8cKO (Cx3cr1-Cre/wt) and WT (Cx3cr1-wt/wt) microglia, where all mice were given Tamoxifen (Figure 5A). While 546 genes were downregulated, 334 were upregulated in IRF8cKO microglia relative to WT cells (Figure 5B). Many DEGs found with IRF8cKO cells overlapped with those in constitutive IRF8KO cells (Figure S5B). Namely, genes denoting microglia identity, such as P2ry12, Siglech, and Aif1 (Iba1), were markedly downregulated in IRF8cKO cells (Figure 5C, 5D). Further, expression of IRF8's downstream transcription factors, Sall1 and Batf3, was noticeably lower in IRF8cKO cells as well (Figure 5C, 5D). Aberrant induction of DAM/NDG genes, as well as Interferon related genes, was also observed in IRF8cKO microglia, similar to IRF8KO cells (Figure 5B, S7B). Likewise, some CLEAR lysosome genes were downregulated in IRF8cKO cells (Figure 5C). Thus, continued IRF8 expression is required for maintaining the transcriptional programs in postnatal microglia.

Single-cell RNA-seq analysis of microglia reveals a possible stem cell-like population with the capacity for proliferation

Next, we performed single-cell RNA-Seq (scRNA-Seq) to assess transcriptome diversity in the microglia population. After filtration, 3,504 WT and 5,303 IRF8KO cells expressing a total of 15,376 genes were selected. UMAP dimension reduction in Figure 6A revealed a total of eleven clusters, three specific for WT and seven for IRF8KO cells[28], as verified by the purity of clusters shown in Figure 6B. Additionally, a cluster referred as "Common" was

found in WT and IRF8KO cells (Figures 6A and B). The SigEMD pipeline identified 6,149 DEGs between WT and IRF8KO cells, which covers 62% (3016/4848) of DEGs found in bulk RNA-seq analyses, lending credence to our transcriptome analyses (Figure S6B). The expression of microglia identity genes was higher in all WT Clusters than in all IRF8KO Clusters (Figure 6C, left). Population heterogeneity was evident within IRF8KO cells, in that Interferon related genes were exclusively found in Clusters 4 (Figure 6C, middle). In line with this, GO analysis for DEGs specifically expressed in IRF8KO Cluster 4 ranked interferon related terms as the top categories (Figure S6C). The overall CLEAR lysosomal gene expression was decreased in IRF8KO cells (Figure 6C, right). It should be mentioned here that IRF8KO Cluster 6 did not express microglia identity genes, nor interferon related genes, while expressing Ly6c2 transcripts that encode Ly6C protein (Figure 6C, S6A). Thus, this cluster likely represents the P3 population in the IRF8KO brain (Figure S2B).

The Common cluster, present in both WT and IRF8KO microglia, exhibited unusual features, in that most microglial genes were not expressed in this cluster (Figure 6D). This finding raised the possibility that the Common cluster represents cells in the early stage of development prior to becoming functional microglia. To delineate a relationship between the Common cluster and other clusters, we applied the PAGA program, which provided a trajectory network (Figure 6E)[29]. This network illustrated a tight connection with the Common cluster to all WT and IRF8KO clusters except IRF8KO Cluster 6. Significantly, cells in the Common cluster highly expressed cell cycle genes such as Cenpf, Cenpa, and Ccnb2, as well as Mki67, a marker for cell

proliferation (Figure 6F). To further substantiate the property of the Common cluster, we examined the expression of genes dependent on IRF8, BATF3, or SALL1 in WT and IRF8KO clusters relative to the Common cluster. The multiple correlation analysis indicated that genes expressed more in WT clusters were relevant to BATF3, SALL1, and IRF8, while those expressed more in IRF8KO clusters were not (Figure 6G). Together, these results would support the idea that the Common cluster represents a predecessor of microglia, possessing the capability to develop into functioning microglia.

Irf8 deletion ameliorates AD pathology in the 5xFAD mouse model

IRF8 has been implicated in AD pathogenesis in mice[49, 50]. IRF8 is also known to take part in neuroinflammation[20, 25]. To gain insight into the role of IRF8 in AD, we compared microglia transcriptome profiles from 9 months old WT and 5xFAD mice, extensively used as a mouse model of AD[51]. We selected 699 genes upregulated in 5xFAD microglia over WT cells, regarded as a 5xFAD associated gene set which contained most DAM/NDG genes described, including Apoe, Tyrobp, and Cst7 (Figure S7A)[10,11]. Many genes in this gene set were also upregulated in IRF8KO microglia without the 5xFAD background (Figure 7A, upper panel), including DAM/NDG genes. These upregulated genes are reported to be expressed early in 5xFAD in the Trem2 independent stage[50, 52]. On the other hand, the rest of the DAM/NDG genes were either unchanged or downregulated in IRF8KO microglia. Microglia from 5xFAD/IRF8KO mice showed essentially the same expression pattern as IRF8KO microglia (Figure 7A). Furthermore, DAM/NDG genes were upregulated in younger three months old IRF8cKO and IRF8KO microglia (Figure S7B, S7C). Thus, Irf8 deficiency

leads to the induction of DAM/NDG genes. Interestingly, the scRNA-seq analyses found that DAM/NDG gene expression was most prominent in Cluster 1 of IRF8KO cells, as evident for the 5xFAD associated gene set, Apoe, and Cst7 (Figure 7B). The IRF8 Cluster 1 exclusively expresses the Runx3 transcription factor, indicating the functional heterogeneity established by the aberrant transcription factor cascade in IRF8KO microglia (Figure 7B).

d'Errico et al. recently reported that Irf8 depletion leads to reduced microglia motility, resulting in reduced spreading of amyloidβ (Aβ) plaques[49]. To further examine the role of IRF8 in AD, histological analyses were performed with a year old 5xFAD/WT and 5xFAD/IRF8KO brains using Thioflavin S (ThioS) and 6E10 antibody[53]. The former detects polymerized Aβ sheets, while the latter stains both polymerized and unpolymerized Aβ. Confocal images revealed that the number of ThioS positive aggregates was similar between 5xFAD/WT and 5xFAD/IRF8KO brains. However, the amyloid load and plaque size were smaller in the 5xFAD/IRF8KO than in the 5xFAD/WT brain (Figure 7C, quantification in 7D). Furthermore, in the 5xFAD/IRF8KO brain, many of the Aβ signals detected by 6E10 (green) did not directly colocalize with Iba1 microglia (Figure S7D, red), whereas extensive colocalization of 6E10 and Iba1 was evident in WT. These results suggest that 5xFAD/IRF8KO microglia are defective in the recognition and perhaps internalization of Aβ in line with the report[49]. We next sought to assess the effect of Irf8 deletion on overall neuronal health in the 5xFAD brain. To this end, we stained cortical regions of a year-old 5xFAD/WT and 5xFAD/IRF8KO brains with the antibody for neuronal nuclei (NeuN). NeuN is a nuclear antigen expressed in normal neurons, and immunoreactivity for NeuN is lost after neuronal damage[54]. NeuN immunoreactivity in the 5xFAD/WT brain

slices was higher than those from 5xFAD/IRF8KO brains (Figure 7E, quantification in 7F). These results indicate that deletion of IRF8 lessens neuronal damage in the 5xFAD brain.

To further assess the complexity of IRF8 action in AD progression, we checked Axl expression. AXL kinase has been reported to play a critical role for microglia's ability to phagocytose Aβ[55]. We found Axl transcript level was lower in IRF8KO microglia than in WT microglia in the presence or absence of 5xFAD background (Figure 7A, S7F). These results suggest that IRF8 could affect AD pathology in a complex manner, along with reduced P2ry12 and Tyrobp, inhibiting Aβ spread on one hand and facilitating phagocytic clearance of Aβ on the other.

Discussion

We demonstrated that IRF8 bound to the microglia genome in a stepwise fashion during postnatal development, as binding was not extensively at P9, increased P14, reaching a maximum at adult. Microglia from P9 to P14 are still immature without full identity and function[24]. Our results indicate that full IRF8 binding causes a final step of maturation, which provides full identity and functionality to microglia. It is interesting to note that IRF8 expression was constant during this period and through adult, suggesting that the stepwise increase in IRF8 binding was not due to an increase in IRF8 but caused by changes in epigenome structure, such as chromatin modification. Such changes may be mediated by an external cue from other cells in the brain. In adult microglia, IRF8 binding was seen as close repeats and neighbored by histone modification marks, H3K27ac and H3K4me1, which denote enhancers.

