

## Article

# A Core Transcription Regulatory Circuitry Defining Microglia Cell Identity Inferred from the Reanalysis of Multiple Human Microglia Differentiation Protocols

Antoine Aubert, François Stüder, Bruno Maria Colombo and Marco Antonio Mendoza-Parra \*

Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France; Antoine.AUBERT@genoscope.cns.fr (A.A.); fstuder@genoscope.cns.fr (F.S.); bruno.colombo@univ-evry.fr (B.M.C.)

\* Correspondence: mmendoza@genoscope.cns.fr; Tel.: +33-01-60-76-42-92

**Abstract:** Microglia, the immune cells in the brain involved in both homeostasis and injury/infection control, play a predominant role in neurodegenerative diseases. In vivo studies on microglia are limited due to the requirement of surgical intervention, which can lead to the destruction of the tissues. Over the last few years, multiple protocols—presenting a variety of strategies—have described microglia differentiation issued from human pluripotent stem cells. Herein, we have reanalyzed the transcriptomes released on six different microglia differentiation protocols and revealed a consensus core of master transcription regulatory circuitry defining microglia identity. Furthermore, we have discussed the major divergencies among the studied protocols and have provided suggestions to further enhance microglia differentiation assays.

**Keywords:** microglia differentiation; transcriptomics; gene regulatory networks; systems biology



**Citation:** Aubert, A.; Stüder, F.; Colombo, B.M.; Mendoza-Parra, M.A. A Core Transcription Regulatory Circuitry Defining Microglia Cell Identity Inferred from the Reanalysis of Multiple Human Microglia Differentiation Protocols. *Brain Sci.* **2021**, *11*, 1338. <https://doi.org/10.3390/brainsci11101338>

Academic Editors: Helene Hirbec and Florence E. Perrin

Received: 13 September 2021

Accepted: 4 October 2021

Published: 11 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Microglia represent about 10% of the cells in the brain. They are localized in the parenchyma and are the most representative immune cells of the central nervous system (CNS) [1]. Microglia derive during early embryogenesis from yolk sac–primitive macrophages and begin to colonize the brain before the emergence of neurons and macroglia. These primitive myeloid progenitors are highly proliferated throughout embryonic life and constitute the resident adult microglia in a healthy brain. These cells persist during adulthood via constant self-renewal. Nevertheless, postnatal hematopoietic progenitors also contribute to the adult microglial pool and are constantly renewed throughout life [2–4].

Microglia are highly dynamic cells that constantly monitor the brain microenvironment. These cells guarantee cell homeostasis by ensuring tissue repair. Once stimulated, microglial cells rapidly proliferate, change morphology, and activate and secrete various immunomodulatory factors (proinflammatory cytokines) [5–7].

Thanks to their ability to scan the microenvironment in the brain, microglia are also involved in the recognition and the elimination of pathogens invading the CNS [8–10] and are expected to play a predominant role in the context of neurodegenerative diseases, including Alzheimer's disease (AD) [11–15], Parkinson's disease (PD) [16], and progressive supranuclear palsy (PSP), which corresponds to a common type of atypical parkinsonism that is based on a four-repeat tauopathy pathology [17,18]. Concerning AD, there is growing evidence that microglia protect against the incidence of this illness, as suggested by the observation of their accumulation around  $\beta$ -amyloid plaques. Furthermore, recent genome-wide association studies (GWAS) have revealed an astonishing accumulation of AD risk loci near or within genes known to be expressed or that are even sometimes specifically in microglia, which suggests a genetic dimension that directly influences microglia behavior in the context of AD [19]. Similarly, recent studies have suggested that

human endogenous retroviruses (HERV) might contribute to microglia activation, in both neurodegenerative diseases as well as in neuropsychological disorders, by regulating the expression of inducible NOS [17,20,21].

Motivated by the necessity of studying human microglia behavior in a normal setting as well as in the context of neurodegenerative diseases, over the last six years, multiple protocols described microglia differentiation strategies issued from human pluripotent stem cells [22]. In most of these studies, the obtained iPSC-derived microglia-like cells (iMGL) were validated by phenotypical assays as well as by the use of comprehensive transcriptome readouts (microarrays [23,24] or RNA-sequencing [25–30]), allowing for the investigation of the presence of the signature genes of microglia. While the authors of these studies systematically found relevant microglia-related expressed genes, a systematic comparison between all of the transcriptomes assessed within microglia differentiation studies is missing.

In this article, we provide a detailed comparative study—covering six different protocols—evaluating (i) the quality of the generated transcriptome data; (ii) their relevant gene expression differences; (iii) the presence of a consensus gene regulatory program among the compared studies as well as their particularities in the context of master transcription factors.

Overall, this comparative study reveals a core gene regulatory program defining the microglia state and provides evidence for the differences observed between the studied protocols. Notably, it provides new avenues to explore master transcription factor over-expression-driven strategies for enhancing microglia differentiation methods, which are supported by the fact that master transcription factors possess the capacity to trigger a developmental program outside of its normal context [31].

## 2. Materials & Methods

### 2.1. Public Transcriptomes

Datasets associated with the six protocols under study were collected from the Gene expression Omnibus (GEO) database. The corresponding articles and their associated GEO series accession number (GSE) are Abud et al. (2017) [26] (GSE97744); Muffat et al. (2016) [32] (GSE85839); Douvaras et al. (2017) [27] (GSE97744); McQuade et al. (2018) [26] (GSE117829); Trudler et al. (2021) [29] (GSE169065); and Chen et al. (2021) [30] (GSE163984).

### 2.2. Data Pre-Processing

Raw fastq files were aligned to the reference human genome GRCh38 using Bowtie2 (v2.4.1). Aligned datasets were qualified for read count saturation using the next generation sequencing quality control generator (NGS-QC Generator) [33,34]. Furthermore, read counts associated with annotated transcribing regions were assessed with the tool featureCounts (v2.0.1) [35] and were assembled within a matrix composed by genes in rows and all datasets (issued from all protocols) in columns. The raw matrix was cleaned by removing genes devoid of read counts, and then one read was added over the whole matrix to avoid division by zero errors.

Finally, technical amplitude differences between compared datasets were corrected by applying a quantile normalization procedure. Data sets were regrouped by cell type and genetic background into subgroup or replicate datasets in each protocol. Each replicate dataset was associated with a number. All of the datasets used in this study are summarized in Supplementary Table S1.

### 2.3. Principal Component Analysis

A principal component analysis (PCA) was performed on the log<sub>10</sub> value quantile-normalized counts. Specifically, PCA was performed on genes that appeared as upregulated on in vivo fetal microglia provided by Abud et al. (6319 genes). The three-dimensional PCA analysis was plotted using the “plot3D” function of the rgl package.

