nature immunology

TECHNICAL REPORT

https://doi.org/10.1038/s41590-018-0110-6

Re-evaluating microglia expression profiles using RiboTag and cell isolation strategies

Zhana Haimon[1], Alon Volaski[1], Johannes Orthgiess[2], Sigalit Boura-Halfon[1], Diana Varol[1], Anat Shemer[1], Simon Yona[1], Binyamin Zuckerman[3], Eyal David[1], Louise Chappell-Maor[1], Ingo Bechmann[4], Martin Gericke[4], Igor Ulitsky[3] and Steffen Jung[1*]

Transcriptome profiling is widely used to infer functional states of specific cell types, as well as their responses to stimuli, to define contributions to physiology and pathophysiology. Focusing on microglia, the brain's macrophages, we report here a side-by-side comparison of classical cell-sorting-based transcriptome sequencing and the 'RiboTag' method, which avoids cell retrieval from tissue context and yields translatome sequencing information. Conventional whole-cell microglial transcriptomes were found to be significantly tainted by artifacts introduced by tissue dissociation, cargo contamination and transcripts sequestered from ribosomes. Conversely, our data highlight the added value of RiboTag profiling for assessing the lineage accuracy of Cre recombinase expression in transgenic mice. Collectively, this study indicates method-based biases, reveals observer effects and establishes RiboTag-based translatome profiling as a valuable complement to standard sorting-based profiling strategies.

Cellular functions are defined by transcriptomes and proteomes. Global gene expression profiling can hence provide insights into specific contributions of distinct cell types to various physiological processes. Macrophages are myeloid immune cells that are strategically positioned to ingest and degrade dead cells, debris and foreign material and to orchestrate inflammation and immune defense. Moreover, emerging evidence supports additional critical tissue macrophage contributions to the establishment and maintenance of organ functions. Studies have highlighted the impact of the tissue environment on macrophage expression signatures and enhancer landscapes[1,2]. Conversely, tissue macrophages lose distinct expression patterns once taken into culture[2], likely as a result of the loss of original environmental cues and exposure to new ones. Accurate expression profiling of cells to infer in vivo functions therefore requires methods that allow efficient and rapid retrieval of phenotypically specified cells or their RNA from intact organs.

Classically, the isolation of defined cell populations from their physiological tissue context involves the preparation of single-cell suspensions followed by flow-cytometry- or magnetic-bead-based cell sorting. Depending on the cell type studied and its respective extent of tissue embedding, release of the cells can require mechanical processing and extensive enzymatic digestion with prolonged incubations. Collectively, these manipulations confer the inherent risk of artifacts. Moreover, cell isolation protocols are often inefficient and prone to introduce bias toward subpopulations. Even optimized isolation protocols fail, for instance, to retrieve more than 10% of microglia cells from an intact mouse brain, estimated to comprise 3 million cells.

To circumvent the need for cell retrieval, alternative approaches were introduced that allow isolation of cell-specific translatomes by immunoprecipitation (IP) of epitope-tagged ribosomes from crude tissue extracts[3,4]. In the RiboTag approach developed by McKnight and colleagues[4], cell-type-specific expression of Cre recombinase is used to activate expression of a hemagglutinin (HA) epitope-tagged ribosomal subunit (RPL22) by deletion of a *loxP*-flanked (floxed) wild-type exon. IP of the tagged ribosomes from whole tissue extracts with anti-HA antibody-coupled magnetic beads enables the pulldown of cell-type-specific ribosome-attached mRNA; i.e., the translatome.

Here we report the application of the RiboTag approach to the study of microglia. Specifically, we compared previously reported *Cx3cr1*^{Cre} and *Cx3cr1*^{CreER} transgenic animals[5] for their potency and specificity to be used in microglial expression profiling using the RiboTag strategy. Side-by-side comparison of translatomes isolated by IP from crude tissue extracts and transcriptomes from sorted microglial cells highlights the advantages and disadvantages of the respective approaches. Whole-cell transcriptomes were found to be contaminated by artifacts induced by tissue dissociation, contamination with cargo and transcripts sequestered away from ribosomes. Finally, we performed a translatome analysis on microglia from animals exposed to acute peripheral endotoxin challenge. Collectively, our results highlight the specifics of RiboTag profiling and establish this method as a valuable complement to standard sorting-based profiling strategies.

Results

Definition of cell type specificity of *CX3CR1*^{Cre} and *CX3CR1*^{CreER} transgenic mice. The RiboTag strategy is a two-component approach relying on the combination of a floxed *Rpl22*^{HA} allele[4] with a cell-type-specific Cre recombinase transgene. Microglia display unique high expression of CX_3CR1[6], and transgenic mice harboring a GFP reporter gene under the promoter of this chemokine receptor have been instrumental in studying microglial morphology and dynamics, as GFP expression in adult mouse brains is restricted to microglia and non-parenchymal macrophages[6,7]. More recently, we

[1]Department of Immunology, Weizmann Institute of Science, Rehovot, Israel. [2]Carl Ludwig Institute of Physiology, University of Leipzig, Leipzig, Germany. [3]Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel. [4]Institute of Anatomy, University of Leipzig, Leipzig, Germany. *e-mail: s.jung@weizmann.ac.il