Indeed, many IRF8 binding sites colocalized with large super-enhancers as well as smaller enhancers. IRF8 binding led to and was necessary for installing chromatin accessibility sites specific for microglia. This finding was verified by the observation that many of the ATAC-seq peaks neighbored microglia identity genes, which were lost in IRF8KO microglia. IRF8KO cells, in turn, gained alternate ATAC-seq sites as enhancers, some neighboring genes upregulated in IRF8KO microglia, including DAM/NDG genes, such as Apoe. We further showed that IRF8 sets DNA methylation patterns associated with the maintenance of microglia identity genes. Thus, our results reveal that IRF8 binding is the primary, initiating event that confers microglia specific epigenome landscape (Figure 8). We would like to emphasize that continuous IRF8 binding is required for maintaining microglia identity and function, since conditional Irf8 deletion after P14 abolished microglia identity and CLEAR lysosomal gene expression similar to straight Irf8 deletion. Moreover, IRF8 was required to activate two transcription factors, Sall1 and Batf3, that induce many genes necessary for microglia identity and function. Both genes were near and under large super-enhancers enriched with IRF8 binding and were associated with ATAC-seq peaks, indicating open chromatin. These epigenome features were completely lost in IRF8KO cells, highlighting that IRF8, acting upstream of these factors, creates two transcription cascades to execute transcriptional programs for adult microglia. These characteristics of IRR8 are somewhat reminiscent of the “pioneer transcription factor” that enables other transcription factors, histone in modifiers, to gain access to the proper DNA/chromatin sites, thereby installing functional regulatory domains, either active or repressive[56]. IRF8 has been shown to regulate the development of innate immunity in the periphery[18,

[19]. In these cells, IRF8 induces downstream transcription factors, such as KLF4, ID2, and BATF3, to achieve full differentiation. We recently extended this line of work and showed that IRF8 promotes H3K27 acetylation in the regulatory regions, leading to changes in chromatin architecture in myeloid progenitors[57].

The scRNA-seq data on the Common cluster are likely relevant to postnatal microglia development and lifespan. This cluster, found in WT and IRF8KO cells, expressed multiple cell cycle and replication genes, not detected in other clusters. Cells in this cluster were Ki67 positive, thus proliferating. PAGA trajectory suggested that this cluster is connected to most of the other clusters, pointing to the possibility that it presents a *predecessor* population capable of self-renewal and differentiation into mature microglia. With a stem-cell-like property, this population may be the source of postnatal and adult microglia, which replenishes microglia continually for life.

Irf8 deletion led to a drastic reduction of microglia identity genes but unleashed the expression of many DAM/NDG genes, including Apoe and Cst7. Expression of these genes is thought to reflect microglia's responses to AD and other neurodegenerative diseases[10, 11]. Interferon-related genes are another group of genes upregulated in IRF8KO cells. This is unexpected, since this is not observed in IRF8KO monocyte/macrophages, nor in dendritic cells[18, 20]. Rather, Irf8 deletion impairs such interferon-related gene induction and weakens anti-microbial resistance in those cells. Nevertheless, interferon-related gene expression in microglia has been noted in AD and other pathological conditions[11]. Despite extensive DAM/NDG genes and Interferon-related gene expression, AD pathology in 5xFAD brains from IRF8KO mice was less severe than that of WT mice, as evidenced by reduced large Aβ plaques in

5xFAD/IRF8KO brains. It has been reported that microglia constantly survey the brain space, recognize Aβ and distribute them to other regions in the brain[49]. IRF8KO microglia, devoid of the ability to scan the brain and unable to recognize external stimuli, would not efficiently interact with Aβ. This possibility was supported by the observation that IRF8KO microglia did not colocalize well with Aβ. d'Errico et al. further reported that Aβ plaques did not spread in the IRF8KO brain as much as in WT counterparts[49]. Pertinent to this report, Spangenberg et al. showed that prolonged depletion of microglia by CSF1R inhibitors led to a dramatic reduction in plaque formation and neuronal loss[58]. These reports are consistent with the view that normal microglia, by promoting the development and seeding of plaques, worsen AD pathology during some phases of AD progression.

Furthermore, the 5xFAD/IRF8KO brain displayed more NeuN immunoreactivity than 5xFAD/WT brain, indicating that neuronal damage was less extensive without IRF8. Reduced neuronal damage seen in 5xFAD/IRF8KO brain may be contributed by the compositional or spacial pattern of Aβ aggregates, which largely remains elusive[49, 59].

In light of reduced AD pathology, extensive DAM/NDG gene expression in 5xFAD/IRF8KO microglia may appear contradictory. However, the DAM/NDG on the signature does not directly indicate AD phenotypes. Therefore, it is likely that individual DAM/NDG genes affect AD pathology differently. Furthermore, it has been recognized that the DAM/NDG signature does not imply a causality of AD.

Lastly, it should be mentioned that the role of IRF8 is complex, and Irf8 deletion would not simply reduce AD pathology, but could cause detrimental

effects, considering that Irf8 deletion leads to broad downregulation of Axl, as well as CLEAR/lysosomal genes, likely involved in Aβ internalization, and lysosomal degradation[55].

In summary, this study demonstrates the essential requirement of IRF8 for the adult development of microglia and its complex role in modulating AD progression.

Data availability: All high-throughput sequence datasets generated in this paper are available in GSE231406. Some gene sets were obtained from publications as denoted individually.

Acknowledgments: We thank Steven Coon, NICHD DIR MGC, for his guidance for scRNAseq, Yuesheng Li in NHLBI sequencing core for whole-genome bisulfite sequencing, and Fayaz Seifuddin for data analysis. We also thank Avinash Bhandola for providing Rosa26-EYFP mice, Lisa William-Simons for cell sorting, and the NICHD animal facility staff for caring for our mice and for other technical assistance. We gratefully acknowledge colleagues in the NIH and elsewhere for valuable advice and discussions. This work was supported by the NICHD Intramural programs ZIA HD008015-13.

Disclosure: There is no conflict of interest.

Author Contributions: K.S and K.O conceived the framework of this work; K.S generated IRF8cKO microglia, devised whole genome analyses, and executed experiments along with R.P; E.L performed immunohistology; D.K. provided advice on experiments; K.S and K.O wrote the manuscript.

Figure Legends

Figure 1 Microglia specific IRF8 occupancy increases postnatally

A. Genome wide distribution of IRF8 peaks in adult microglia (top) and peritoneal macrophages (bottom). The genomic position of each peak was determined with HOMER.
B. HOMER *de novo* motif analysis for IRF8 binding sites in microglia. Motifs with a match score >0.85 are shown.
C. An MA plot visualizing the difference between microglia IRF8 peak intensity and that of peritoneal macrophages (n=3). The differential peaks with FDR<0.05 are indicated in pink.
D. Top 10 Gene ontology terms for microglia specific IRF8 peaks (top) and macrophage specific IRF8 peaks (bottom), analyzed with GREAT. Sequences covering 500kb distal from the TSS to 1kb from TES were set as an associating region. The terms of the GO biological process and the corresponding binomial FDR values were shown.
E. The number of IRF8 peaks identified in P9, P14, and adult microglia.
F. Schematic illustration of the distance between two neighboring IRF8 peaks; the shortening peak intervals represent an array-like appearance of peaks (left). The CDF plot shows that the frequency of short interval peaks increases along with the postnatal microglia development and more than peritoneal macrophages. (right; $p=2.2x10^{-16}$, Kolmogorov-Smirnov two-sided Test).
G. IGV examples of IRF8 peaks in developing and adult microglia and peritoneal macrophages at microglia specific genes (Sall1 and Cx3cr1; left) or macrophage specific genes (Clec4n, Lyz1, and Lyz2; right).

Figure 2 Identification of IRF8-driven transcription factor cascades in shaping microglia-specific transcriptional programs

A. Volcano plot for RNA-seq comparing WT and IRF8KO microglia. Genes upregulated in IRF8KO microglia are indicated in pink (n=2,356), and genes downregulated in IRF8KO cells are in blue (n=2,682, FDR<0.01). Representative genes are shown with a label.
B. Heatmaps for indicated gene sets. "Microglia identity" and "Interferon-related" gene sets were referred to as "cluster 6: Microglia" and "Interferon-Related" in the previous publication[11] The CLEAR lysosome gene set was obtained from elsewhere[42].
C. Transcription factors downregulated in IRF8KO microglia (FDR<0.01 cut-off) are shown. A list of transcription factor genes was obtained from the previous publication[14]. Genes that show more than four log-scale reductions in IRF8KO cells are boxed, and Sall1 and Batf3 in the blue box were further investigated.
D. Volcano plot for RNA-Seq data comparing WT and Batf3KO microglia. The genes upregulated in Batf3KO cells are shown in pink (n=172) and downregulated genes in blue (n=274, FDR<0.01). The representative genes are indicated in the plot.
E. Correlation biplot showing that Batf3 dependent genes correlate with those dependent on IRF8. The DEGs from Batf3KO microglia RNA-Seq were analyzed with the Kendall rank correlation test. Microglia identity genes shared between Batf3KO DEGs and IRF8KO DEGs are labeled red.

F. Correlation biplot showing that Sall1 dependent and IRF8 dependent genes overlap. All up and down DEGs from Sall1KO microglia RNA-Seq obtained from the previous publication[41] were analyzed with the Kendall rank correlation test. The microglia identity genes shared between Sall1KO DEGs and IRF8KO DEGs are labeled purple, and those of interferon-related genes are red.