#### 2.4. Differential Gene Expression Analysis

Differential gene expression analysis between samples was performed using Deseq2 (v.1.32.0) on the quantile normalized count matrix. Datasets associated with a given protocol were compared to their corresponding iPSC readouts when available. When the iPSC data were not provided (Muffat et al. (2016), Douvaras et al. (2017), and McQuade et al. (2018) data sets), the iPSC datasets provided by Abud et al. (2017) were used as a reference for the differential expression analysis. Differentially expressed genes were considered for fold-changes > 2 and presenting an adjusted  $p$ -value ( $\text{padj}$ ) < 0.01 (Supplementary Table S2).

To determine, the composition of the microglia identity of each iMGL population, upregulated genes were compared to their in vivo counterpart (fetal and adult) to determine the percentage of commonly up regulated genes between each iMGL population compared to two in vivo controls (fetal and adult microglia from Abud et al.). This value of similarity was tempered by the divergence between the in vivo microglia and the iMGL (the percentage of up regulated genes in iMGL that were not differently expressed in their in vivo counterpart). This value allowed us to take the variability of the number of up regulated genes between samples into account.

$$\%Commun\ genes = \frac{\text{Number of comun upregulated genes}}{\text{Number of upregulated genes (MGL)}} \times 100$$

$$\%Divergence = \left(1 - \frac{\text{Number of comun upregulated genes}}{\text{Number of upregulated genes (iMGL)}}\right) \times 100$$

Microglia-specific markers were defined by selecting upregulated genes that were present in the fetal microglia dataset (6154 genes) but that were absent from all other control samples (CD14+/CD16- monocytes, CD14+/CD16+ inflammatory monocytes, myeloid dendritic cells, neuronal precursors, and neuronal-related datasets). This Microglia-specific marker subset (201 genes; Supplementary Table S3) was further confirmed by gene ontology analysis using ENRICH [36] and was used to profile gene expression changes within the iMGL samples. In each case, datasets were clustered by performing Euclidian biclustering. Heatmaps were plotted using the “pheatmap” package on R.

#### 2.5. Reconstruction of Gene Co-Regulatory Networks

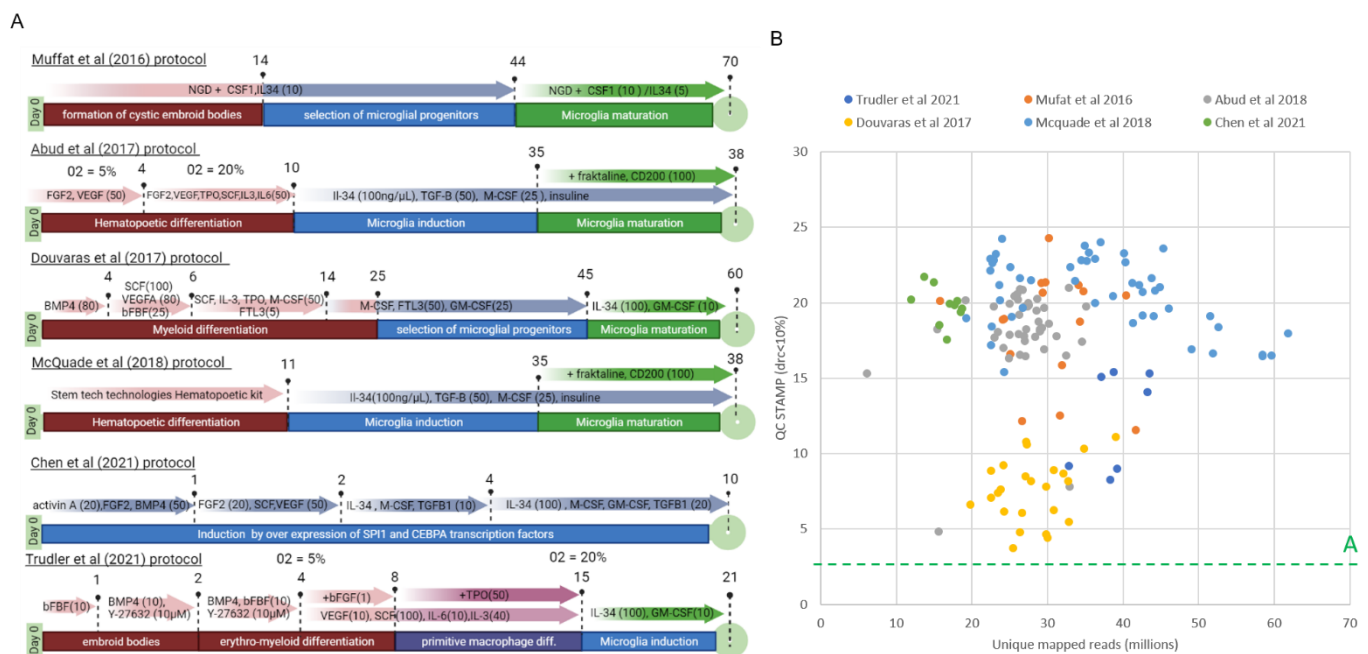
Differential gene expression readouts (differential log<sub>2</sub> fold-change levels higher than 2; smaller than −2) associated with each of the studied protocols were integrated with transcription factor-target gene relationships inferred from human CAGE readouts assessed by the FANTOM consortium covering more than 300 human cell/tissue types [37]. This integrative processing was performed within TETRAMEER, which also allowed us to infer master regulator TFs by modelling transcription signaling regulation propagation [38]. TETRAMEER allowed us to rank TFs on the basis of their capacity to regulate down-stream target genes that are associated with each of the studied protocols. Ranked TF-TG co-regulatory networks issued from all six protocols were integrated into a single network. Gene ontology analysis on the common master regulatory TFs was performed using ENRICH [36].

Inferred consensus master transcription factors were compared with master players described in three further studies: Gosselin et al. [39], which was based on in vivo microglia isolated from surgically resected brain tissue and that had been cultured in an in vitro environment; Banerjee et al. [40], who described a protocol for iMGL generation from peripheral monocyte samples; and Grubman et al. [41], who described an enhanced maturation of iMGL precursors when co-cultured with human neuronal precursors (ReN) (Supplementary Figure S4). Specifically, the raw read counts provided by Grubman et al. (GSE125872) were used to query for genes being over-expressed in the iMG + ReN conditions relative to their corresponding iPSC sample. Similarly, the microglial signature genes provided by Banerjee et al. in their Supplementary Table S2 were used for this comparison. Finally, master transcription factors associated with the study of Gosselin et al. were collected from Supplementary Table S2 retrieved in Banerjee et al. [40] as well as from the information provided on the corresponding publication.