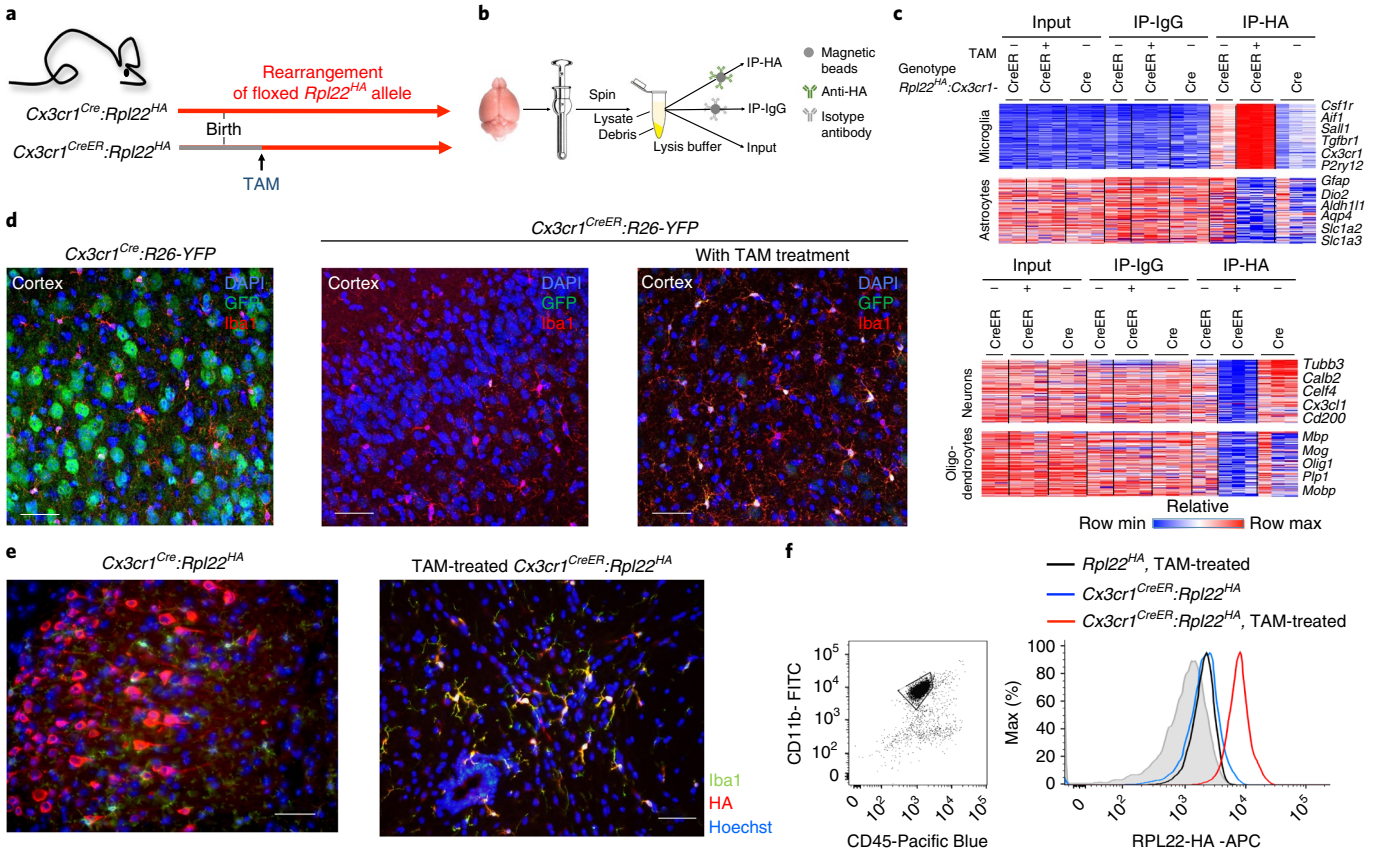

Fig. 1 | RiboTag analysis reveals that *Cx3cr1*^{Cre} mice but not *Cx3cr1*^{CreER} animals display rearrangements in neurons. a, The *Cx3cr1*^{Cre} and *Cx3cr1*^{CreER} systems. **b**, The IP protocol, including brain homogenization, centrifugation to remove cell debris and incubation with magnetic beads and relevant antibodies. **c**, Heat maps of RNA-seq data comparing IPs obtained from brains of *Cx3cr1*^{Cre}:*Rpl22*^{HA} and *Cx3cr1*^{CreER}:*Rpl22*^{HA} mice, represented by lists of genes of microglia (115), neurons (97), astrocytes (95) and oligodendrocytes (98) showing enrichment and de-enrichment of mRNAs of specific cell types in the different samples. Reference datasets in ref. [9]. Each column represents an individual mouse, $n = 2$ for *Cx3cr1*^{CreER} no TAM, $n = 3$ for *Cx3cr1*^{Cre} and *Cx3cr1*^{CreER} with TAM. **d**, Microscopic analysis of cortical brain sections from *Cx3cr1*^{Cre}:*R26-YFP* mice (left) and *Cx3cr1*^{CreER}:*R26-YFP* mice (TAM treated (right) or untreated controls (middle)), stained for IBA-1, YFP and DAPI, showing neuronal expression of YFP in *Cx3cr1*^{Cre} brains and microglia-restricted YFP expression in *Cx3cr1*^{CreER} brains. The animals analyzed are F_1 offspring of the intercross of homozygous *Cx3cr1*^{CreER} or *Cx3cr1*^{Cre} animals and homozygous *R26-YFP* mice. Representative of 2 independent experiments. Scale bars, 200 µm. **e**, Immunofluorescence staining of tissue sections of *Cx3cr1*^{Cre}:*Rpl22*^{HA} (left) and TAM-treated *Cx3cr1*^{CreER}:*Rpl22*^{HA} (right) mice, stained for IBA1, HA and Hoechst, showing neuronal expression of HA in *Cx3cr1*^{Cre} cortex and microglia-restricted HA expression in *Cx3cr1*^{CreER} spinal cord. Scale bars: 200 µm left, 50 µm right. Representative of 2 independent experiments. **f**, Flow cytometry analysis showing HA staining in microglia (CD11b^{+} CD45^{int}, gated on Ly6C/G^{-} DAPI^{-}) of *Rpl22*^{HA} TAM-treated mice (black line) and *Cx3cr1*^{CreER}:*Rpl22*^{HA} mice, untreated (blue line) and TAM-treated (red line). Shadowed histogram represents isotype (IgG) control. Representative data of 3 repeats.

introduced two mouse strains that display Cre recombinase activity under the control of the *Cx3cr1* promoter, either constitutively (*Cx3cr1*^{Cre} mice) or following tamoxifen (TAM)-mediated activation of an estrogen receptor-fused latent Cre recombinase (*Cx3cr1*^{CreER} mice)[5,8] (Fig. 1a). To implement the RiboTag method for the study of microglia, we generated *Cx3cr1*^{Cre}:*Rpl22*^{HA} and *Cx3cr1*^{CreER}:*Rpl22*^{HA} mice, all homozygous for the *Rpl22*^{HA} allele. We then performed high-throughput RNA sequencing (RNA-seq) on RNA isolated from whole brain tissue (input), RNA retrieved by an isotype control IP (IP-IgG) and RNA retrieved by anti-HA IP (IP-HA) of brain extracts from the two mouse strains (Fig. 1b). To assess the cell type specificity of the obtained translatomes, we compared them to published neuron- and glia-specific gene expression signatures[9]. Translatomes retrieved from *Cx3cr1*^{Cre}:*Rpl22*^{HA} and TAM-treated *Cx3cr1*^{CreER}:*Rpl22*^{HA} mice showed an enrichment for mRNAs encoding microglial proteins, such as Sall1, Csf1r, Trem2, Aif1 (Iba-1) and CD11b, which represent a small fraction in the total input, confirming rearrangement of the floxed *Rpl22*^{HA} allele in microglia (Fig. 1c). Conversely, key astrocyte and oligodendrocyte transcripts, such as *Gfap*, *Aldh1l1*, *Aqp4* and *Mbp*, *Mog*, *Olig1*, *Plp1*, respectively, were de-enriched in both translatomes as expected (Fig. 1c). Translatomes retrieved by ribosome IP from brain homogenates of *Cx3cr1*^{Cre}:*Rpl22*^{HA} but not TAM-treated *Cx3cr1*^{CreER}:*Rpl22*^{HA} mice also exhibited a prominent neuronal signature, including mRNAs encoding calbindin 2, CX_3CL1 and CELF (Fig. 1c and Supplementary Fig. 1). This suggested activation of RPL22-HA expression in neuronal cells of *Cx3cr1*^{Cre}:*Rpl22*^{HA} mice. Analysis of *Cx3cr1*^{Cre}:*Rosa26*^{YFP} animals, which harbor a floxed reporter allele, revealed prominent neuronal labeling, comparable to that recently reported for *LysM*^{Cre} mice[10] (Fig. 1d and Supplementary Fig. 2a). Moreover, immunohistochemical analysis of *Cx3cr1*^{Cre}:*Rpl22*^{HA} animals detected neuronal staining by anti-HA antibodies in spinal cord and brain sections, including in Purkinje cells in the cerebellum, in line with the observed *Calb2* expression (Fig. 1e and Supplementary Fig. 2b). In contrast, RPL22-HA expression in TAM-treated *Cx3cr1*^{CreER}:*Rpl22*^{HA} mice was restricted to microglia, as demonstrated by co-staining for IBA-1 (Fig. 1e and Supplementary Fig. 2b). Since *Cx3cr1*^{gfp} mice lack GFP labeling in adult neurons[6,11], rearrangements in *Cx3cr1*^{Cre} mice are likely due to a transient and yet-to-be-defined window of *Cx3cr1* promoter activity during neuronal development. In support of this

notion, one of the *Cx3cr1*^{Cre}*:Rpl22*^{HA} mice analyzed also displayed astrocyte and oligodendrocyte transcripts, in line with the shared neuroectodermal origin of these glia cells and neurons (Fig. 1c). Of note, rearrangements in neurons of *Cx3cr1*^{Cre}*:Rpl22*^{HA} animals were observed irrespective of whether the floxed allele and the Cre transgene went together through the germline (data not shown).

Cx3cr1^{CreER}*:Rpl22*^{HA} mice that were TAM-treated postnatally displayed brain macrophage-restricted activation of the RiboTag. Some enrichment for the microglia translatome was also observed without TAM treatment in these mice, and rare YFP$^+$ cells could be detected in untreated *Cx3cr1*^{CreER}*:Rosa26*^{YFP} animals (Fig. 1c,d middle panel), corroborating reports of leakiness of the CreER system[12]. However, as confirmed by flow cytometric analysis (Fig. 1f), robust rearrangement and microglial expression of the HA epitope-tagged ribosome subunit in *Cx3cr1*^{CreER}*:Rpl22*^{HA} mice were dependent on TAM induction in our facility. Collectively, these data illustrate the value of the RiboTag profiling approach for assessing the accuracy of Cre transgenic mouse models and investigating specific cell types, including fate mapping and conditional mutagenesis.

Comparison of RiboTag profiling to cell-sorting-based transcriptomics. Having established the value of *Cx3cr1*^{CreER}*:Rpl22*^{HA} mice, we next compared translatomes and transcriptomes of sorted and unsorted microglia. Specifically, we divided individual brains of TAM-treated *Cx3cr1*^{CreER}*:Rpl22*^{HA} mice: one hemisphere underwent direct tissue homogenization followed by IP-HA or an IP-IgG control to define the method-related background. The second hemisphere was subjected to the classical microglia isolation protocol involving tissue digestion followed by cell sorting of microglial cells (defined as DAPI$^-$ Ly6C/G (Gr1)$^-$ CD11b$^+$ CD45^{int}) (Fig. 2a and Supplementary Fig. 3). A fraction of the sorted microglia was taken for direct mRNA isolation to yield the whole transcriptome (Sort); another fraction was lysed and subjected to IP-HA to retrieve the translatome of sorted cells (Sort-IP) (Fig. 2a). This experimental setup allowed comparison of translatomes of sorted and unsorted microglia alongside whole transcriptomes of sorted microglia from the same brain, and thus investigation of the impact of the isolation protocol on gene expression.

Unbiased *k*-means clustering of the significantly differentially expressed genes between at least two sample groups (IP-HA vs. IP-IgG, IP-HA vs. Sort, IP-HA vs. Sort-IP and Sort vs. Sort-IP, fold change > 2, $P < 0.05$) revealed 2,508 differentially expressed genes, which could be divided into four clusters (Fig. 2b). Cluster IV was discerned as RiboTag method-related background, since mRNA reads in the nonspecific IP-IgG were higher than in the specific IP-HA and absent from sorted samples. IP-specific genes were selected for being significantly higher in IP-HA than IP-IgG (fold change > 2, $P < 0.05$), and all genes below this threshold were removed from the analysis. Cluster I comprised 913 mRNAs enriched in the specific IP-HA compared with IP-IgG and present in both samples of sorted microglia. This cluster includes established microglia signature genes, such as *Aif1*, *Irf8*, *Sall1*, *Cx3cr1*, *Tgfbr*, *Tmem119* and *Hexb* (Fig. 2b and Supplementary Fig. 4), indicating that the retrieval methods we used are comparable. Cluster II was represented by 525 mRNAs that were highly abundant in sorted samples (both in translatome and transcriptome) but not present in the direct IP-HA. Cluster III comprised 282 mRNAs prominently enriched in the direct IP-HA, but less abundant in the sorted microglia. Clusters II and III highlight differences between the retrieval methods, as well as discrepancies between transcriptomes and translatomes, and will be the focus of the remainder of this study.