G. A model illustrating IRF8 dependent transcription cascade. IRF8 induces expression of downstream TFs, BATF3 and SALL1. These TFs, along with IRF8, then activate indicated target genes. This dual cascade would cover all three gene sets, microglia identity, interferon related, and CLEAR lysosomal genes.

Figure 3 Adult IRF8 is an active enhancer that establishes an enhancer landscape in adult microglia

A. A Venn diagram showing the overlap among IRF8 peaks, H3K4me1, and H3K27ac marks.

B. Heatmaps showing H3K27ac (left) and H3K4me1 (right) signal intensity centered on IRF8 peaks ± 3.0kb in adult WT and IRF8KO microglia. Shown above heatmaps are average peak intensities.

C. H3K27ac ranked super enhancers in adult WT microglia. IRF8 dependent genes neighboring super enhancers are indicated in red.

D. Venn Diagram depicting the overlap of H3K27ac super enhancers and IRF8 peaks in adult WT microglia.

E. IRF8 peak counts of adult microglia, P9 microglia, and peritoneal macrophages in the H3K27ac^{high} regions.

F. Representative IGVs visualizing the distribution of IRF8 peaks, H3K4me1, and H3K27ac marks near/at Sall1 (top) and Batf3 genes (bottom). An H3K27ac^{high} region is marked by a red bar. IRF8 bound densely to the Batf3 regulatory region (orange bar), distinct from super enhancers. See also Fig.S3C-G for further description.

Figure 4 IRF8 confers chromatin accessibility on postnatal microglia and sets DNA methylation patterns

A. MA plot showing ATAC-seq signals comparing WT and IRF8KO microglia (n=2). The differentially accessible regions (FDR<0.05) are colored in pink. The number of gained or lost accessible regions in IRF8KO microglia was shown at the corresponding corner of the area.
B. The genome-wide distribution of gained and lost ATAC-seq peaks as determined by HOMER.
C. HOMER *de novo* motif analysis for gained and lost ATAC-seq peaks. The top 4 motifs are shown.
D. The number of IRF8, H3K27ac, and H3K4me1 peaks (WT or IRF8KO specific) overlapped with lost or gained ATAC-seq peaks. See Fig.S3A for the WT or IRF8KO specific peak identification.
E. Volcano plot showing a correlation between ATAC-seq peaks and RNA-seq data. Microglia identity genes (representative "down" genes) and Interferon related genes (representative "up" genes) are labeled in red and purple, respectively. The nearest genes for gained or lost ATAC-seq peaks correlate with genes up or down-regulated in IRF8KO microglia.

F. The number of differentially methylated single nucleotide loci (DML; FDR<0.05) and the differential regions (DMR; p<0.001). Numbers in the graph are differentially methylated and demethylated loci/regions in WT cells.

G. Genome-wide distribution of the differentially demethylated (top) and methylated (bottom) regions identified with HOMER.

H. Numbers of IRF8, H3K27ac, and H3K4me1 (WT or IRF8KO specific) peaks in differentially methylated or demethylated regions.

I. Representative IGVs visualizing IRF8. H3K4me1, H3K27ac, ATAC-seq, PU.1 peaks and methylated DNA (CpG) status on Sall1 (top), Batf3 (middle), and Apoe (bottom) in WT and IRF8KO cells. Methylated and demethylated regions are indicated by black bars.

Figure 5 Postnatal deletion of Irf8 causes gene expression changes analogous to constitutive IRF8KO microglia

A. The strategy for conditional Irf8 deletion in postnatal microglia. P12 or four weeks old Irf8$^{flox/flox}$Cx3cr1Cre$^{ERT2/+}$Rosa26^{YFP} mice were injected with Tamoxifen 5 times, followed by eight weeks of rest. Microglia were isolated and used for RNA-seq analysis.

B. Volcano plot for transcriptome profiles of WT and IRF8cKO microglia. Up and down regulated genes in IRF8cKO microglia are shown in pink (n=334) and blue (n=546, FDR<0.01), respectively. Representative up and down regulated genes are indicated in the plot.

C. Heatmaps comparing expression levels of the indicated gene sets between WT and IRF8cKO microglia.

D. Expression of Aif1 (Iba1) and Batf3 genes in WT and IRF8cKO microglia. Each triangle represents logarithmized counts+1 from RNA-seq data of different biological replicates.

Figure 6 scRNA-seq analysis reveals transcriptome heterogeneity and the presence of a progenitor like population

A. UMAP derived from WT and IRF8KO microglia aggregates. Leiden clusters were colored and labeled as indicated. The data were obtained from the pool of three individual WT and IRF8KO microglia preparations.
B. Percentages of WT and IRF8KO genotypes in indicated clusters.
C. UMAP presentation of indicated gene sets (compare with heatmaps in Figure 2B).
D. Gene expression heatmap showing the expression of microglia identity gene set in each WT and IRF8KO cluster. Note that expression of all identity genes is absent in IRF8KO Cluster 6.
E. PAGA network depicted with the default parameters. Each node indicates the corresponding Leiden cluster.
F. Rank plot aligning the top 20 genes by scoring expression level. The genes in the Common cluster were compared to those of all the other clusters.
G. Heatmap displaying the correlation coefficients between Common cluster and other clusters for IRF8, BATF3, or SALL1 dependent gene sets.

Figure 7 IRF8KO microglia lessens AD pathology in the 5xFAD mouse model, despite spontaneous DAM/NDG gene induction

A. Heatmap showing the 5xFAD associated gene set expression in WT, 5xFAD/WT, IRF8KO, and 5xFAD/IRF8KO microglia. Data are from RNA seq analysis of microglia from nine months old mice. Data were clustered by K-mean methodology (k=2), and representative genes were labeled on the right.
B. UMAP presentation of 5xFAD associated gene set and its representatives, Apoe and Cst7, from the scRNA-seq analysis. Runx3 is a transcription factor exclusively expressed in IRF8 Cluster1, indicating an association of RUNX3 with the transcriptome of this cluster.
C. Representative histology images showing core-dense plaques in the cortex region of a year old 5xFAD/WT or 5xFAD/IRF8KO brain. Brain sections were stained with Thioflavin S (green) and 6E10 antibody (red). Scale bar; 200 μm.
D. Quantification of the number of ThioS+ aggregates (left), amount of Aβ deposition (middle), and size of Aβ plaques (right). Values represent the average of five fields from each brain. Data: Mean +/- SEM; *: p-value <0.05.
E. Representative histology images of NeuN staining in the cortex area of one year old 5xFAD/WT or 5xFAD/IRF8KO brain. Scale bar; 200 μm. On the right are enlarged images from the bracketed regions on the left.
F. Quantification of NeuN positive cells. Values are the average of five fields from three brains. *: p<0.05.

Figure 8 Model for the role of IRF8 in postnatal microglia development

At the early postnatal stage (P9 microglia), IRF8 binding to the microglia genome is sparse (Top). After P14 through adulthood, IRF8 binding becomes denser, forming arrays along with histone modifications (Middle). Full IRF8 binding leads to the formation of microglia specific enhancers (large and small), opens relevant regions of chromatin and stably demethylates enhancer regions of DNA (bottom), which, taken together, establishes transcriptional programs in microglia.

References

1. Rivest, S. Regulation of innate immune responses in the brain. *Nat Rev Immunol* 9, 429-439 (2009).

2. Prinz, M., Jung, S. & Priller, J. Microglia Biology: One Century of Evolving Concepts. *Cell* 179, 292-311 (2019).

3. Ueno, M. *et al.* Layer V cortical neurons require microglial support for survival during postnatal development. *Nat Neurosci* 16, 543-551 (2013).

4. Fourgeaud, L. *et al.* TAM receptors regulate multiple features of microglial physiology. *Nature* 532, 240-244 (2016).

5. Fantin, A. *et al.* Tissue macrophages act as cellular chaperones for vascular anastomosis downstream of VEGF-mediated endothelial tip cell induction. *Blood* 116, 829-840 (2010).

6. Schafer, D.P. *et al.* Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron* 74, 691-705 (2012).

7. Wang, X. *et al.* Requirement for Microglia for the Maintenance of Synaptic Function and Integrity in the Mature Retina. *J Neurosci* 36, 2827-2842 (2016).

8. Reshef, R. *et al.* The role of microglia and their CX3CR1 signaling in adult neurogenesis in the olfactory bulb. *Elife* 6 (2017).

9. Nimmerjahn A, K.F., Helmchen F. . Resting microglial cells are highly dynamic surveillants of brain parenchyma in vivo. *Science* 308(5726), 1314-1318 (2005).

10. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* 169, 1276-1290 e1217 (2017).

11. Friedman, B.A. *et al.* Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models. *Cell Rep* 22, 832-847 (2018).

12. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* 330, 841-845 (2010).