### 3. Results

#### 3.1. Comparative Reanalysis of Whole Transcriptome Readouts Generated within Six iPSC-Derived Microglia (iMGL) Differentiation Studies

Concerned by the variety of strategies proposed for generating microglia cells from human iPSC differentiation and their potential similarities and discrepancies on their performance, we have focused our attention on six protocols, namely the 70-day long protocol provided by Muffat and co-authors [32]; the protocol proposed by Abud and co-authors that based on a hematopoietic intermediate but that also required hypoxic conditions [25] as well as its improved version relying on a commercial solution issued from a stem cell technologies company [26]; the protocol proposed by Douvaras and colleagues that is based on an initial myeloid differentiation [27]; the 21-day protocol proposed by Trudler et al. that mimics normal yolk sac and erythromyeloid embryonic development [29]; and the extremely fast protocol (10 days) proposed by Chen et al., which is focused on the forced expression of the master transcription factors SPI1 (SFFV pro-viral integration 1) and CEBPA (CCAAT Enhancer Binding Protein Alpha) [30] (Figure 1A).



**Figure 1.** Overview of the iMGL protocols and the quality of their RNA-Seq libraries. **(A)** Short summary of the human iMGL differentiation protocols compared in this study. Protocols are organized by chronological order of publication in the literature. The different steps of the cell differentiation were highlighted and associated with the corresponding treatment (when not specified the unit's concentration is ng/μL). Each time point corresponds either to a key step in the cellular differentiation or a change in medium composition. The O<sub>2</sub>% in the medium was also added when modified during the protocol. When not specified, the O<sub>2</sub> concentration is considered to be normoxia (O<sub>2</sub> = 20%). **(B)** Scatter plot of the quality of each RNA-Seq libraries associated with the compared protocols as assessed by the NGS-QC Generator tool compared to their corresponding unique mapped sequenced reads. The green dashed line demarcates the quality "A" threshold assessed by the NGS-QC Generator database, issued from the analysis of more than 90 thousand public datasets (<http://ngs-qc.org/> accessed on 1 October 2021).

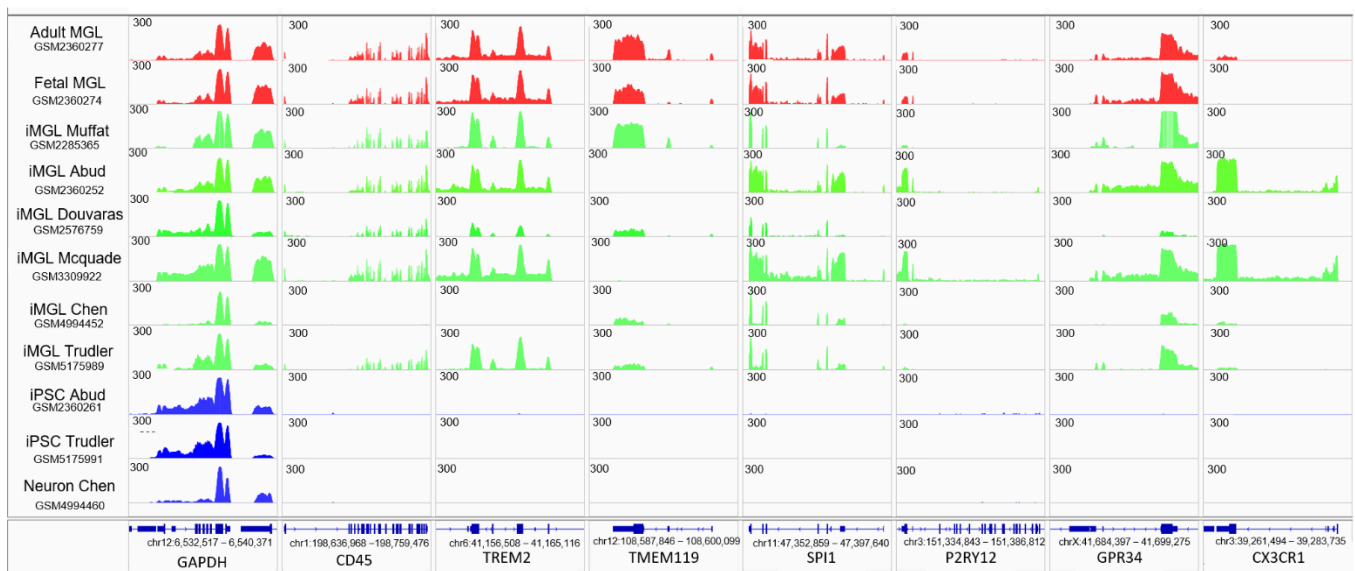
Each of these studies generated whole transcriptome readouts associated with the iMGL state, their hematopoietic intermediates (HPC) when available, their related iPSC precursors as well as positive controls issued from fetal and adult microglia cells, and negative control samples corresponding to CD14+/CD16− monocytes (CD14M), CD14+/CD16+ inflammatory monocytes (CD16M), myeloid dendritic cells (Blood DCs), neuronal precursors (NPC), and neuronal transcriptome data. Overall, as part of this comparative analysis,

we collected 148 transcriptomes collected from the public gene expression omnibus (GEO) repository and reprocessed them with the help of our in-house qcGenomics pipeline [33].

As part of the primary analysis, we evaluated the quality of the collected datasets. This was performed using our previously described QC STAMP-scoring methodology, which assesses the sequencing-depth saturation point by comparing the level of read-count pattern reproducibility between the original RNA-seq profile and those where only a subset of the total mapped reads are used for reconstruction [34,42]. As illustrated on Figure 1B, most of the reanalyzed datasets present QC STAMP scores comprised between 15 and 25 despite of their mapped reads, while the transcriptomes generated by Douvaras and colleagues present a relatively lower quality scoring level. In all cases, the assessed QC STAMP scores are within the highest levels (quality A) estimated over more than 90,000 public datasets (see our database <http://ngs-qc.org/database.php> (accessed on 1 October 2021) and Supplementary Figure S1), indicating that the processed RNA-seq readouts are of high quality.

### 3.2. iMGL Transcriptome Datasets Issued from Different Protocols Present Variable Degree of Similarity with Control Microglia Samples

With the aim of evaluating the microglia differentiation performance between the compared protocols, we have first evaluated the gene expression levels of well-known microglia markers within the analyzed transcriptomes. As illustrated on Figure 2, significant gene expression is observed for the G protein-coupled receptor GPR34 and the transcription factor SPI1 in all iMGL transcriptomes, which is in agreement with the transcription levels observed on the fetal and adult microglia readouts. On the contrary, expression differences are observed for the gene CD45 and TREM2, where no signals were observed in datasets generated by Chen et al. Similarly, TMEM119 is devoid of readouts on the McQuade and Abud's iMGL datasets; P2RY12 do not present read counts on Muffat, Douvaras, Chen and Trudler's iMGL datasets, and the expression of the CX3CR1 gene can be seen in the Abud and McQuade iMGL data.