Transcripts over-represented in microglia translatomes. Cluster III (Fig. 2b) is defined by genes highly expressed in IP-HA samples relative to Sort samples and could be further subdivided according to transcript abundance in the Sort-IP samples (Fig. 2c): cluster III-a (72 mRNAs) were low in both samples of the sorted cells (Sort and Sort-IP); clusters III-b and III-c (210 mRNAs) were low in the Sort samples but highly expressed in the Sort-IP samples (Fig. 2c).

Cluster III-a can be largely explained by the presence of non-parenchymal macrophage mRNAs. Non-parenchymal brain macrophages, including perivascular, meningeal and choroid plexus macrophages, can be discriminated from CD11b$^+$ CD45^{int} microglia as CD11b$^+$ CD45^{hi} cells and therefore excluded by fluorescence-activated cell sorting (Fig. 2d). CX_3CR1 is expressed both in microglia and in non-parenchymal brain macrophages (Fig. 2d)[13]. Accordingly, both CD45^{int} microglia and CD45^{hi} macrophages of TAM-treated *Cx3cr1*^{CreER}*:Rpl22*^{HA} mice express the HA-tagged Rpl22 isoform (Fig. 2e). Moreover, some of the non-parenchymal brain macrophages are long-lived like microglia and hence the population does not lose the rearranged alleles as monocytes do[13]. Single-cell transcriptomics have shown that non-parenchymal brain macrophages differ in gene expression from microglia[13,14]. Accordingly, cluster III-a included *Cd163*, *F13a1*, *Cbr2*, *Mrc1 and Lyve1* (Fig. 2f,g).

The combined clusters III-b and III-c comprise mRNAs that are enriched in both IP-HA and Sort-IP translatomes over the whole transcriptomes, suggesting their functional importance for the cells. These include mRNAs encoding proteins related to metabolism (*Gpx1*, *Sdhc*), vesicular transport (*Clta*, *Kdelr1*, *Ykt6*), sphingolipid metabolism (*Gm2a*, *Psap*) and lipids (*Apoe*), as well as components of the GABA-receptor signaling cascade (*Gabarap*, *Gnai2*) (Fig. 2h). Specific functions of these genes in microglia remain to be explored.

Collectively, these results highlight the value of a multifaceted approach, combining the RiboTag and cell sorting strategies to improve cell type specificity. In addition, the RiboTag approach allows us to focus specifically on genes that are actively being translated and contributing to the cellular proteome at a particular time and location.

Transcripts over-represented in transcriptome cluster I: isolation artifacts. Cluster II (Fig. 2b) consists of 525 mRNAs that were high in Sort but low in IP-HA. This cluster can be further subdivided according to transcript abundance in the Sort-IP samples (Fig. 3a). Cluster II-a comprised 190 transcripts similarly expressed in Sort-IP and Sort samples and was found to include mRNAs that are related to immune activation, such as *Cd86*, *Cd53*, *Tlr4*, and *Tlr7* (Fig. 3b). Ingenuity pathway analysis (IPA) of significantly upregulated genes in Sort vs. IP-HA (fold change > 2, $P < 0.05$) showed upregulation of pathways such as "Production of NOS and ROS," "Phagocytosis" and "TLR signaling" (Fig. 3c).

Since these transcripts are high in both translatomes and transcriptomes of sorted cells, we assume that they reflect cell activation resulting from the isolation process, as recently reported from another system[15]. Commonly used macrophage isolation protocols, such as the one we applied for the microglia retrieval, include enzymatic tissue digestion at 37 °C, a step that could cause cell activation and transcriptome alterations. Moreover, enzymes employed in these digests might contain endotoxin contaminations that could activate cells. To probe for the potential impact of these manipulations, we compared transcriptomes of sorted cells that were isolated from the same brain with or without collagenase and DNase digestion, as well as the RiboTag approach. Surprisingly, though, both isolation procedures resulted in comparable transcriptional profiles, indicated by differential expression of 472 and 267 genes (clusters II and III, respectively), as compared to the relevant IP samples (Supplementary Fig. 5a). Global correlation of gene expression of the samples retrieved with or without incubation was high ($r^2 = 0.99$), as compared to the correlation of sorted and IP samples ($r^2 = 0.1$) (Supplementary Fig. 5b). Similarities were also apparent in a correlation matrix compiled from data of independent experiments (Supplementary Fig. 5c). Collectively, these data establish

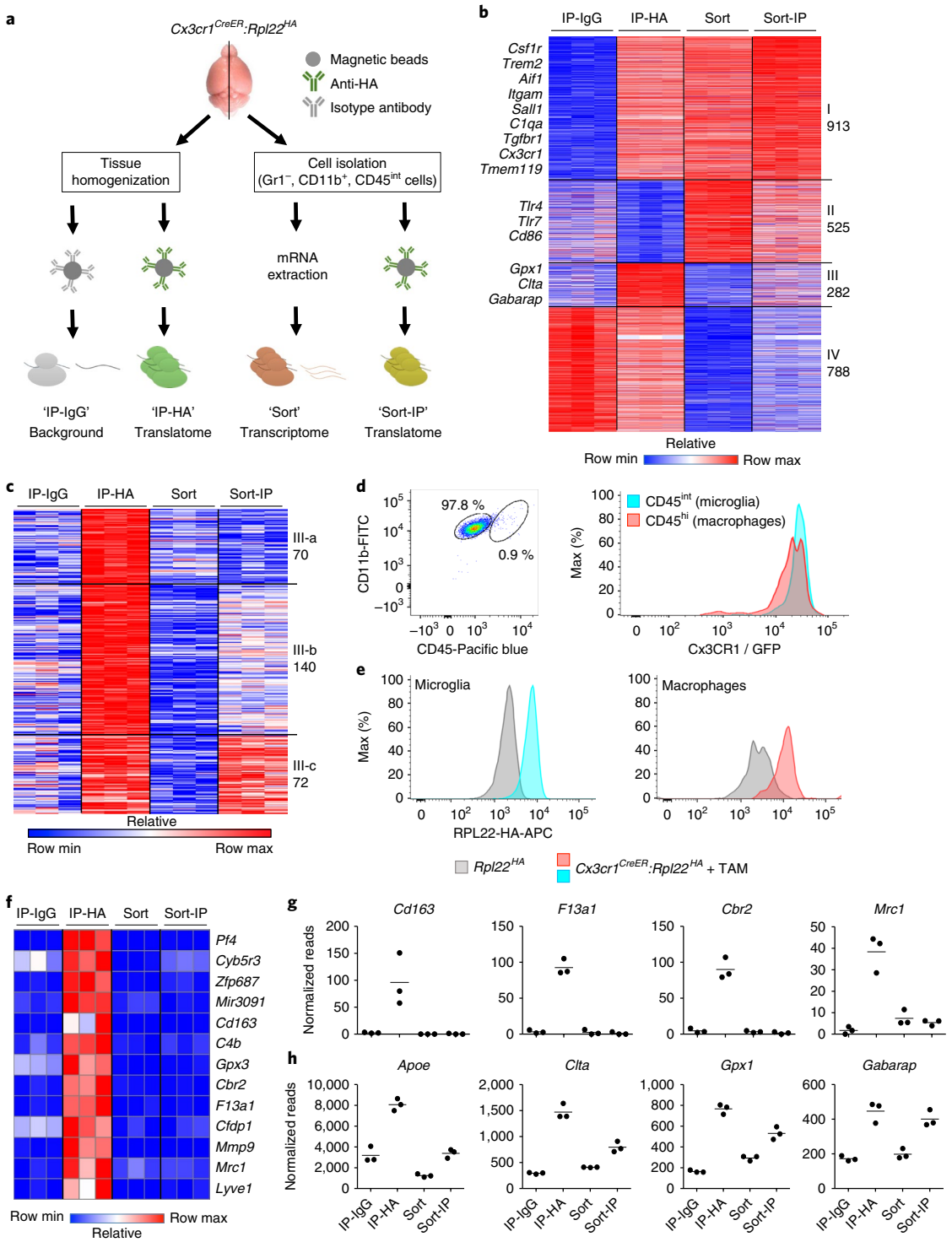

Fig. 2 | Comparison of cell-sorting-based protocol and the RiboTag method for profiling microglia. a, Experimental protocol comparing RiboTag and cell-sorting-based strategies. **b**, Heat map of RNA-seq data of samples obtained in a. Genes selected by maximum value > 100 normalized reads (3,186 out of 17,406 genes), significantly changed (fold change > 2, $P < 0.05$) between the following: IP-HA vs. IP-IgG, Sort vs. IP-HA, Sort-IP vs. IP-HA and Sort vs. Sort-IP, representing 2,508 genes. $n = 3$, individual mice. Statistical test was part of the DESeq2 package, using adjusted P values. **c**, Heat map representing k-means reclustering of genes in cluster III from **b**, showing genes high in IP-HA and low in Sort samples. **d**, Fluorescence-activated cell sorting (FACS) dot plot (left) showing separation of microglia ($CD45^{int}$) from other brain macrophages (MΦ; $CD45^{hi}$) by flow cytometry. Histogram (right) of microglia and MΦ isolated from $Cx3cr1^{GFP}$ mice indicating high *Cx3cr1* promoter-driven GFP expression in both populations. Representative of 3 independent experiments. **e**, FACS histogram of HA staining in microglia (left) and MΦ (right) in control $Rpl22^{HA}$ mice (gray) or TAM-treated $Cx3cr1^{CreER}$:$Rpl22^{HA}$ mice (cyan or red). Representative of 2 independent experiments. **f**, Heat map of RNA-seq data of representative non-parenchymal brain macrophage genes, showing enrichment in IP-HA but not in sorted samples. **g**, Graphs showing normalized reads of example genes from **f**. Each dot represents an individual mouse, $n = 3$, line represents mean. **h**, Graphs showing normalized reads of example genes from cluster III-b and III-c in **c**, showing functional genes enriched in IP and Sort-IP. Each dot represents an individual mouse, $n = 3$, line represents mean.