13. Kierdorf K, E.D., Goldmann T, Sander V, Schulz C, Perdiguero EG, Wieghofer P, Heinrich A, Riemke P, Hölscher C, Müller DN, Luckow B, Brocker T, Debowski K, Fritz G, Opdenakker G, Diefenbach A, Biber K, Heikenwalder M, Geissmann F, Rosenbauer F, Prinz M. . Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat Neurosci.* 16(3), 273-280 (2013).

14. Matcovitch-Natan, O. *et al.* Microglia development follows a stepwise program to regulate brain homeostasis. *Science* 353, aad8670 (2016).

15. Lavin, Y. *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159, 1312-1326 (2014).

16. Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* 159, 1327-1340 (2014).

17. Mancino, A. *et al.* A dual cis-regulatory code links IRF8 to constitutive and inducible gene expression in macrophages. *Genes Dev* 29, 394-408 (2015).

18. Das, A., Chauhan, K.S., Kumar, H. & Tailor, P. Mutation in Irf8 Gene (Irf8(R294C)) Impairs Type I IFN-Mediated Antiviral Immune Response by Murine pDCs. *Front Immunol* 12, 758190 (2021).

19. Kurotaki D, O.N., Nishiyama A, Yamamoto M, Ban T, Sato H, Nakabayashi J, Umehara M, Miyake N, Matsumoto N, Nakazawa M, Ozato K, Tamura T. . Essential role of the IRF8-KLF4 transcription factor cascade in murine monocyte differentiation. *Blood.* 7;121(10), 1839-1849. (2013).

20. Gupta, M. *et al.* IRF8 directs stress-induced autophagy in macrophages and promotes clearance of Listeria monocytogenes. *Nat Commun* 6, 6379 (2015).

21. Holtschke T, L.J., Kanno Y, Fehr T, Giese N, Rosenbauer F, Lou J, Knobeloch KP, Gabriele L, Waring JF, Bachmann MF, Zinkernagel RM, Morse HC 3rd, Ozato K, Horak I. . Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. *Cell.* 18;87(2), 307-317. (1996).

22. Masuda, T. *et al.* Specification of CNS macrophage subsets occurs postnatally in defined niches. *Nature* 604, 740-748 (2022).

23. Konishi, H. *et al.* Astrocytic phagocytosis is a compensatory mechanism for microglial dysfunction. *EMBO J* 39, e104464 (2020).

24. Bennett, M.L. *et al.* New tools for studying microglia in the mouse and human CNS. *Proc Natl Acad Sci U S A* 113, E1738-1746 (2016).

25. Yoshida, Y. *et al.* The transcription factor IRF8 activates integrin-mediated TGF-beta signaling and promotes neuroinflammation. *Immunity* 40, 187-198 (2014).

26. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676-682 (2012).

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

28. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19, 15 (2018).

29. Wolf, F.A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* 20, 59 (2019).

30. Skene, P.J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 6 (2017).

31. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137 (2008).

32. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38, 576-589 (2010).

33. Yang, Y. *et al.* Leveraging biological replicates to improve analysis in ChIP-seq experiments. *Comput Struct Biotechnol J* 9, e201401002 (2014).

34. Ross-Innes, C.S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389-393 (2012).

35. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42, W187-191 (2014).

36. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307-319 (2013).

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

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

39. Feng, H., Conneely, K.N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res* 42, e69 (2014).

40. Wang, H. *et al.* A reporter mouse reveals lineage-specific and heterogeneous expression of IRF8 during lymphoid and myeloid cell differentiation. *J Immunol* 193, 1766-1777 (2014).

41. Buttgereit, A. *et al.* Sall1 is a transcriptional regulator defining microglia identity and function. *Nat Immunol* 17, 1397-1406 (2016).

42. Palmieri, M. *et al.* Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways. *Hum Mol Genet* 20, 3852-3866 (2011).

43. Tamura, T. *et al.* IFN regulatory factor-4 and -8 govern dendritic cell subset development and their functional diversity. *J Immunol* 174, 2573-2581 (2005).

44. Hildner K, E.B., Purtha WE, Diamond M, Matsushita H, Kohyama M, Calderon B, Schraml BU, Unanue ER, Diamond MS, Schreiber RD, Murphy TL, Murphy KM. . Batf3 deficiency reveals a critical role for CD8alpha+ dendritic cells in cytotoxic T cell immunity. *Science* 14;322(5904):1097-100. (2008).

45. Dey, A. *et al.* BRD4 directs hematopoietic stem cell development and modulates macrophage inflammatory responses. *EMBO J* 38 (2019).

46. Mattei, A.L., Bailly, N. & Meissner, A. DNA methylation: a historical perspective. *Trends Genet* 38, 676-707 (2022).

47. Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* 25, 1010-1022 (2011).

48. Yona, S. *et al.* Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* 38, 79-91 (2013).

49. d'Errico, P. *et al.* Microglia contribute to the propagation of Abeta into unaffected brain tissue. *Nat Neurosci* 25, 20-25 (2022).

50. Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med* 26, 131-142 (2020).

51. Oakley, H. *et al.* Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J Neurosci* 26, 10129-10140 (2006).

52. Ulland, T.K. & Colonna, M. TREM2 - a key player in microglial biology and Alzheimer disease. *Nat Rev Neurol* 14, 667-675 (2018).

53. Rajamohamedsait, H.B. & Sigurdsson, E.M. Histological staining of amyloid and pre-amyloid peptides and proteins in mouse tissue. *Methods Mol Biol* 849, 411-424 (2012).

54. Gusel'nikova VV, K.D. NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker. *Acta Naturae.* 7(2):42-7. (2015).

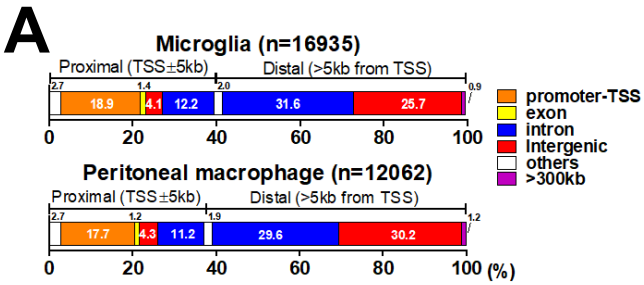
55. Huang, Y. *et al.* Microglia use TAM receptors to detect and engulf amyloid beta plaques. *Nat Immunol* 22, 586-594 (2021).

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

57. Kurotaki, D. *et al.* Chromatin structure undergoes global and local reorganization during murine dendritic cell development and activation. *Proc Natl Acad Sci U S A* 119, e2207009119 (2022).

58. Spangenberg, E. *et al.* Sustained microglial depletion with CSF1R inhibitor impairs parenchymal plaque development in an Alzheimer's disease model. *Nat Commun* 10, 3758 (2019).

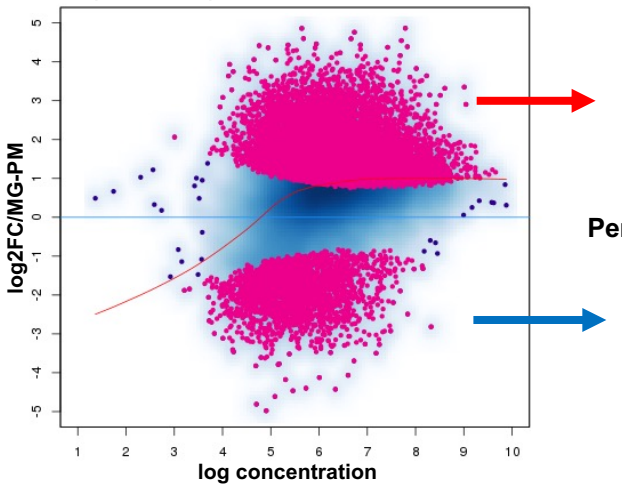
59. Sanchez-Varo, R. *et al.* Plaque-Associated Oligomeric Amyloid-Beta Drives Early Synaptotoxicity in APP/PS1 Mice Hippocampus: Ultrastructural Pathology Analysis. *Front Neurosci* 15, 752594 (2021).