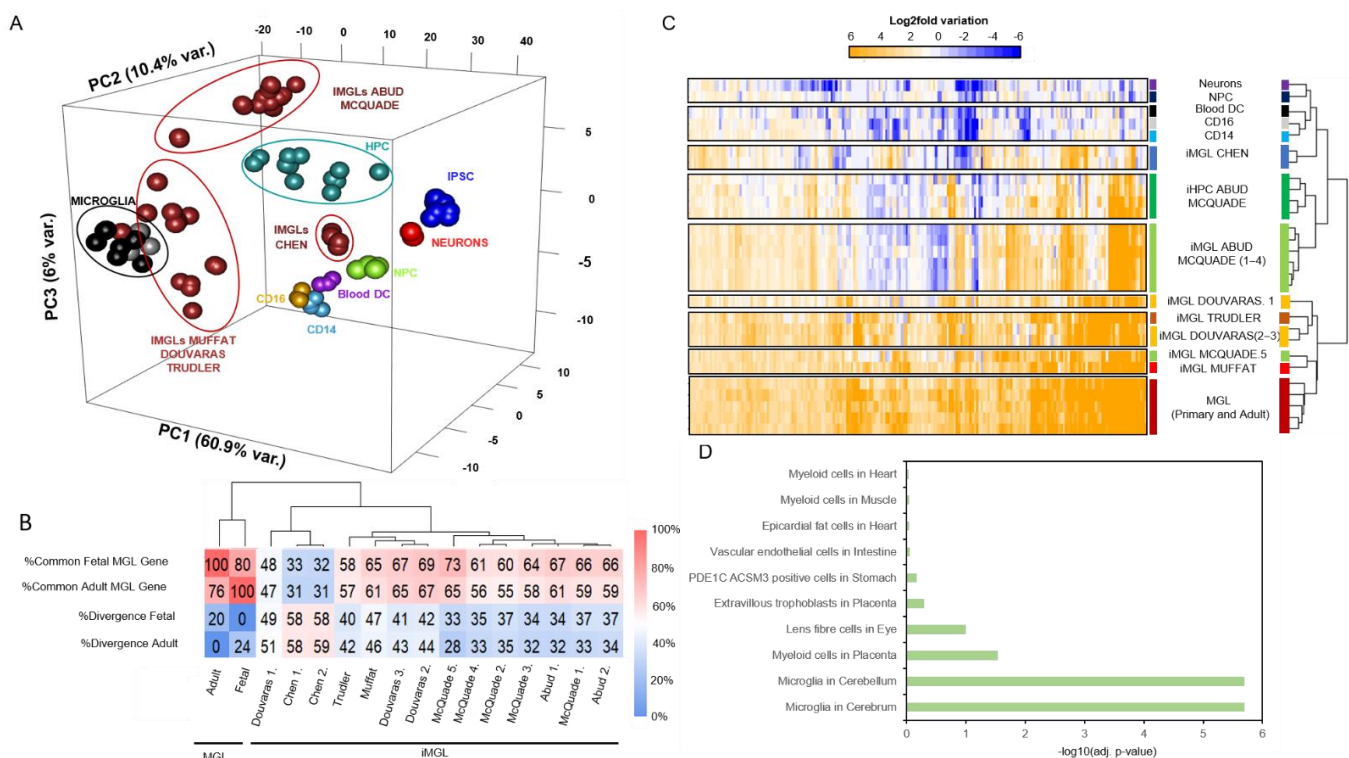


**Figure 2.** RNA-sequencing read coverage on microglial specific gene loci. Datasets corresponding to control in vivo microglia samples (adult and fetal cells) are shown in red, induced microglia-like (iMGL) datasets associated with each of the studied protocols are shown in green, and a composed negative control corresponding to the iPSC and neurons is shown in blue. To enhance their comparability, the read count amplitude for all samples has been set to 300.

These major differences observed on a small subset of genes strongly support the need for large scale analyses able to decrease potential technical differences and to highlight the relevance of each of these protocols for generating iMGL cells. For this, a principal



component analysis (PCA) was used to compare all of the iMGL datasets as well as the various positive and negative control samples (Figure 3A). The use of a 3-dimensional PCA representation (PC1: 60.9%; PC2: 10.4%; PC3: 6% variance, respectively) allowed us to reveal the similarities between the microglia control datasets and the iMGL samples issued from Muffat, Douvaras, and Troudler's protocols, which are located far away from the human iPS, neuronal precursors (NPCs), and neurons as well as the hematopoietic datasets (CD14, CD16, Blood DC). Surprisingly the iMGLs issued from Chen's protocol appeared distant from the microglia control datasets and rather proximal to the hematopoietic precursor (HPC) and hematopoietic datasets. Similarly, most of the iMGLs issued from Abud and McQuade's protocols appeared to be preferentially distal from microglia control samples, proximal to HPCs, but distant from hematopoietic samples, suggesting that they are more similar to microglia than hematopoietic samples.



**Figure 3.** Differences and similarities between the iMGL gene expression profiles issued from six differentiation protocols. (A) Three-dimensional principal component analysis (3D PCA) showing iMGL in dark red, in vivo microglia (fetal and adult microglial cells in black and grey, respectively), induced hematopoietic progenitor cells (iHPC) in azure, CD14+/CD16– quiescent monocyte in green, CD14+/CD16+ inflammatory monocytes in yellow, blood dendritic cells in violet, neurons in red, and iPSC in blue. The first principal component (PC1 60.9% variance) represents the hematopoietic differentiation process, principal component 2 (PC2 10.4% variance) allows the separation of the in vivo microglia cells and the iMGL datasets, and principal component 3 (PC3 6% variance) allows the discrimination between the different monocytic population trajectories. The PCA graph reveals the stratification of all iMGL samples on three major clusters. The first cluster is composed by iMGL samples issued from protocols developed by Muffat et al., Douvaras et al., and Troudler et al. and presents a striking similarity with the in vivo microglia control samples. The second cluster, separated from the first cluster because of PC2, is composed by iMGL samples issued from protocols developed by Abud et al. and McQuade et al. Finally, the third cluster present in the monocytic region of the graph presented the iMGL samples generated by Chen et al. (B) Similarity and divergency matrix (in percent) for each of the iMGL datasets compared to the control microglia samples. This analysis is based on the upregulated genes (2-fold change; adjusted  $p$ -value < 0.01) assessed in each of the iMGL samples relative to the corresponding induced pluripotent stem cell (iPSC) precursor samples. (C) Gene expression levels (log2 fold-change, Euclidian clustered heatmap) assessed on iMGL and the corresponding control samples on a subset of 201 microglia-specific genes. (D) Cell ontology of the 201 microglia-specific genes.