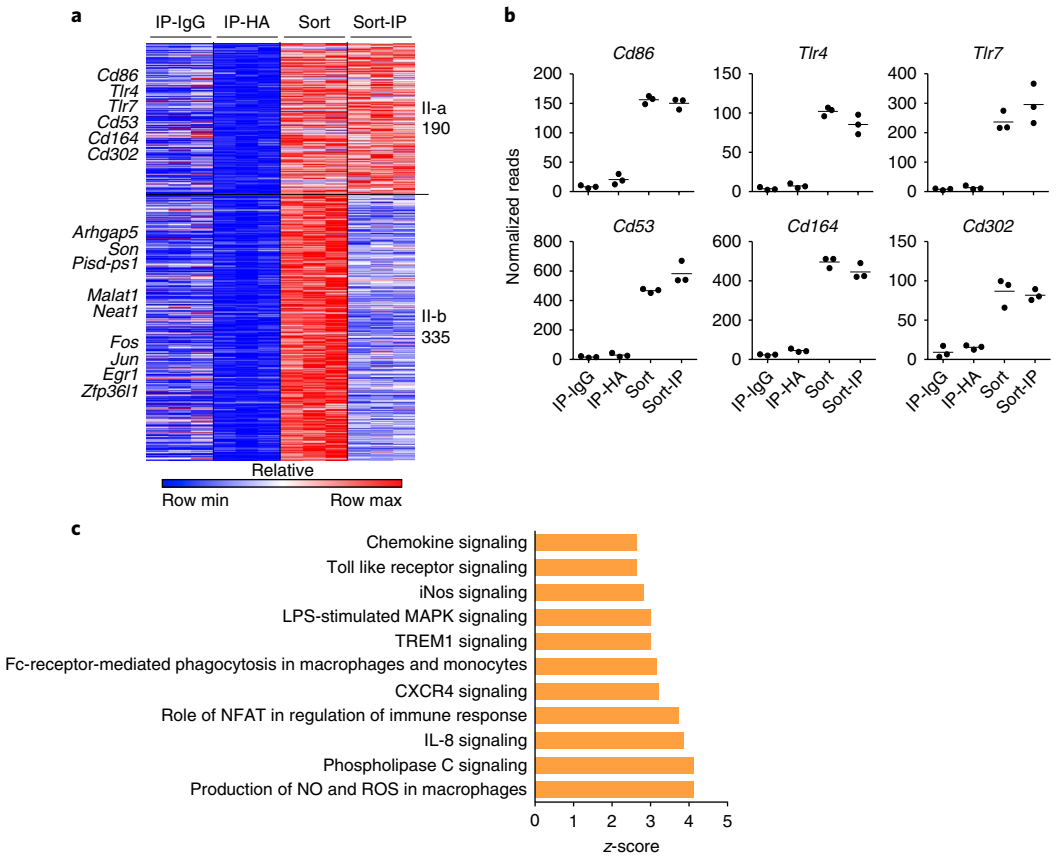

Fig. 3 | Microglial isolation results in cell activation. a, Heat map representing *k*-means reclustering of genes from cluster II in Fig. 2b, showing mRNAs high in Sort and low in IP-HA samples. **b**, Graphs showing normalized reads of example genes from cluster II-a in **a**, showing high expression of immune activation-related genes in sorted samples. Each dot represents an individual mouse, $n=3$, line represents mean. **c**, Ingenuity pathway analysis (IPA) of genes significantly higher (fold change > 2, $P<0.05$, according to DESeq2 statistical analysis) in Sort compared to IP-HA ($n=3$), showing activated pathways related to immune response represented by activation *z*-score as calculated by IPA software.

that the artifact is reproducible and suggest that it is introduced by extraction of the cells from their native environment, rather than subsequent manipulation.

Transcripts over-represented in the transcriptome cluster II: cargo contaminants and sequestered RNAs. Cluster II-b spanned 335 genes that were low in the Sort-IP compared with Sort samples (Fig. 3a), suggesting translatome–transcriptome differences. Microglia are specialized phagocytes that, like other macrophages, take up dead cells and cell debris for clearance[16]. Although we did not formally rule out other sources of contamination, whole-cell transcriptomes could hence include genetic material from recently ingested neighboring cells. Indeed, almost half of the mRNAs in cluster II-b (157 out of 335 genes) were likely to be derived from such external sources (Fig. 4a). Examples include *Arhgap5*, *Son* and *Pisd-ps1*, which are reportedly transcribed in astrocytes and neurons[9] (Fig. 4b).

Long noncoding mRNAs (lncRNAs) are enriched in nuclei[17], where some of them act in transcriptional regulation. As expected, representatives of these lncRNAs, such as *Malat1* and *Neat1*, were identified in the whole cell transcriptomes, while absent from translatomes, and appeared in cluster II-b (Fig. 4c).

Gene expression is controlled at the level of transcription and translation. The latter comprises specific mechanisms that prevent mRNAs from their integration into ribosomes, including nuclear retention and sequestration into dedicated membraneless cytosolic ribonucleoprotein complexes[18,19]. The content of these organelles, such as processing bodies (P-bodies) and stress granules, is only beginning to be defined[19,20]. However, sequestered mRNAs have been reported to be longer and to comprise extended 3′ untranslated regions (UTRs), as well as to display lower splicing efficiencies[19,21]. When analyzed for these three parameters, mRNAs defined by cluster II-b showed significant presence of these hallmarks, as compared to all other clusters (Fig. 4d). Moreover, transcripts of cluster II-b also showed significant overlap with the list of nuclear retained mRNAs reported for other cellular systems[22] (Fig. 4e). Among the protein-coding mRNAs that seem sequestered from immediate translation, we found *Fos*, *Jun*, *Egr1* and *Zfp36l1* (Fig. 4f), which are immediate-early genes that have been described to be induced within minutes after activation. Of note, these mRNAs appear also in the translatome of the sorted cells, suggesting that they move to the ribosomes during the isolation procedure. Collectively, the discrepancies we observed between microglia translatomes and whole-cell transcriptomes can be explained by the inclusion of cargo-derived transcripts and mRNAs sequestered to nuclei or P-bodies (Fig. 4g). These data highlight the value of the RiboTag approach for retrieving functionally relevant mRNAs.

RiboTag analysis of microglial response to peripheral LPS challenge. Arguably, method-related artifacts, such as the ones associated with microglia isolation, could be neutralized if controls and experimental samples were prepared using the same approach. However, this assumes that artifacts introduced by the isolation are not affected by biological treatments and challenges. To