B

Rank	Motif	p-value	%T/B	Name
1		1e-1765	48.06 /17.76	Sfpi1 (0.969)
2		1e-378	29.60 /16.67	IRF8 (0.881)

C



D

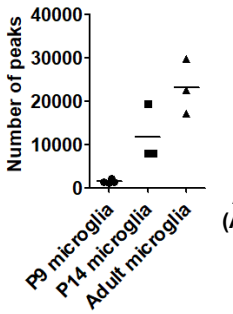
Microglia specific IRF8 (n=9679)

# Term Name	FDR
positive regulation of immune system process	1.19210e-111
cell activation	3.09162e-99
leukocyte activation	4.62254e-84
regulation of immune response	1.59865e-77
regulation of cell activation	5.56143e-77
regulation of lymphocyte activation	2.94470e-69
regulation of leukocyte activation	1.50084e-68
regulation of hemopoiesis	4.11654e-67
regulation of leukocyte differentiation	3.70227e-66
leukocyte differentiation	5.52185e-61

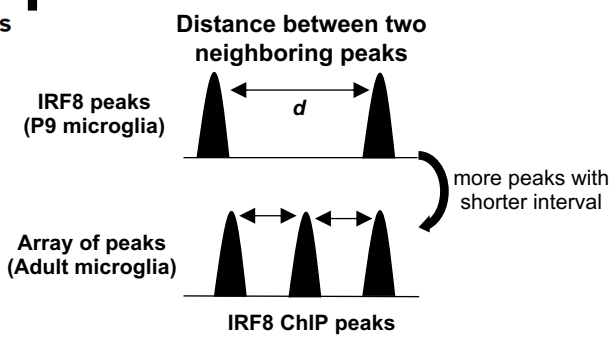
Peritoneal macrophage specific IRF8 (n=2251)

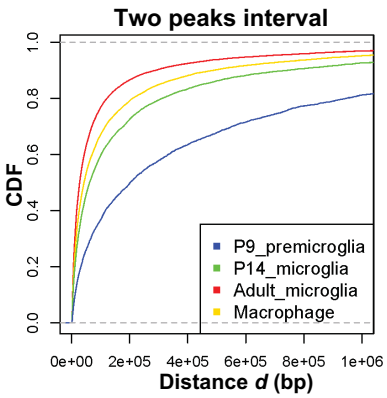
# Term Name	FDR
regulation of immune system process	3.99626e-40
regulation of MAPK cascade	8.16172e-29
inflammatory response	1.53254e-28
response to molecule of bacterial origin	4.08295e-28
defense response	4.10928e-28
positive regulation of immune system process	9.64933e-27
response to other organism	3.52883e-26
response to bacterium	4.63784e-26
response to external biotic stimulus	4.65288e-26
response to lipopolysaccharide	1.25573e-25

E



F





G

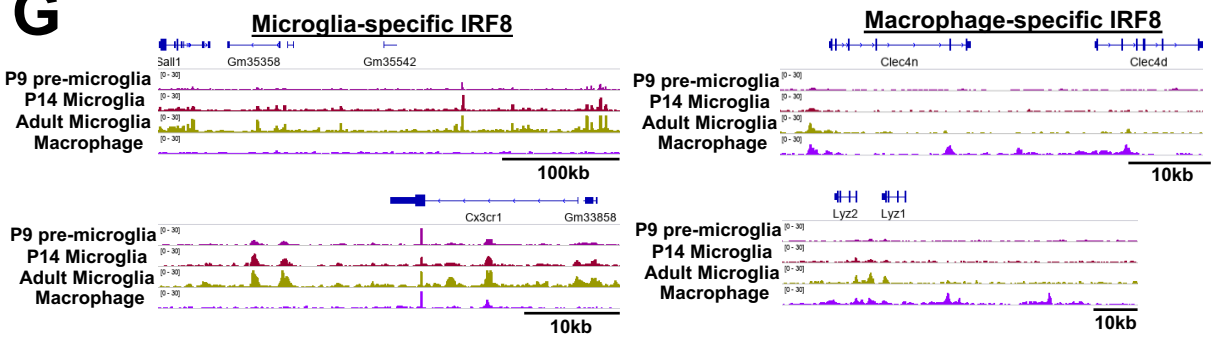


Fig.1 Saeki et al.

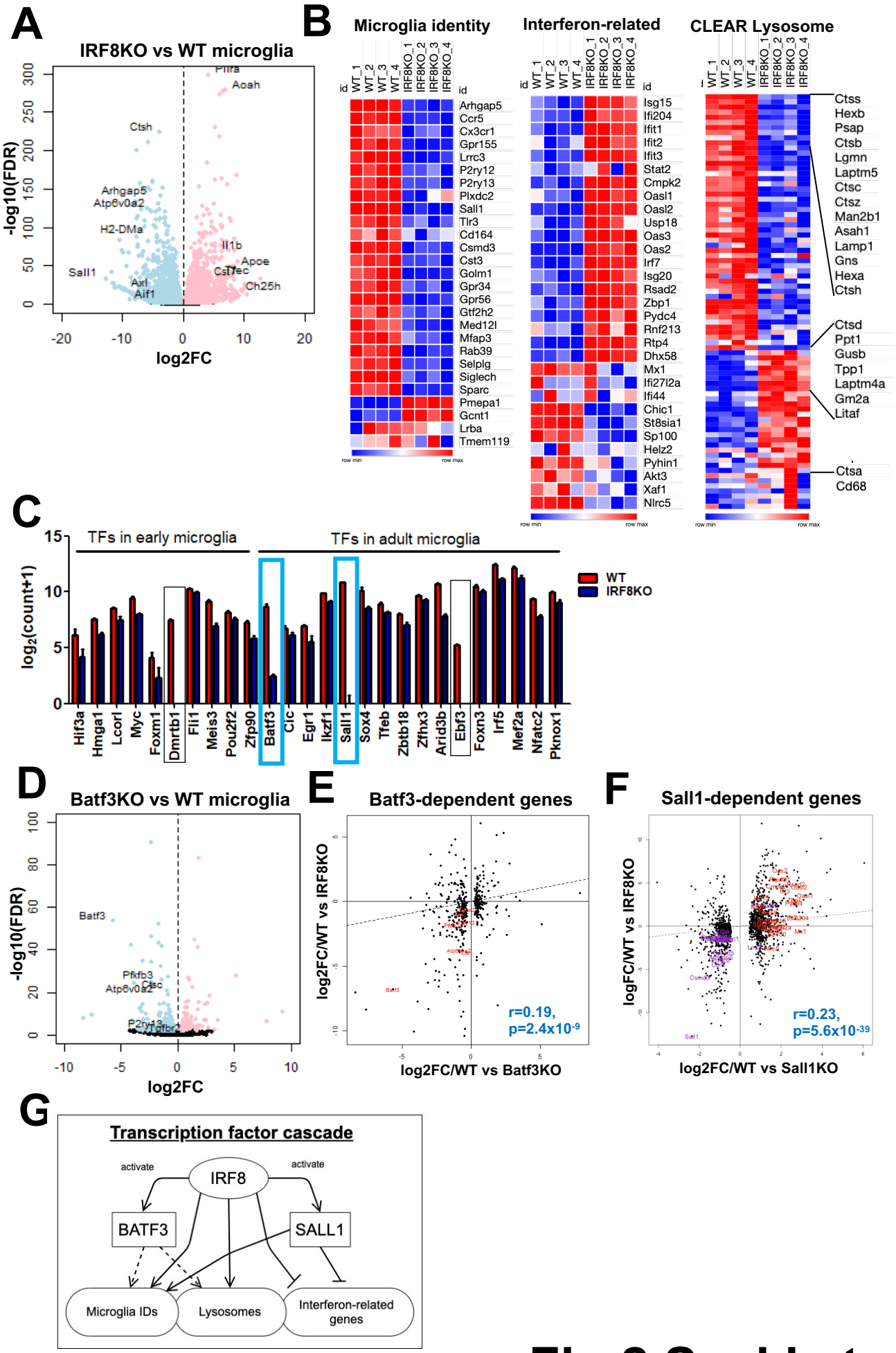


Fig.2 Saeki et al.