To further evaluate the potential differences between the transcriptomes associated with the various iMGL samples, a differential expression analysis was performed relative to their corresponding induced pluripotent stem cell precursors (Supplementary Figure S2 and Table S2). The analysis of the number of upregulated genes (2-fold change; adjusted  $p$ -value < 0.01) demonstrated > 55% of commonalities between most of the iMGL replicates issued from McQuade, Muffat, Douvaras, and Trudler's protocols and the microglia control samples (Figure 3B). Indeed, only one of the iMGL replicates issued from the Douvaras' protocol shared less than 50% of their upregulated genes with the microglial control samples (fetal and adult). Surprisingly, the iMGL samples issued from Chen's protocol only shared ~30% of their upregulated genes with the microglia control samples, which is in agreement with the low similarity observed in the PCA analysis. Furthermore, the fraction of upregulated genes that diverge relative to either the fetal or adult microglia samples revealed that iMGL samples issued from Chen's protocol presented the highest discrepancies (~58%).

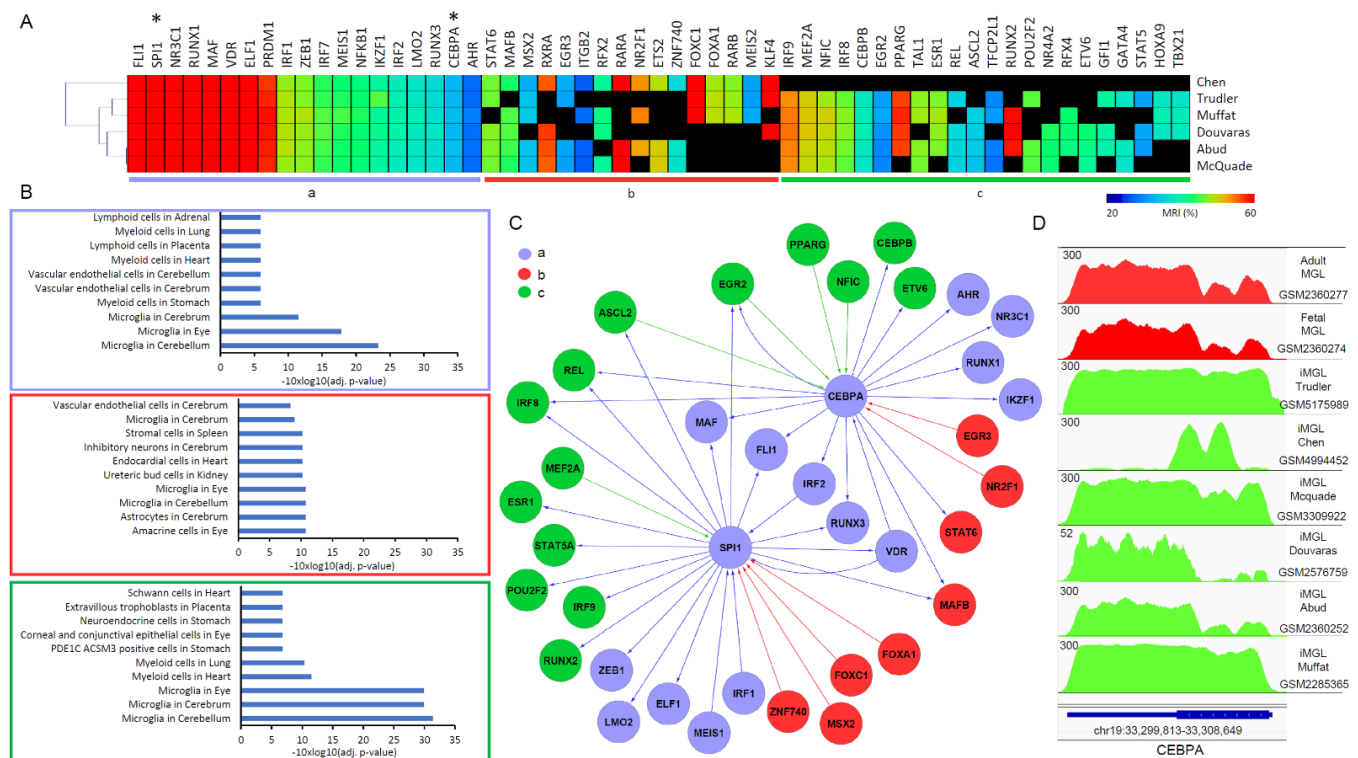
To take in consideration the fold change levels when performing these comparative analyses, we first selected the upregulated genes within the microglia control samples and then subtracted the upregulated gene commonly expressed in negative control samples (CD14M, CD16M, Blood DC, NPC and Neurons) (Supplementary Table S1). This set of 201 genes also appeared to be strongly upregulated (>4-fold change) in the iMGL samples issued from Muffat, Douvaras, and Trudler's protocols, where ~70% to 80% of the microglial genes were upregulated (Figure 3C and Supplementary Table S3). The iMGL samples issued from Abud and McQuade's protocols presented 36% upregulated genes, while the iMGL samples issued from Chen's protocol only presented 23% of these selected microglia-specific upregulated genes. Remarkably, one of the iMGL replicate datasets issued from the protocol described by McQuade et al. (McQuade.5) showed a strong similarity with the fetal and adult microglia control samples. On the contrary, one of the replicates generated by Douvaras et al. (Douvaras.1) showed a lowered similarity to in vivo control samples (Figure 3C). A gene ontology analysis performed on the set of 201 genes confirmed their enrichment on microglia-related terms at a high confidence level (Figure 3D).

### 3.3. A Master TF Co-Regulatory Network Driving Microglia Differentiation Reveal Strengths and Weaknesses of the Compared Protocols

Microglia differentiation, when induced by whatever protocol, might reshape the gene regulatory programs defining stem cell towards a microglial state. To reconstitute the gene regulatory programs involved in microglia differentiation, we have integrated the differential iMGL gene expression readouts with TF-TG relationships issued from the integration of human CAGE readouts assessed by the FANTOM consortium covering more than 300 human cell/tissue types [37]. This integrative effort has been performed with TETRAMEER, our previously developed Cytoscape application [38], leading to the reconstruction of iMGL-related gene regulatory networks associated with each of the compared protocols. Considering that the reconstructed networks present TF-TG relationships issued from a variety of cell/tissue types, TETRAMEER applies an in silico transcription regulation signaling cascade initiated from each of the TF to conserve those coherent with the assessed gene expression information related to the microglial state. This in silico modeling strategy allows us to compute a master regulatory index for each of the TFs that is defined as the fraction of target genes (direct and indirect) that can be controlled by a given TF relative to the total number of nodes within the GRN (Supplementary Figure S3).