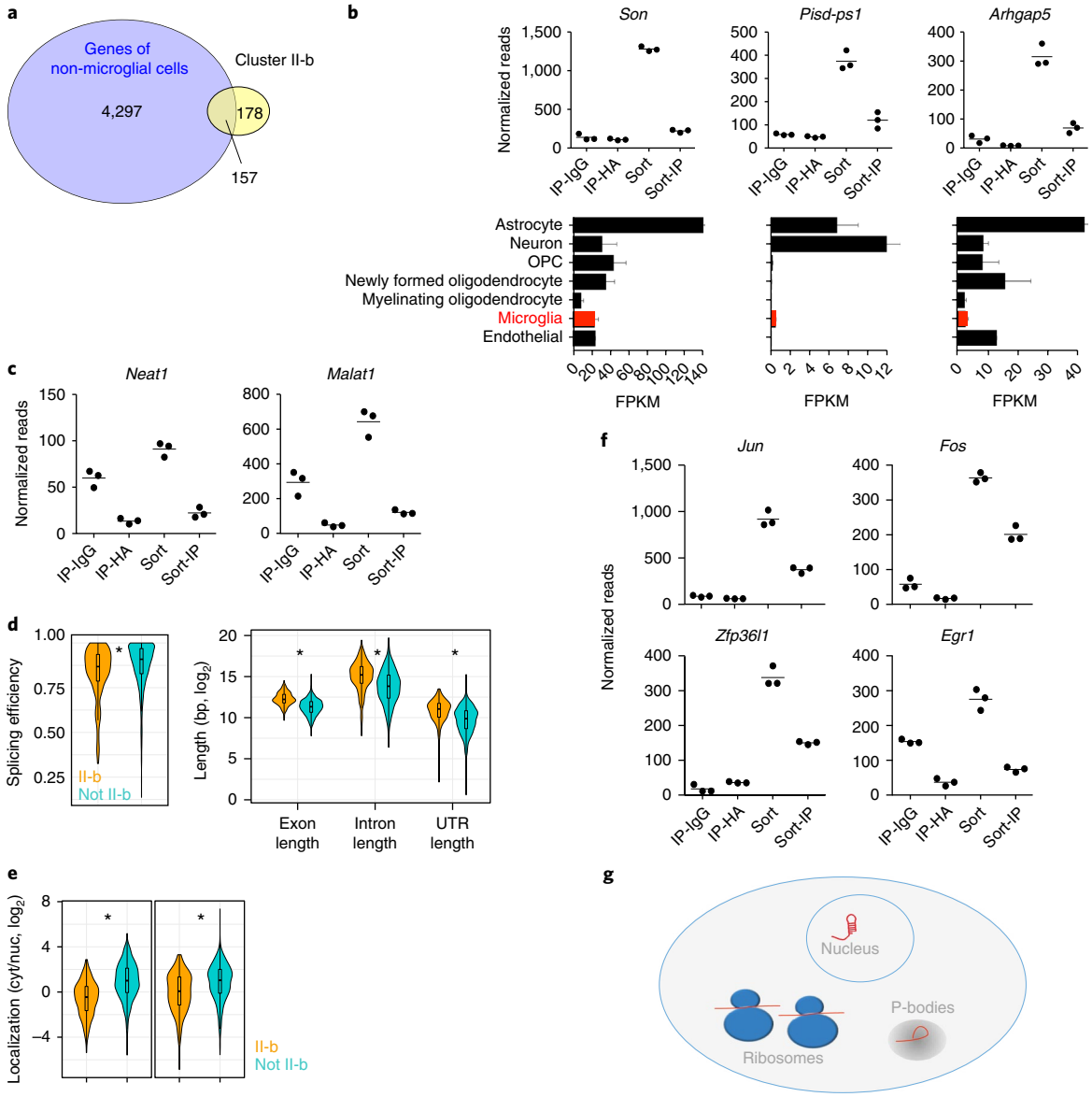

Fig. 4 | Microglia transcriptomes, but not translatomes, include cargo-derived mRNAs and mRNAs sequestered in nuclei and P bodies. a, Venn diagram of overlapping genes of cluster II-b (Fig. 2c) (yellow) with genes of non-microglial cells, selected according to ref. [9]. Non-microglial genes were selected by having maximum value > 10, microglia reads < 30 and maximum value fold change > 3 over microglia. Genes fitting these criteria represent cargo from ingested cells. **b**, Graphs showing normalized reads of example genes from list of shared genes in **a**, showing high cargo contamination in Sort samples, but not in IP. Each dot represents an individual mouse, $n = 3$ (top). Expression of genes above in different brain cells, data obtained from ref. [9] (bottom). Error bars represent s.d. FKPM, fragments per kilobase of exon per million reads mapped. **c**, Graphs showing normalized reads of example genes of long noncoding RNAs that reside within the nucleus and are present only in Sort but not IP samples. Each dot represents an individual mouse, $n = 3$, line represents mean. **d**, Violin plot representing splicing efficiency (left) and gene length (right) of genes in cluster II-b (orange) compared to genes not in II-b (green), showing that genes in cluster II-b are less efficiently spliced and have longer genes and longer 3′ UTRs than the rest of genes in the dataset, suggesting nuclear retention. Splicing efficiency was computed by comparing intron-spanning and intron-crossing reads[27]. Wilcoxon test; false discovery rate correction was performed for right panel. Splicing efficiency $^*P = 4.124 \times 10^{-7}$, exon length $^*P = 6.636 \times 10^{-63}$, intron length $^*P = 1.575 \times 10^{-27}$, UTR length $^*P = 1.865 \times 10^{-35}$. **e**, Violin plot representing cellular localization of genes in other tissues (left, liver; right, MIN6 pancreatic beta cell line) with established nuclear (nuc) and cytoplasmic (cyt) fractions. Genes in cluster II-b (orange) are more nuclear than other genes (green). Wilcoxon test and false discovery rate correction. Liver $^*P = 7.731 \times 10^{-49}$, MIN6 $^*P = 2.980 \times 10^{-17}$. n (number of genes) = 316 (liver II-b), 1,970 (liver not II-b), 306 (MIN6 II-b), 1,846 (MIN6 not II-b). Liver and MIN6 datasets were based on 2 independent experiments[22]. **f**, Graphs showing normalized reads of immediate-early genes found in Sort but not in IP-HA (in cluster II-b, microglial genes), suggesting sequestration from translation in unsorted cells. Each dot represents an individual mouse, $n = 3$, line represents mean. **g**, Different states of mRNAs in the cell: nuclear retention, sequestration from translation in P-bodies or active translation in the ribosomes.

examine this issue, we performed the RiboTag protocol on TAM-treated *Cx3cr1*^{CreER}:*Rpl22*^{HA} mice following an intraperitoneal lipopolysaccharide (LPS) injection (2.5 mg/kg). Brain hemispheres of individual LPS- and PBS-treated animals were subjected to either homogenization or microglial isolation and sorting, and processed (Fig. 2a) and used to generate a summary heat map of the