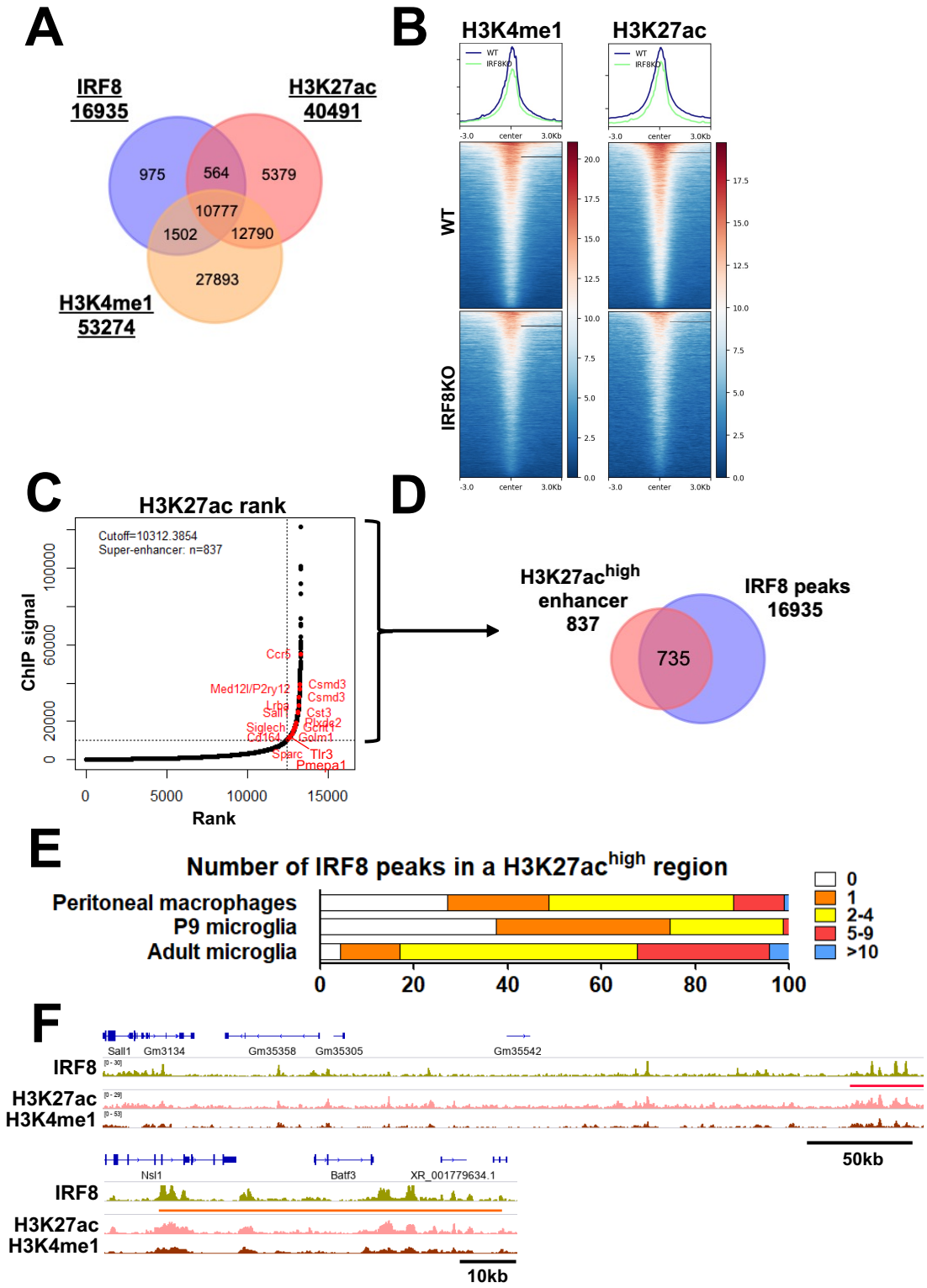
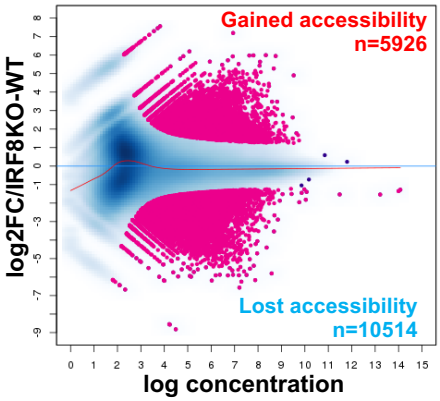


Fig.3 Saeki et al.

A

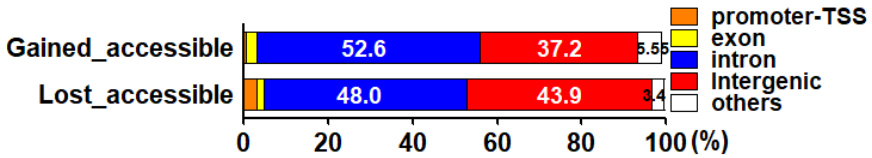


D

IRF8 peaks, H3K4me1 and H3K27ac marks in lost and gained ATAC-seq peaks

	Lost accessible regions (n=10514)	Gained accessible regions (n=5926)
IRF8 peaks (n=16935)	4103	158
WT-specific H3K4me1 (n=3882)	1489	8
KO-specific H3K4me1 (n=4952)	15	984
WT-specific H3K27ac (n=2819)	1210	7
KO-specific H3K27ac (n=1275)	1	163

B



C

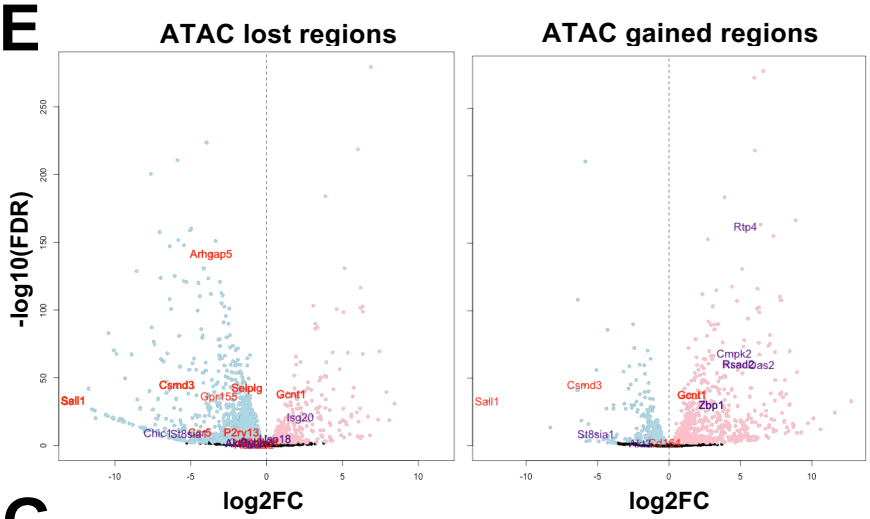
Lost accessible regions

Motif	$\log_{10}(p)$	T/B (%)	Best Match
GGAAGTGAAACT	-5.7e+03	40.37/3.32	PU.1-IRF8
CGAAACTA	-7.1e+02	17.63/5.59	Irf4
TTCCGCTT	-4.4e+02	19.09/8.44	Stat2
TGTGACACCACA	-5.8e+01	0.24/0.01	RUNX2

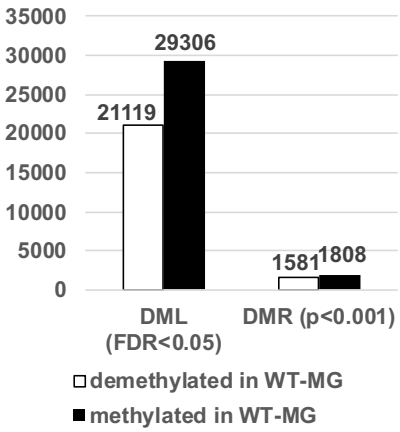
Gained accessible regions

Motif	$\log_{10}(p)$	T/B (%)	Best Match
AAAAGAGGAAGT	-3.2e+03	55.32/18.35	Spfi1
ATTGCACAATAC	-1.7e+03	26.15/6.92	Cebp
CGGATGAGTCAC	-1.4e+03	22.03/5.88	AP-1(bZIP)
TGTGGTCA	-7.4e+02	21.97/8.89	RUNX(Runt)

E



F



G

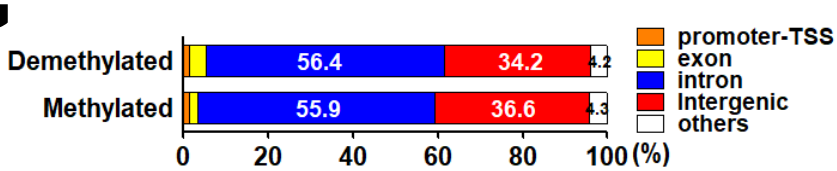


Fig.4 Saeki et al.

H IRF8 peaks, H3K4me1 and H3K27ac marks in methylated and demethylated regions

	Methylated in WT(n=1808)	Demethylated in WT(n=1581)
IRF8 peaks (n=16935)	41	721
WT-specific H3K4me1 (n=3882)	5	230
KO-specific H3K4me1 (n=4952)	254	1
WT-specific H3K27ac (n=2819)	8	222
KO-specific H3K27ac (n=1275)	33	0