Assuming that all six protocols are expected to drive towards the same microglial state, we compiled all of the TFs presenting significant MRI scores within all of the protocols and selected those that are present on at least three of the six cases. This analysis led to 57 TFs, which were classified into three groups. The first group is composed of 19 TFs that were present on all six protocols. The second subset is composed of 22 TFs expressed on the transcriptomes issued from most of the protocols, except the one described by Chen and colleagues. Finally, the third group is composed of 16 TFs that were upregulated on transcriptomes issued from the protocol provided by Chen and colleagues as well as some

of the other protocols (Figure 4A). A gene ontology enrichment analysis performed for each of these groups revealed that only the master TFs retrieved on all of the protocols (first group) or those retrieved on protocols other than that using specific TFs-overexpression (second group) are significantly enriched for microglia-related terms (Figure 4B). This microglia-related TFs enrichment is also supported by a comparison with master players retrieved in three other studies, either issued from in vivo microglia collected from surgically resected brain tissue [39], from matured microglia due to their co-culturing with human neuronal precursors [41], or differentiated from peripheral monocytes [40] (Supplementary Figure S4).



**Figure 4.** In silico prediction of a consensus master transcription regulatory circuitry driving microglia differentiation. (A) Master transcription factors predicted from the reconstruction of gene regulatory networks (GRN) and conserved in at least three over six of the studied protocols. Heatmap corresponds to their master regulatory index (MRI), which corresponds to the fraction of downstream target players (direct or indirect) that can be controlled by the TF within the reconstructed GRN (in percent). Transcription factors were organized into three groups: (a) those that were retrieved in the datasets issued from all protocols; (b) those retrieved from datasets generated by Chen et al. and at least two other protocols; and (c) those retrieved from at least three protocols except Chen's protocol. The transcription factors SPI1 and CEBPA, which were over-expressed in the protocol described by Chen et al., are part of group (a), as highlighted by the “\*”. (B) Cell ontology analysis was performed for each of the TF groups described in (A). (C) SPI/CEBPA co-regulatory network displaying their first TF neighbors within the master players listed in (A). The color-code associated with each node corresponds to the groups described in (A). Similarly, edges are colored in agreement with the color of the source node. (D) RNA-sequencing read count enrichment at the CEBPA locus for iMGL datasets issued from the different protocols (green) as well as the fetal and adult control samples (red). Notice the shortened read count enrichment pattern observed for the sample associated with the protocol described by Chen et al. relative to all other datasets. Enrichment read count levels were set to 300 for their enhanced comparability.

Considering that the specific TF-overexpression protocol used by Chen and colleagues targets the factors CEBPA and SPI1, we focused our attention to their corresponding regulome. As illustrated in Figure 4A, both CEBPA and SPI1 are retrieved over-expressed in all of the protocols and present strong master regulatory indexes (SPI1: MRI > 70%; CEBPA: MRI > 30%). Their related TF co-regulatory network reveals an important number



of co-regulated master players, including MAF, FLI1, IRF2, RUNX3, and the cellular receptor for Vitamin D3, VDR; all of them retrieved among all six protocols (Figure 4C).

Interestingly, another group of SPI1/CEBPA co-regulated players, including IRF9, MEF2A, NFIC, IRF8, CEBPB, EGR2, REL, and ETV6, were found expressed in most protocols, with the exception of the one described by Chen and colleagues (Figure 4A,C). Specifically, the transcription factor IRF8 (interferon regulatory factor 8), known to be essential for proper microglia reactivity [43] as well as for its development [44], and the NF- $\kappa$ B family member REL (c-rel), known to be expressed in the brain and playing relevant roles in protection from neurodegenerative-associated stimuli [45], appeared to be co-regulated by SPI1 and CEBPA but not expressed within the protocol described by Chen et al. (Figure 4C). Other master TFs such as CEBPB, which was previously described as a major transcription factor regulating the expression of genes involved in inflammatory responses [46], and the ETV6 transcriptional repressor, which is known to play a critical role in hematopoiesis [47], are described as being regulated by CEBPA within the reconstructed regulome but not expressed within the protocol described by Chen et al. (Figure 4C).

These cumulative observations strongly support for a suboptimal activation of the core transcription regulatory circuitry driving microglial differentiation when using the protocol described by Chen and colleagues, which is most likely due to the partial activity associated with the over-expressed factor CEBPA. In fact, the read count enrichment at the CEBPA locus observed in the RNA-seq samples issued from the CEBPA over-expression generated by the protocol described by Chen and colleagues, reveals a shorter gene expression version than that observed in RNA-seq samples issued from all other protocols as well as those related to the fetal and adult microglia control datasets (Figure 4D). This shorter CEBPA transcript coincides also with a lower molecular weight being reported by the authors (268 amino acids) relative to the molecular weight described in public repositories (358 amino acids).

#### 4. Discussion and Conclusions

There is increasing evidence that microglia, the most performant immune cells in the brain, is involved in both homeostasis and injury/infection control and that it also plays a predominant role in neurodegenerative diseases. Indeed, microglia activation is dependent on neuroinflammation and might also be related to the HERV, which increases NO production [17]. This being said, studies on microglia *in vivo* are limited due to the necessity of surgical intervention, which can lead to the tissue destruction and the consequent activation of the microglia. Furthermore, their transfer to a tissue culture environment results in the rapid and extensive downregulation of microglia-specific genes [39]. Histological studies, on the other hand, may be helpful to provide clues regarding cell involvement in their environment; nevertheless, they are limited in terms of evaluating the functions of living cells [48].

In order to overcome this problem, different laboratories have moved towards the generation of microglia *in vitro* from circulating progenitors and stem cells. The first attempts successfully obtained microglia-like (MGL) cells from circulating monocytes and bone marrow-derived stem cells [49,50]. Nevertheless, these approaches represent non-ideal models since the microglia that were obtained do not share the same embryonic origin as the microglia in the brain, and these cells present a relatively advanced states of differentiation, making them unsuitable for asking questions about the regulatory networks governing their development. More recent approaches have focused on the differentiation of stem cells, either those issued from embryonic sources or from the reprogramming of differentiated material (ESC and iPSC, respectively) [22].

The first article, published in 2016, described a 74-days long protocol describing the differentiation of iPSC into microglia-like (iMGL) cells through the intermediate passage, generating neural progenitors [32]. Since then, several other protocols have been described [23–29], most of which are 30 days shorter than the previous one, notably because they do not require the neural progenitor step and because the generation of hematopoi-

etic progenitor cells (HPC) can be achieved by mimicking the in vivo microglial cells ontogeny [3]. Finally, Chen and colleagues, described a novel approach that produces the differentiation of iMGL by overexpressing the transcription factors SPI1 and CEBPA, both of which are known to play a major role in microglial cell fate during embryogenesis [30].