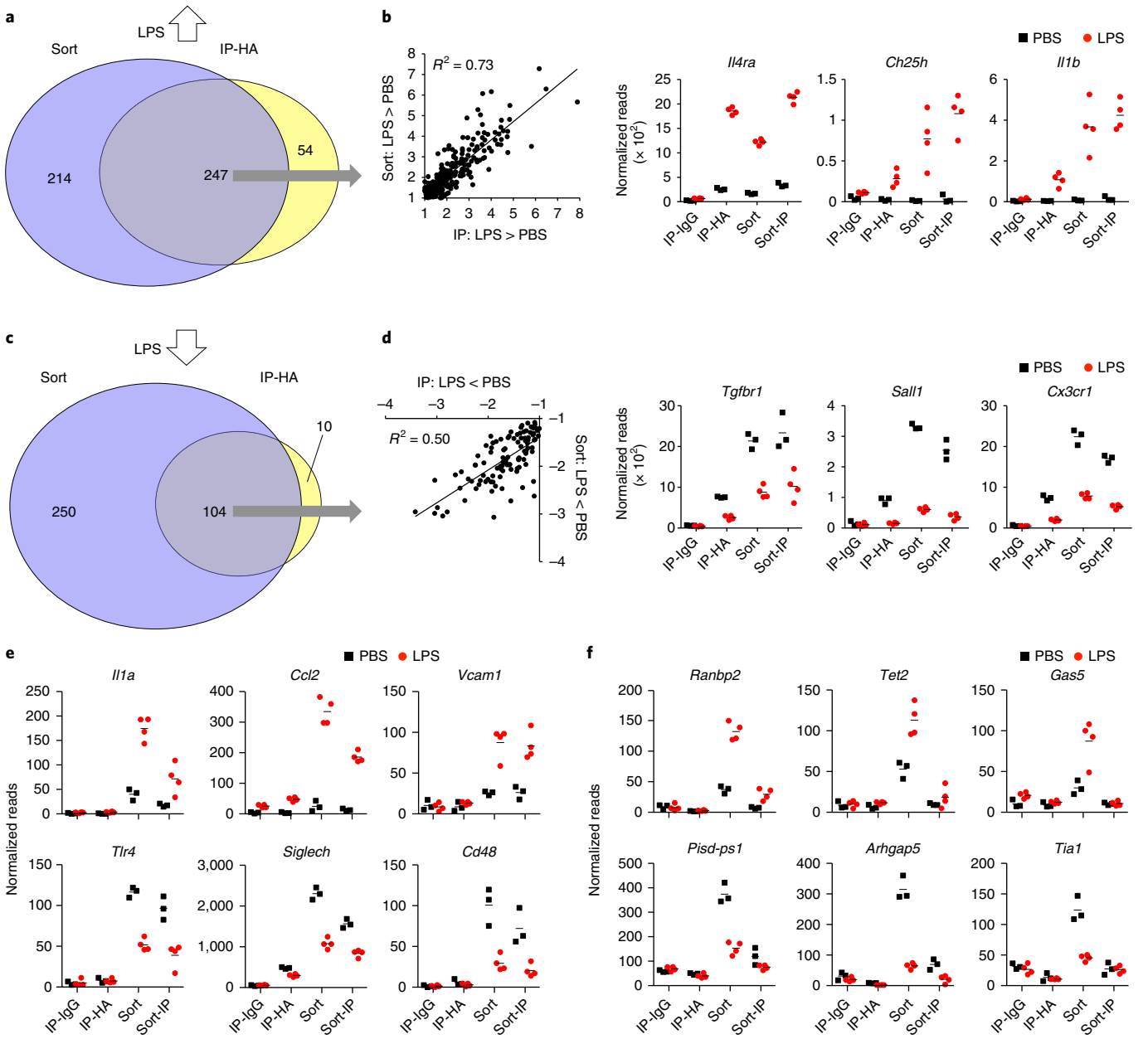

Fig. 5 | Analysis of microglia isolated from mice challenged with LPS. a, Venn diagram of overlapping genes upregulated by LPS treatment in IP-HA (blue) and in Sort (yellow), showing 54 genes upregulated in IP-HA only, 247 genes shared between methods and 214 genes upregulated by sorting only. **b**, Correlation analysis of 247 shared genes (from **a**) upregulated as a result of LPS in both methods, representing $\log_2$ fold change of significantly changed genes ($\log_2$ fold change > 1, $P < 0.05$, as calculated by DESeq2 statistical analysis) in LPS vs. PBS in each of the methods (left); graphs showing normalized reads of example genes detected as upregulated with LPS by both methods (right). Each dot represents an individual mouse, $n = 3$ in PBS group, $n = 4$ in LPS group, line represents mean. For IP-HA, genes were selected by first being enriched over control IP-IgG, and then by LPS > PBS. **c**, Venn diagram of overlapping genes downregulated by LPS treatment in IP-HA (blue) and in Sort (yellow), showing 10 genes downregulated in IP-HA only, 104 genes shared between methods and 250 genes downregulated by sorting only. **d**, Correlation analysis of 104 shared genes (from **c**) downregulated as a result of LPS in both methods, representing $\log_2$ fold change of significantly changed genes ($\log_2$ fold change < –1, $P < 0.05$, as calculated by DESeq2 statistical analysis) in LPS vs. PBS in each of the methods (left); graphs showing normalized reads of example genes detected as downregulated with LPS by both methods (right). Each dot represents an individual mouse, $n = 3$ in PBS group, $n = 4$ in LPS group, lines represent mean. For IP-HA, genes were selected by first being enriched over control IP-IgG, and then by LPS < PBS. **e**, Graphs showing normalized reads of example genes related to immune activation that were upregulated (top) or downregulated (bottom) as a result of LPS treatment in sorted samples only but not detected in IP-HA, showing differential susceptibility of biologically treated samples to artifacts introduced by isolation method. Each dot represents an individual mouse, $n = 3$ in PBS group, $n = 4$ in LPS group, lines represent mean. **f**, Graphs showing normalized reads of example genes originated from ingested cargo that were upregulated (top) or downregulated (bottom) as a result of LPS treatment in Sort only but not detected in IP-HA or Sort-IP, showing that the whole translatome includes LPS-dependent changes that do not originate in microglia, thus introducing false information. Each dot represents an individual mouse, $n = 3$ in PBS group, $n = 4$ in LPS group, lines represent mean.

RNA-seq data (Supplementary Fig. 6). To define the effect of the isolation method on microglia of LPS-treated and PBS-treated mice, we performed separate analyses for differentially expressed genes between PBS and LPS treatment in each method and then assessed the respective overlap. The majority of genes detected as up- or downregulated by the endotoxin challenge in the IP sample were shared with the sorted samples (Fig. 5a,c). Mutual genes correlated between methods and showed a similar trend of up and downregulation (Fig. 5b,d), indicating that bona fide LPS-induced changes are seen with both methods. Upregulated genes included *Il4ra*, *Ch25h* and *Il1b* (Fig. 5b), while microglial signature genes such as *Tgfbr1*, *Sall1* and *Cx3cr1* were downregulated (Fig. 5d).

Notably, a considerable number of mRNAs changed upon LPS treatment only in the sorted samples, but not in the anti-HA IP from whole-brain extract (46% of the upregulated and 71% of the downregulated genes) (Fig. 5a,c). Transcripts that were detected as changed upon LPS treatment in sorted samples (both Sort and Sort-IP) but not in IP-HA included genes related to immune activation such as *Il1a*, *Ccl2* and *Vcam1* that were upregulated (Fig. 5e) and *Tlr4*, *Siglech* and *Cd48* that were downregulated (Fig. 5e). Since they were found in the Sort-IP samples as well, these mRNAs are likely translated in sorted but not in unsorted microglia. These data establish that the artifact introduced by the cell isolation is affected by the state of the animals from which the cells are retrieved. Moreover, mRNAs that are defined as cargo contamination, owing to their presence in the transcriptomes but absence from IPs of the sorted cells, changed following the LPS challenge (Fig. 5f). For example *Pisd-ps1*, *Arhgap5* and *Tia1* were downregulated and *Ranbp2*, *Tet2* and *Gas5* were upregulated upon LPS challenge (Fig. 5f). All of these genes were reported to be more highly expressed by other brain cells, rather than microglia (Supplementary Fig. 7)[9], and are absent from the translatomes of either sorted or unsorted microglia in our dataset.

Collectively, these results indicate that data retrieved from sorted microglia include false information that originates from sorting-related immune activation and cargo contamination from ingested cells. Importantly, our data establish that identical processing of control and test samples does not necessarily neutralize these artifacts, since they themselves are affected by the biological treatment. Taking these findings together, our study highlights the advantages and disadvantages of classical sorting-based cell isolation protocols and the RiboTag approach (Supplementary Table 1).

Discussion

Here we compared experimental approaches for retrieving microglial expression signatures from brains of untreated and challenged animals. Specifically, we defined strengths and weaknesses of the classical cell isolation and sorting-based protocols and the RiboTag strategy[4], which relies on polyribosome IP from crude tissue extracts. Below we discuss the pros and cons of the respective techniques.

The RiboTag strategy was originally introduced by McKnight and colleagues and applied to expression profiling of neurons and Sertoli cells[4]. Cell type specificity of the approach depends on the accuracy of the Cre driver that is combined with the *Rpl22*^{HA} allele. This aspect is highlighted in our study by the side-by-side comparison of *Cx3cr1*^{Cre} and *Cx3cr1*^{CreER} animals, which revealed the superiority of the inducible system for achieving brain macrophage specificity and excluding neurons. However, as reported earlier and confirmed in this study, even *Cx3cr1*^{CreER} mice target not only microglia, but also non-parenchymal macrophages[13]. In situations where subpopulations can be phenotypically discriminated, as in the case of CD45^{int} microglia and CD45^{hi} perivascular macrophages, sorting-based approaches, potentially combined with a RiboTag analysis as in this study, can hence be advantageous. This emphasizes the need for the development of new Cre transgenic lines targeting microglia, including combinatorial approaches, such as the split-Cre strategy[23], to improve the cell-type or lineage specificity. Importantly, current tests for accuracy of Cre transgenic lines are based on their combination with reporter alleles and the analysis of resulting double transgenic animals by flow cytometry.

As in the case of *Cx3cr1*^{Cre} animals reported here, or *LysM*^{Cre} mice, in which expression was assumed to be myeloid cell-specific but found to also target neurons[10], this approach is insufficient. Our results demonstrate that the RiboTag approach provides a useful complementary method of determining Cre line specificity, in particular for cell types such as neurons and endothelial cells that are notoriously difficult to isolate for flow cytometric analysis.

Macrophage expression signatures tend to be contaminated by material the phagocytes ingested from their surroundings. Since the RiboTag strategy retrieves only mRNAs that are associated with HA-epitope-tagged host cell ribosomes, it excludes such exogenous material and hence allows identification of bona fide macrophage mRNAs.

Although not directly addressed in this study, an additional benefit of the RiboTag approach is the fact that it can be used to determine the effect of conditional mutations. Specifically, Cre-mediated rearrangements result in parallel mutagenesis and induction of the *Rpl22*^{HA} allele. For instance this approach was used to define the impact of an *Mecp2* deficiency on macrophages in brown adipose tissue (BAT), comparing translatomes retrieved from BAT of *Cx3cr1*^{CreER}:*Rpl22*^{HA} and *Cx3cr1*^{CreER}:*Rpl22*^{HA}:*Mecp2*$^{fl/y}$ mice[24].

Ribosome-associated mRNAs, as retrieved by the RiboTag approach, are considered to reflect the translatome. Notably, though, the microglial response to LPS using an analogous experimental system showed that mRNAs induced by the challenge can be prevented from translation by binding of a splicing factor to their 3′ UTRs[25]. Like other emerging layers of post-transcriptional expression control, this mechanism, which was revealed by a combined transcriptome and peptidome analysis on immunoprecipitated ribosomes, requires further study.

Above we listed advantages of the RiboTag approach, but an inherent weakness of this protocol is its reliance, for confident assessment of gene expression in the targeted population, on an enrichment of the specific mRNA over the input; i.e., the whole-tissue extract. The RiboTag approach hence precludes statements about the expression of genes that are equally expressed in the target cells and the surrounding tissue. For the assessment of these coexpressed genes, sorting-based strategies might be superior, although these also bear caveats, as outlined below.

Side-by-side comparison of the translatomes isolated by IP from crude tissue extracts and from sorted microglial cells with whole-cell transcriptomes revealed shortcomings of the latter. First, we noted a prominent activation signature that is presumably introduced during the process of extracting cell from their tissue context. This artifact comprised proinflammatory genes, such as *Cd86*, *Tlr4* and *Tlr7*, and will have to be considered in microglia profiling studies. Importantly, this robust and reproducible artifact could not be discerned when control and test samples were processed similarly. Rather, we found that the isolation procedure had a differential impact on microglia retrieved from either challenged or unchallenged animals. Of note, artifacts such as the ones we report here that are introduced during the isolation procedure were shown to be significantly reduced when the transcription inhibitor actinomycin D was included during cell preparation[26].