I

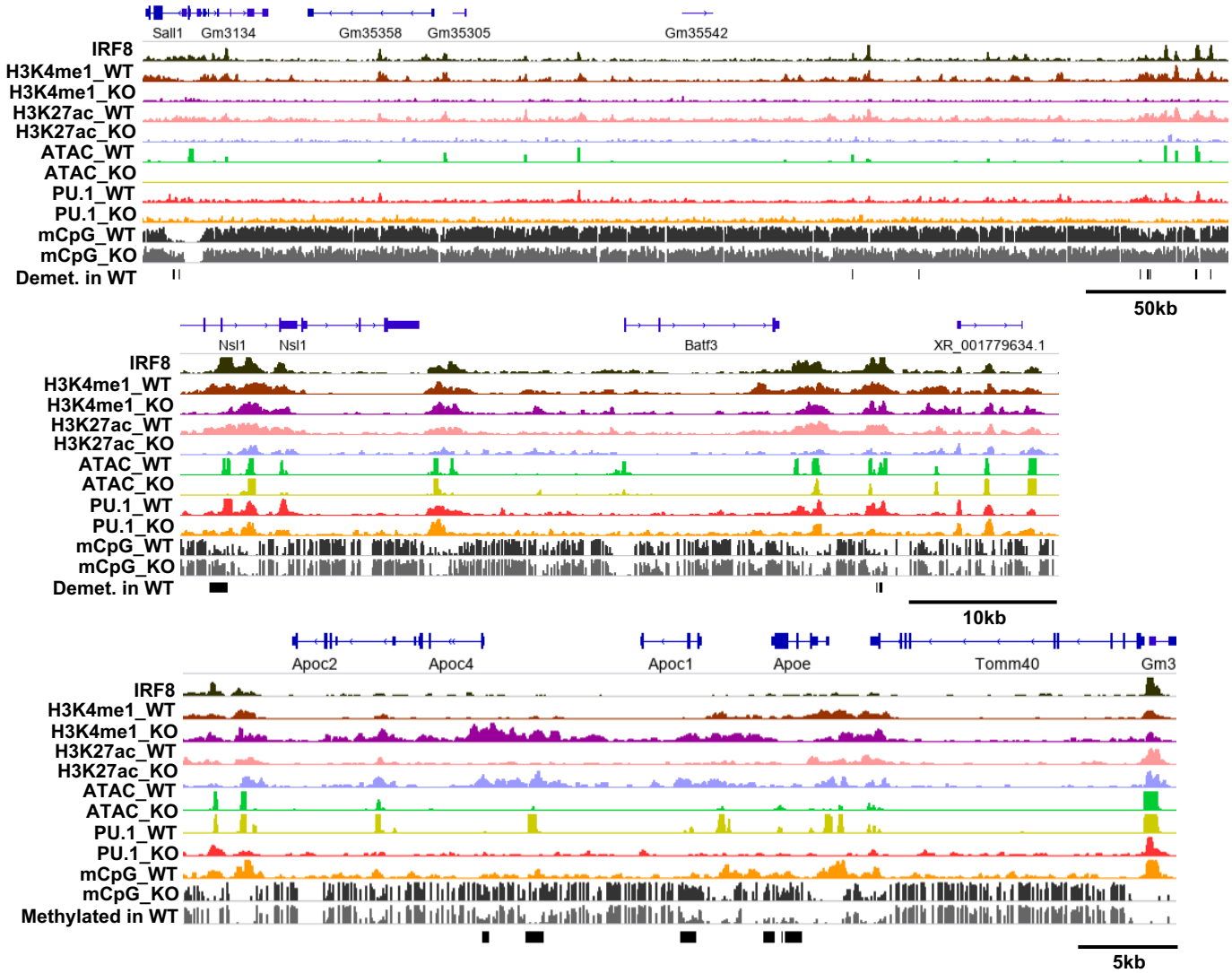


Fig.4 Saeki et al.

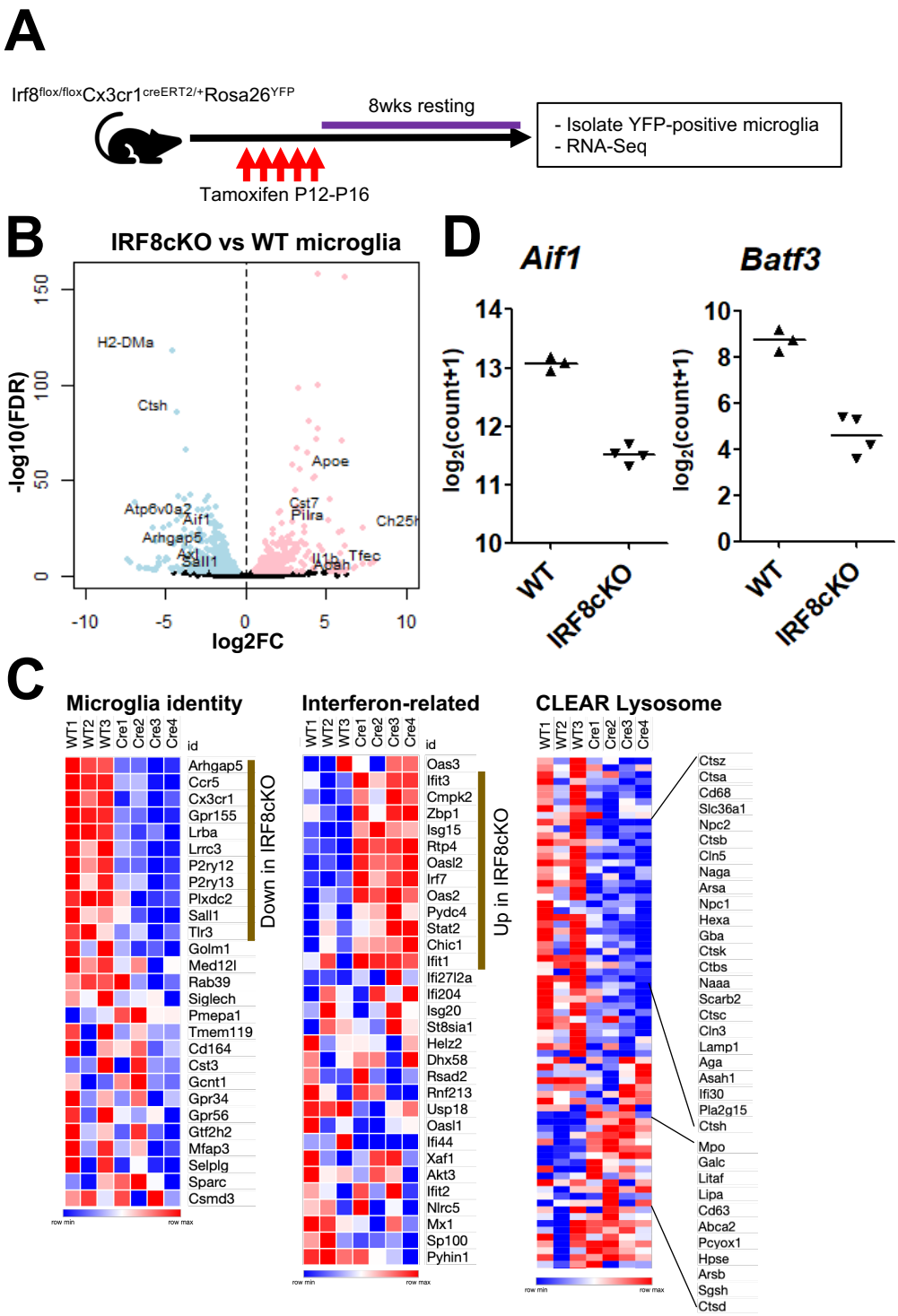


Fig.5 Saeki et al.