The transcriptomic reanalysis effort described in this article allowed us to highlight the commonalities but also the differences in gene programs between all six of the compared protocols. Importantly, this effort revealed a core transcription regulatory circuitry driving microglial differentiation composed by 19 master TFs that were over-expressed on all protocols, including SPI1 and CEBPA (Figure 4A). A second subset of master regulators that were expressed on all protocols (IRF9, MEF2A, NFIC, IRF8, CEBPB, EGR2) except the one described by Chen and colleagues was observed. Importantly, these factors are also part of the reconstructed SPI1/CEBPA regulome, supporting their direct connection to the inner core transcription regulatory circuitry (Figure 4A,C).

While all of these findings fully support the hypothesis of Chen and colleagues, which indicates that microglia differentiation can be driven by over-expressing SPI1 and CEBPA, the major discrepancies between this last protocol and all others seems to have origin on a truncated version of the CEBPA factor used during such study. In this context, it would be essential to reconsider the strategy, this time using endogenous promoter activation systems, such as the use of RNA-guided gene activation by CRISPR-Cas9-based transcription factors [51].

Finally, the presented core transcription regulatory circuitry provides means to enhance microglia differentiation protocols, but also to evaluate potential deregulated core transcription regulatory circuits, which might arise in the context of neurodegenerative diseases, which, for instance, can be modeled with the help of 3-dimensional cerebral organoid systems.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/brainsci11101338/s1>, Figure S1: Step by Step visualization of RNA-seq readouts on QC Genomics platform, Figure S2: Volcano plots revealing the differentially expressed genes (DEGs) in iMGL datasets generated by the various differentiation protocols, Figure S3: Master regulatory Transcription factors predicted by TETRAMER for each of the differential expression data associated to each of the iMGL differentiation protocols, Figure S4: Consensus master transcription regulatory circuitry driving microglia differentiation, Table S1: Public Datasets and their associated identifiers (GSM) used in this comparative study (Excel file), Table S2: Differential gene expression readouts (DeSeq2) (Excel file), Table S3: Microglia-specific up-regulated genes identified in this study (Excel file), Table S4: Master regulatory indexes per TF assessed by TETRAMER (Excel file).

**Author Contributions:** Conceptualization, M.A.M.-P.; data curation, F.S.; formal analysis, A.A. and F.S.; funding acquisition, M.A.M.-P.; investigation, A.A.; methodology, A.A.; software, F.S.; supervision, M.A.M.-P.; writing—original draft, B.M.C. and M.A.M.-P.; writing—review and editing, A.A. and B.M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the institutional bodies CEA, CNRS, and Université d'Evry-Val d'Essonne and by the "Fondation pour la Recherche Médicale" (FRM; funding ALZ° 201912009904). F.S. is supported by the funding agreement 20I9-L22 from the "Institut National du Cancer (INCa).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Datasets reanalyzed in this study were collected from the Gene expression Omnibus (GEO) database. The corresponding GEO series accession numbers are: GSE97744, GSE85839, GSE97744, GSE117829, GSE169065 and GSE163984.