Taken together, our study shows that cell isolation coupled with sorting-based methods and the RiboTag approach each have strengths and weaknesses, which should be considered when designing experiments and drawing conclusions. Cell isolation bears the risk of artifacts that might significantly confound transcriptome-based studies, including single-cell analysis. Our study should hence caution experimentalists and make them aware of the 'observer effect', which is well established in physics but often less appreciated in biology.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41590-018-0110-6.

Received: 24 December 2017; Accepted: 5 April 2018;
Published online: 18 May 2018

References

1. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
2. Gosselin, D. et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
3. Heiman, M. et al. A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
4. Sanz, E. et al. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl Acad. Sci. USA* **106**, 13939–13944 (2009).
5. Yona, S. et al. Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* **38**, 79–91 (2013).
6. Jung, S. et al. Analysis of fractalkine receptor CX_3CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol. Cell. Biol.* **20**, 4106–4114 (2000).
7. Davalos, D. et al. ATP mediates rapid microglial response to local brain injury in vivo. *Nat. Neurosci.* **8**, 752–758 (2005).
8. Goldmann, T. et al. A new type of microglia gene targeting shows TAK1 to be pivotal in CNS autoimmune inflammation. *Nat. Neurosci.* **16**, 1618–1626 (2013).
9. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
10. Orthgiess, J. et al. Neurons exhibit Lyz2 promoter activity in vivo: Implications for using LysM-Cre mice in myeloid cell research. *Eur. J. Immunol.* **46**, 1529–1532 (2016).
11. Kim, K.-W. et al. In vivo structure/function and expression analysis of the CX3C chemokine fractalkine. *Blood* **118**, e156–e167 (2011).
12. Fonseca, M. I. et al. Cell-specific deletion of C1qa identifies microglia as the dominant source of C1q in mouse brain. *J. Neuroinflammation* **14**, 48 (2017).
13. Goldmann, T. et al. Origin, fate and dynamics of macrophages at central nervous system interfaces. *Nat. Immunol.* **17**, 797–805 (2016).
14. Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
15. van den Brink, S. C. et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
16. A-Gonzalez, N. et al. Phagocytosis imprints heterogeneity in tissue-resident macrophages. *J. Exp. Med.* **214**, 1281–1296 (2017).
17. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
18. Braunschweig, U. et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
19. Khong, A. et al. The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell* **68**, 808–820.e5 (2017).
20. Hubstenberger, A. et al. P-body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell* **68**, 144–157.e5 (2017).
21. Tebaldi, T. et al. Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. *BMC Genomics* **13**, 220 (2012).
22. Bahar Halpern, K. et al. Nuclear retention of mRNA in mammalian tissues. *Cell Rep.* **13**, 2653–2662 (2015).
23. Hirrlinger, J. et al. Split-cre complementation indicates coincident activity of different genes in vivo. *PLoS One* **4**, e4286 (2009).
24. Wolf, Y. et al. Brown-adipose-tissue macrophages control tissue innervation and homeostatic energy expenditure. *Nat. Immunol.* **18**, 665–674 (2017).
25. Boutej, H. et al. Diverging mRNA and protein networks in activated microglia reveal SRSF3 suppresses translation of highly upregulated innate immune transcripts. *Cell Rep.* **21**, 3220–3233 (2017).
26. Wu, Y. E., Pan, L., Zuo, Y., Li, X. & Hong, W. Detecting activated cell populations using single-cell RNA-seq. *Neuron* **96**, 313–329.e6 (2017).
27. Tilgner, H. et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).

Acknowledgements

We thank all members of the Jung laboratory for discussion, the staff of the Weizmann Animal Facility, members of the FACS facility for expert advice, G. Friedlander for help with bioinformatics, and C. Glass for sharing sequencing data. S.J. was supported by the Israeli Science Foundation (887/11), an Infect-ERA grant, the European Research Council (Adv ERC 340345), and the Deutsche Forschungsgemeinschaft (CRC/TRR167 'NeuroMac'). I.B and M.G. were supported by the Deutsche Forschungsgemeinschaft DFG-SFB 1052/1: 'Obesity mechanisms' (projects A09 and B09).

Author contributions

Z.H. and S.J. conceived the project and designed the experiments; Z.H., A.V. and J.O. performed the experiments; L.C.-M. performed RNA-seq and E.D. analyzed the data. I.U. and B.Z. performed bioinformatics analysis. S.Y., A.S., D.V., S.B.-H., I.B. and M.G. advised on experiments; Z.H. and S.J. wrote the paper; S.J. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-018-0110-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.J.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Animals. Mice were maintained on a 12 h light/dark cycle, and food and water were provided ad libitum. All animals were on a C57BL/6JOlaHsd background, maintained in specific-pathogen-free conditions and handled according to protocols approved by the Weizmann Institute Animal Care Committee (IACUC), as per international guidelines. The strains used included *Cx3cr1*^{Cre} mice (JAX stock # 025524 B6J.B6N(Cg)-*Cx3cr1*$^{tm1.1(cre)Jung/J}$)[5], *Cx3cr1*^{CreER} mice (JAX stock # 020940 B6.129P2(C)-*Cx3cr1*$^{tm2.1(cre/ERT2)Jung/J}$)[5], (JAX stock # 011029 B6N.129-*Rpl22*$^{tm1.1Psam/J}$)[4] and Rosa26-YFP reporter mice[28]. The presented RiboTag data were generated with animals homozygous for the *Rpl22*^{HA} allele and heterozygous for the modified *Cx3cr1* alleles. Mice heterozygous for the *Rpl22*^{HA} allele yielded similar results (data not shown).

Microglia isolation protocols. Mice were anesthetized with Pental (1:2 in PBS) and were perfused with PBS. Brains were dissected, coarsely chopped and incubated for 20 min at 37 °C in 1 ml HBSS solution containing 2% BSA, 1 mg/ml Collagenase D (Sigma) and 50 µg/ml DNase1 (Sigma). In the middle of the incubation, homogenates were pipetted for further dissociation. Next the homogenates were filtered through a 150-µm mesh, washed with cold FACS buffer (2% FCS, 1 mM EDTA in PBS without Ca^{2+} or Mg^{2+}) and centrifuged at 2,200 r.p.m. (970*g*) at 4 °C for 5 min. For the enrichment of microglia, the cell pellet was resuspended with 3 ml of 40% Percoll solution and centrifuged at 2,200 r.p.m. (970*g*), no acceleration and braking, at room temperature for 15 min. Next, the cell pellet was resuspended, passed through 80-µm mesh, washed with 5 ml FACS buffer and centrifuged at 1,400 r.p.m. (400*g*) at 4 °C for 5 min, followed by antibody labeling and flow cytometry analysis. For the protocol excluding the digestion, brains were chopped and then filtered through a 150-µm mesh. The subsequent steps were as above, but without the enzymatic digestion.

Tamoxifen (TAM) treatment. To induce gene recombination in CreER transgenic mice, tamoxifen (TAM) was dissolved in warm corn oil (Sigma) and administered orally via gavage for four times every other day. All animals were TAM-treated first at 4–6 weeks of age. Each oral application consisted of 5 mg at a concentration of 0.1 mg/µl. Mice were examined 8 weeks after treatment. For the LPS treatment, mice were injected intraperitoneally (i.p.) with a single dose of LPS (2.5 mg/kg; *E. coli* 0111:B4; Sigma); controls received or the same volume of vehicle solution (PBS).

Flow cytometry and cell sorting. Antibodies against CD11b (M1/70, Biolegend cat. no. 101205, RRID: AB_312788), Ly6C/G (Gr-1) (RB6-8C5, Biolegend cat no. 108411, AB_313376), CD45 (30-F11, Biolegend cat. no. 103112, RRID: AB_312976) were used. Samples were flow sorted using an AriaIII (BD Biosciences, BD Diva Software) cell sorter. Analysis was performed on a Fortessa (BD Biosciences, BD Diva Software) and analyzed with FlowJo software (Treestar).

Histology. Mice were anesthetized with Pental (1:2) and were perfused with PBS. Brains and spinal cords were excised and fixed for 4 h in 2% paraformaldehyde. Brains were incubated for 72 h in 30% sucrose and spinal cords in 18% EDTA before OCT (TissueTek) imbedding and freezing. Postfixed and stained 15–30 µm frozen sections were blocked in 2% horse serum for 2 h and incubated overnight at 4 °C with the primary antibody. Sections were washed three times in PBSX1 0.02% TritonX (Sigma) and exposed to secondary antibody for 1 h at 4 °C. Before covering, samples were washed three times and incubated for 5 min with Hoechst. Sections were analyzed with an Olympus BX51 confocal laser scanning microscope. Image acquisition was processed by Olympus image browser software. The following primary antibodies were used: rat monoclonal anti-HA (1:100, Sigma, cat. no. 12158167001 ROCHE), rabbit polyclonal anti-Iba1 (1:250, Wako, cat. no. 019-19741), mouse polyclonal anti-NeuN (1:100, Millipore, cat. no. ABN90) and Hoechst 33342 (1:10,000, Invitrogen). The following secondary antibodies were used: donkey anti-rat IgG (H+L) Cy2 cat. no. 712-225-153 and donkey anti-rat IgG (H+L) Cy3 cat. no. 712-165-153; donkey anti-rabbit IgG (H+L) Cy2 cat no. 711-225-152 and donkey anti-rabbit IgG (H+L) Cy3 cat. no. 711-165-152; donkey anti-mouse IgG (H+L) Cy5 cat. no. 715-175-150.