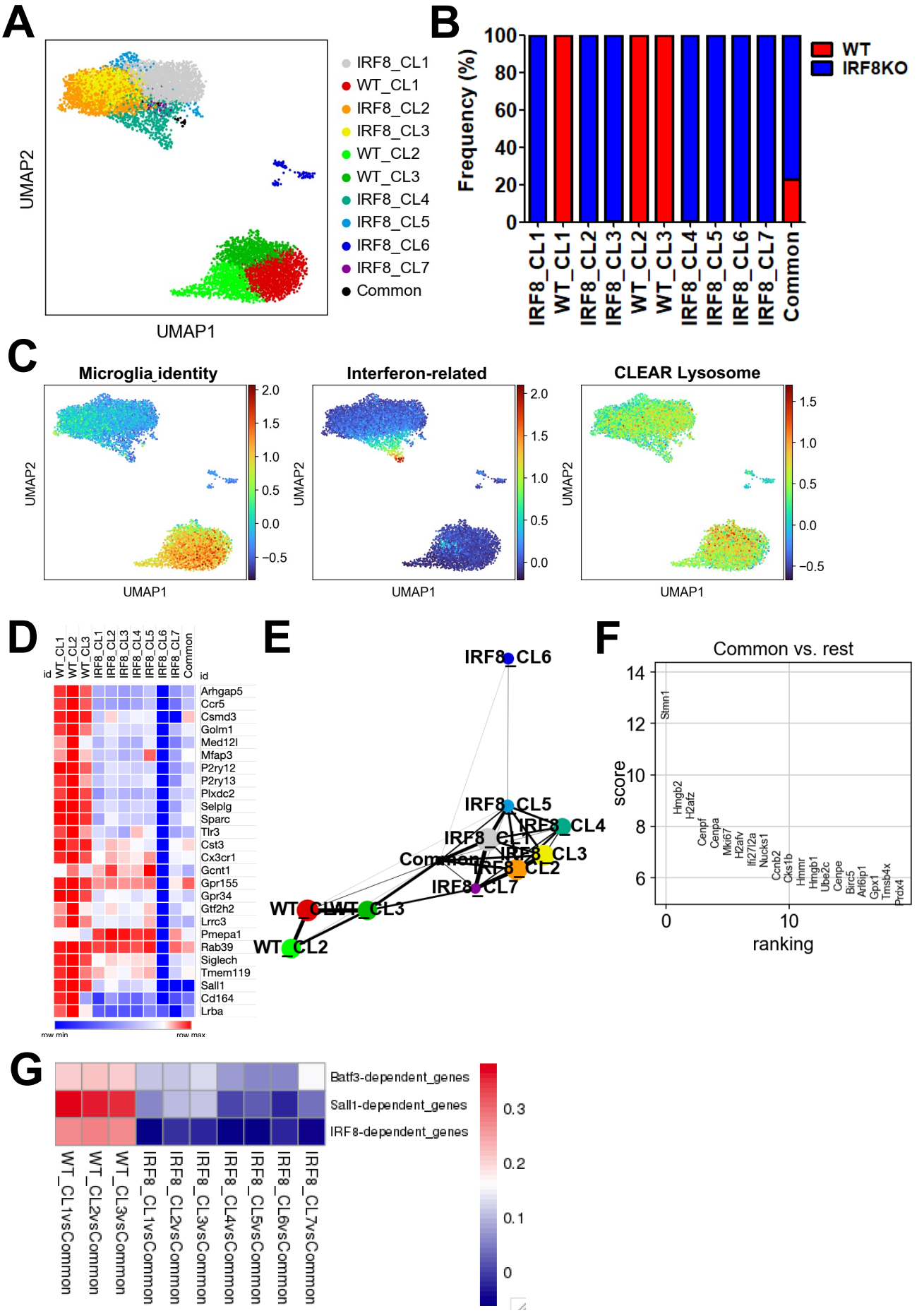


Fig.6 Saeki et al.

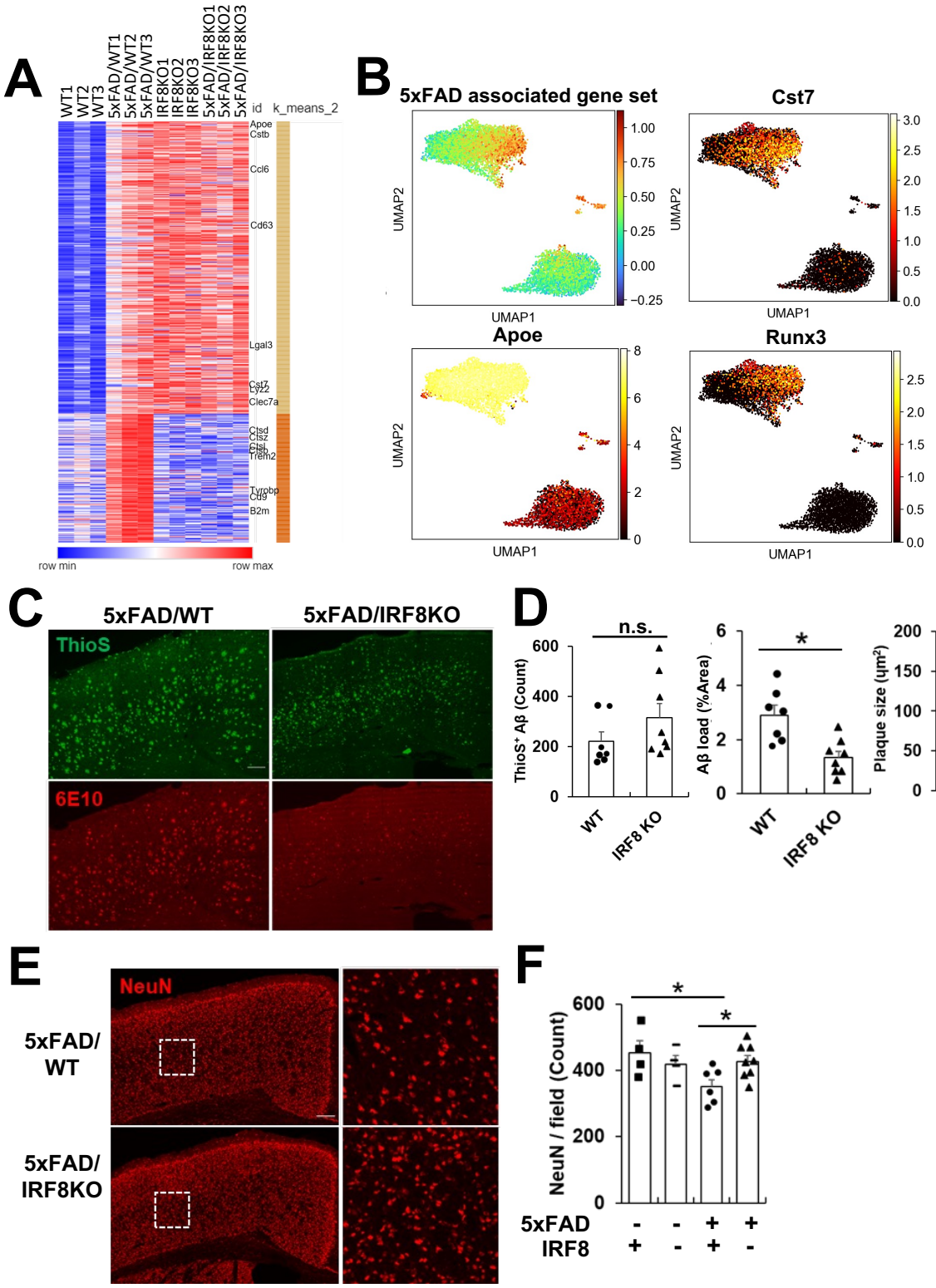


Fig.7 Saeki et al.

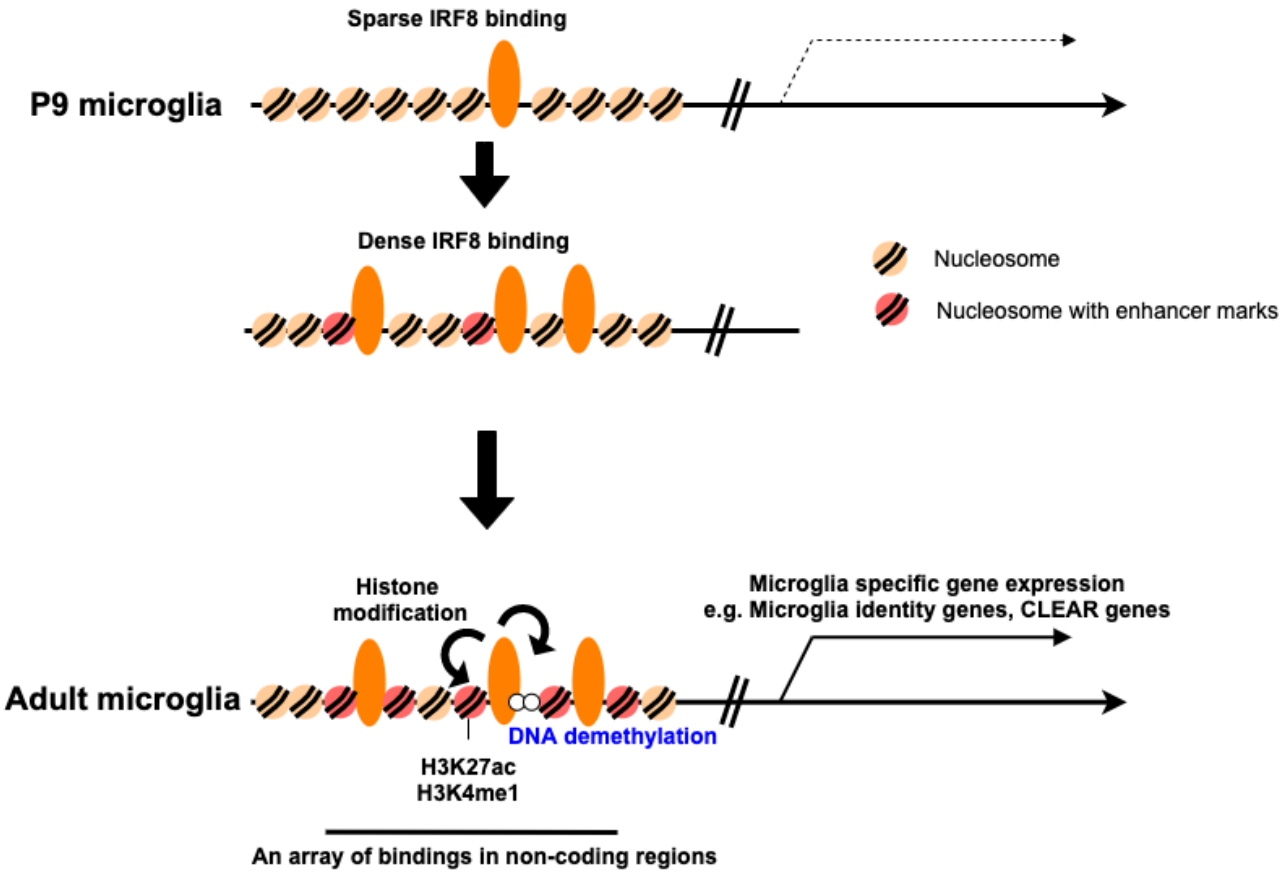


Fig.8 Saeki et al.