Immunofluorescence. Mouse brains were fixed in 4% paraformaldehyde and subsequently paraffin-embedded by standard protocols. AT sections (7 µm thick) were deparaffinized in xylene and rehydrated through descending grades of ethanol to water. Antigen retrieval was performed in Tris/EDTA buffer at pH 9.5 twice for 5 min at 95 °C. Sections were rinsed in PBS supplemented with 0.3% Triton-X (Sigma-Aldrich, St Louis, Missouri, USA) three times for 5 min. Nonspecific binding sites were blocked using 1% bovine serum albumin in PBS supplemented with 0.3% Triton-X for 30 min at room temperature. The primary antibodies chicken anti-GFP (1:100; Abcam, cat. no. ab13970) and rabbit anti-Iba1 (1:200; Wako; Richmond, VA, cat. no. 019-19741) were incubated overnight at 4 °C. To detect the primary antibodies, AlexaFluor 488 goat anti-chicken (cat. no. A11039) and Alexa Fluor 568 goat anti-rabbit (cat. no. A11036; 1:200 each; all from Invitrogen; Karlsruhe, Germany) were applied for 1 h at room temperature. Autofluorescence of the tissue was quenched by using prewarmed 0.3% Sudan black for 2 min. Nuclear counterstain was performed by using 4′,6-diamidino-2-phenylindole (DAPI; 1:10,000 in PBS) for 5 min, followed by three buffer rinses. Finally, sections were embedded with Dako immunofluorescence mounting medium (Dako, Glostrup, Denmark). Images were taken using a FV1000 confocal laser scanning microscope (Olympus, Hamburg, Germany).

Ribosome immunoprecipitation (IP). Samples were extracted from mice, flash-frozen in liquid nitrogen and stored at −80 °C until use. Samples were homogenized on ice in ice-cold homogenization buffer (50 mM Tris, pH 7.4, 100 mM KCl, 12 mM $MgCl_2$, 1% NP-40, 1 mM DTT, 1:100 protease inhibitor (Sigma), 200 units/ml RNasin (Promega) and 0.1 mg/ml cycloheximide (Sigma) in RNase free DDW) 10% w/v with a Dounce homogenizer (Sigma) until the suspension was homogeneous. To remove cell debris, 1 ml of the homogenate was transferred to a microcentrifuge tube and centrifuged at 10,000 *g* at 4 °C for 10 min. Supernatants were transferred to a fresh microcentrifuge tube on ice, and then 10 µl was removed for input fraction analysis and 5 µl (125 µg) of anti-HA antibody (H9658, Sigma) or 5 µl (1 µg) of mouse monoclonal IgG1 antibody (Sigma, Cat# M5284) was added to the supernatant, followed by 4 h of incubation with slow rotation in a cold room at 4 °C. (Subsequent further calibration of the amounts of the antibodies showed that 5 µg of anti-HA antibody and mouse monoclonal IgG1 antibody (Merck, Cat# PP100) yielded similar results, including transcripts enriched in the IgG IP samples seen in cluster IV of Fig. 2b (data not shown)). Meanwhile, Dynabeads Protein G (Thermo Fisher Scientific), 100 µl per sample, were equilibrated to homogenization buffer by washing three times. At the end of 4 h of incubation with antibody, beads were added to each sample, followed by incubation overnight at 4 °C. After not more than 12 h, samples were washed three times with high-salt buffer (50 mM Tris, 300 mM KCl, 12 mM $MgCl_2$, 1% NP-40, 1 mM DTT, 1:200 protease inhibitor, 100 units/ml RNasin and 0.1 mg/ml cycloheximide in RNase free DDW), 5 min per wash in a cold room on a rotator. At the end of the washes, beads were magnetized and excess buffer was removed, 150 µl lysis buffer was added to the beads and RNA was extracted with Dynabeads mRNA Direct purification kit (Thermo Fisher). RNA was eluted in 6 µl H_2O and taken for RNA-seq. For Sort-IP samples (RiboTag IP after sorting), ~50×10^3–100×10^3 cells were sorted into cold PBS and centrifuged at 400 *g* for 5 min at 4 °C. Supernatant was removed and the pellet was resuspended in 1 ml of lysis buffer, and then the rest of IP was continued as above.

RNA sequencing. RNA-seq of populations was performed as described previously[1]. In brief, 5,000 microglial cells were sorted into 50 µl of lysis buffer (Life Technologies) and stored at −80 °C. mRNA was captured with Dynabeads oligo(dT) (Life Technologies) according to the manufacturer's guidelines. We used a derivation of MARS-seq[29]. Library concentration was measured with a Qubit fluorometer (Life Technologies) and mean molecule size was determined with a 2200 TapeStation instrument. RNA-seq libraries were sequenced using Illumina NextSeq-500.

Data analysis. Raw reads were mapped to the genome (NCBI37/mm9) using hisat (version 0.1.6). Only reads with unique mapping were considered for further analysis. Gene expression levels were calculated using the HOMER software package (analyzeRepeats.pl rna mm9 -d < tagDir > -count exons -condenseGenes -strand + -raw)[30]. Normalization and differential expression analysis was done using the DESeq2 R package (Bioconductor, https://bioconductor.org/packages/release/bioc/html/DESeq2.html). Differential expressed genes were selected using a twofold change cutoff between at least two populations and $P < 0.05$ adjusted for multiple gene testing. Gene expression matrix was clustered using a *k*-means algorithm (Matlab function kmeans) with correlation as the distance metric. Heat maps were generated using Genee software.

Analysis of gene features. For comparison of features between genes in cluster II-b and other genes, we used gene models from GENCODE vM13. Splicing efficiency was computed as by Tilgner and colleagues[27] using poly(A)$^+$ RNA-seq data from microglia[2]. Cytoplasmic/nuclear expression levels in liver and MIN6 cells were obtained using RNA-seq data from Bahar Halpern and colleagues[22] quantified using RSEM with GENCODE vM13 gene models. Cytoplasmic/nuclear ratios were computed using DESeq2.

Statistical analysis. In all experiments, data are presented as mean ± s.d., if not stated otherwise. Statistical tests were selected based on appropriate assumptions with respect to data distribution and variance characteristics. Statistical significance was defined as $P < 0.05$. Sample sizes were chosen according to standard guidelines. Number of animals is indicated as *n*. Of note, the sizes of the tested animal groups were also dictated by availability of the transgenic strains and litter sizes, allowing littermate controls. Pre-established exclusion criteria were based on IACUC guidelines. Animals of the same age, sex and genetic background were randomly assigned to treatment groups. The investigator was not blind to the mouse group allocation, although tested samples were assayed blindly.

Accession codes. The accession codes for the RNA-seq datasets reported in this paper can be found at GEO: GSE114001.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data that support the findings of this study are available as Source Data for Figs. 1, 2, 3a, 4a and 5a,c and Supplementary Figs. 5a and 6a. Other data are available from the corresponding author upon reasonable request.

References

28. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
29. Jaitin, D. A. et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
30. 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).

nature research

Corresponding author(s): Steffen Jung , NI-A25510D

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

sample sizes were chosen according to guidelines as published in NATURE METHODS | VOL.10 NO.12 . Of note, sizes of the tested animal groups were also dictated by availability of the transgenic strains and litter sizes, allowing littermate controls.

2. Data exclusions

Describe any data exclusions.

no data were excluded from analysis

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

Experiments were replicated at least once and included repeats

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Animals of the same age, sex and genetic background were randomly assigned to treatment groups.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The investigator was not blinded to the mouse group allocation. Tested samples were blindly assayed.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Gene expression levels were calculated using the HOMER software package. Normalization and differential expression analysis was done using the DESeq2 R-package.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.	no restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	IP anti-HA antibody (H9658, Sigma); Flow cytometry Antibodies against CD11b (M1/70), Ly6C/G (Gr-1) (RB6-8C5), CD45 (30-F11) purchased from Biolegend or eBioscience; Histology: Rat anti HA, mouse anti NeuN (MAB 377, Milipore),Rabbit anti IBA-1 (Wako).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	NA
b. Describe the method of cell line authentication used.	NA
c. Report whether the cell lines were tested for mycoplasma contamination.	NA
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	NA

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.	JAX stock # 025524 B6J.B6N(Cg)-Cx3cr1tm1.1(cre)Jung/J 5 JAX stock # 020940 B6.129P2(C)-Cx3cr1tm2.1(cre/ERT2)Jung/J) 5 JAX stock # 011029 B6N.129-Rpl22tm1.1Psam/J)4 Rosa26-YFP reporter mice (Srinivas et al., 2001).

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	NA