**Conflicts of Interest:** The authors have no competing interests to declare.

## References

1. Sousa, C.; Biber, K.; Michelucci, A. Cellular and molecular characterization of microglia: A unique immune cell population. *Front. Immunol.* **2017**, *8*, 198. [[CrossRef](#)]
2. Ginhoux, F.; Greter, M.; Leboeuf, M.; Nandi, S.; See, P.; Gokhan, S.; Mehler, M.F.; Conway, S.J.; Ng, L.G.; Stanley, E.R.; et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **2010**, *330*, 841–845. [[CrossRef](#)]
3. Ginhoux, F.; Lim, S.; Hoeffel, G.; Low, D.; Huber, T. Origin and differentiation of microglia. *Front. Cell. Neurosci.* **2013**, *7*, 45. [[CrossRef](#)]
4. Tong, C.K.; Vidyadaran, S. Role of microglia in embryonic neurogenesis. *Exp. Biol. Med.* **2016**, *241*, 1669–1675. [[CrossRef](#)]
5. Luo, X.; Chen, S.-D. The changing phenotype of microglia from homeostasis to disease. *Transl. Neurodegener.* **2012**, *1*, 9. [[CrossRef](#)] [[PubMed](#)]
6. Bachiller, S.; Ferrer, I.J.; Paulus, A.; Yang, Y.; Swanberg, M.; Deierborg, T.; Boza-Serrano, A. Microglia in neurological diseases: A road map to brain-disease dependent-inflammatory response. *Front. Cell. Neurosci.* **2018**, *12*, 488. [[CrossRef](#)]
7. Cserép, C.; Pósfai, B.; Lénárt, N.; Fekete, R.; László, Z.I.; Lele, Z.; Orsolits, B.; Molnár, G.; Heindl, S.; Schwarcz, A.D.; et al. Microglia monitor and protect neuronal function through specialized somatic purinergic junctions. *Science* **2020**, *367*, 528–537. [[CrossRef](#)] [[PubMed](#)]
8. Kielian, T.; Mayes, P.; Kielian, M. Characterization of microglial responses to *Staphylococcus aureus*: Effects on cytokine, costimulatory molecule, and Toll-like receptor expression. *J. Neuroimmunol.* **2002**, *130*, 86–99. [[CrossRef](#)]
9. Town, T.; Nikolic, V.; Tan, J. The microglial “activation” continuum: From innate to adaptive responses. *J. Neuroinflammation* **2005**, *2*, 24. [[CrossRef](#)] [[PubMed](#)]
10. Michell-Robinson, M.; Touil, H.; Healy, L.; Owen, D.; Durafourt, B.; Bar-Or, A.; Antel, J.; Moore, C.S. Roles of microglia in brain development, tissue maintenance and repair. *Brain* **2015**, *138*, 1138–1159. [[CrossRef](#)]
11. Wojtera, M.; Sikorska, B.; Sobow, T.; Liberski, P.P. Microglial cells in neurodegenerative disorders. *Folia Neuropathol.* **2005**, *43*, 311–321.
12. Cho, S.-H.; Sun, B.; Zhou, Y.; Kauppinen, T.; Halabisky, B.; Wes, P.; Ransohoff, R.M.; Gan, L. CX3CR1 Protein signaling modulates microglial activation and protects against plaque-independent cognitive deficits in a mouse model of Alzheimer disease. *J. Biol. Chem.* **2011**, *286*, 32713–32722. [[CrossRef](#)]
13. De Pablos, R.M.; Herrera, A.J.; Espinosa-Oliva, A.M.; Sarmiento, M.; Muñoz, M.F.; Machado, A.; Venero, J.L. Chronic stress enhances microglia activation and exacerbates death of nigral dopaminergic neurons under conditions of inflammation. *J. Neuroinflamm.* **2014**, *11*, 34. [[CrossRef](#)]
14. Hansen, D.V.; Hanson, J.E.; Sheng, M. Microglia in Alzheimer’s disease. *J. Cell Biol.* **2018**, *217*, 459–472. [[CrossRef](#)]
15. Chagas, L.; Sandre, P.C.; Ribeiro, N.C.A.R.; Marcondes, H.; Silva, P.O.; Savino, W.; Serfaty, C.A. Environmental signals on microglial function during brain development, neuroplasticity, and disease. *Int. J. Mol. Sci.* **2020**, *21*, 2111. [[CrossRef](#)] [[PubMed](#)]
16. Lecours, C.; Bordeleau, M.; Cantin, L.; Parent, M.; Di Paolo, T.; Tremblay, M. Microglial implication in Parkinson’s disease: Loss of Beneficial physiological roles or gain of inflammatory functions? *Front. Cell. Neurosci.* **2018**, *12*, 282. [[CrossRef](#)] [[PubMed](#)]
17. Alster, P.; Madetko, N.; Kozirowski, D.; Friedman, A. Microglial Activation and Inflammation as a Factor in the Pathogenesis of Progressive Supranuclear Palsy (PSP). *Front. Neurosci.* **2020**, *14*, 893. [[CrossRef](#)] [[PubMed](#)]
18. Kozirowski, D.; Figura, M.; Milanowski, L.M.; Szlufik, S.; Alster, P.; Madetko, N.; Friedman, A. Mechanisms of neurodegeneration in various forms of parkinsonism—Similarities and differences. *Cells* **2021**, *10*, 656. [[CrossRef](#)]
19. Hemonnot, A.-L.; Hua, J.; Ulmann, L.; Hirbec, H. Microglia in Alzheimer disease: Well-known targets and new opportunities. *Front. Aging Neurosci.* **2019**, *11*, 233. [[CrossRef](#)] [[PubMed](#)]
20. Xiao, R.; Li, S.; Cao, Q.; Wang, X.; Yan, Q.; Tu, X.; Zhu, Y.; Zhu, F. Human endogenous retrovirus W env increases nitric oxide production and enhances the migration ability of microglia by regulating the expression of inducible nitric oxide synthase. *Virology* **2017**, *32*, 216–225. [[CrossRef](#)]
21. Gruchot, J.; Kremer, D.; Küry, P. Neural cell responses upon exposure to human endogenous retroviruses. *Front. Genet.* **2019**, *10*, 655. [[CrossRef](#)] [[PubMed](#)]
22. Speicher, A.M.; Wiendl, H.; Meuth, S.G.; Pawlowski, M. Generating microglia from human pluripotent stem cells: Novel in vitro models for the study of neurodegeneration. *Mol. Neurodegener.* **2019**, *14*, 46. [[CrossRef](#)] [[PubMed](#)]
23. Pandya, H.; Shen, M.J.; Ichikawa, D.M.; Sedlock, A.B.; Choi, Y.; Johnson, K.R.; Kim, G.; Brown, M.; Elkhalloun, A.G.; Maric, D.; et al. Differentiation of human and murine induced pluripotent stem cells to microglia-like cells. *Nat. Neurosci.* **2017**, *20*, 753–759. [[CrossRef](#)]
24. Hänseler, W.; Sansom, S.N.; Buchrieser, J.; Newey, S.E.; Moore, C.S.; Nicholls, F.; Chintawar, S.; Schnell, C.; Antel, J.; Allen, N.; et al. A highly efficient human pluripotent stem cell microglia model displays a neuronal-co-culture-specific expression profile and inflammatory response. *Stem Cell Rep.* **2017**, *8*, 1727–1742. [[CrossRef](#)] [[PubMed](#)]
25. Abud, E.M.; Ramirez, R.N.; Martinez, E.S.; Healy, L.; Nguyen, C.H.; Newman, S.A.; Yeromin, A.V.; Scarfone, V.; Marsh, S.; Fimbres, C.; et al. iPSC-derived human microglia-like cells to study neurological diseases. *Neuron* **2017**, *94*, 278–293.e9. [[CrossRef](#)] [[PubMed](#)]
26. McQuade, A.; Coburn, M.; Tu, C.H.; Hasselmann, J.; Davtyan, H.; Blurton-Jones, M. Development and validation of a simplified method to generate human microglia from pluripotent stem cells. *Mol. Neurodegener.* **2018**, *13*, 67. [[CrossRef](#)] [[PubMed](#)]
27. Douvaras, P.; Sun, B.; Wang, M.; Kruglikov, I.; Lалlos, G.; Zimmer, M.; Terrenoire, C.; Zhang, B.; Gandy, S.; Schadt, E.; et al. Directed differentiation of human pluripotent stem cells to microglia. *Stem Cell Rep.* **2017**, *8*, 1516–1524. [[CrossRef](#)]
28. Takata, K.; Kozaki, T.; Lee, C.; Thion, M.S.; Otsuka, M.; Lim, S.; Utami, K.H.; Fidan, K.; Park, D.S.; Malleret, B.; et al. Induced-pluripotent-stem-cell-derived primitive macrophages provide a platform for modeling tissue-resident macrophage differentiation and function. *Immunity* **2017**, *47*, 183–198.e6. [[CrossRef](#)]

29. Trudler, D.; Nazor, K.L.; Eisele, Y.S.; Grabauskas, T.; Dolatabadi, N.; Parker, J.; Sultan, A.; Zhong, Z.; Goodwin, M.S.; Levites, Y.; et al. Soluble  $\alpha$ -synuclein–antibody complexes activate the NLRP3 inflammasome in hiPSC-derived microglia. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2025847118. [[CrossRef](#)]
30. Chen, S.-W.; Hung, Y.-S.; Fuh, J.-L.; Chen, N.-J.; Chu, Y.-S.; Chen, S.-C.; Fann, M.-J.; Wong, Y.-H. Efficient conversion of human induced pluripotent stem cells into microglia by defined transcription factors. *Stem Cell Rep.* **2021**, *16*, 1363–1380. [[CrossRef](#)]
31. Davis, T.L.; Rebay, I. Master regulators in development: Views from the Drosophila retinal determination and mammalian pluripotency gene networks. *Dev. Biol.* **2017**, *421*, 93–107. [[CrossRef](#)]
32. Muffat, J.; Li, Y.; Yuan, B.; Mitalipova, M.; Omer, A.; Corcoran, S.; Bakiasi, G.; Tsai, L.-H.; Aubourg, P.; Ransohoff, R.M.; et al. Efficient derivation of microglia-like cells from human pluripotent stem cells. *Nat. Med.* **2016**, *22*, 1358–1367. [[CrossRef](#)] [[PubMed](#)]
33. Blum, M.; Cholley, P.-E.; Malysheva, V.; Nicaise, S.; Moehlin, J.; Gronemeyer, H.; Mendoza-Parra, M.A. A comprehensive resource for retrieving, visualizing, and integrating functional genomics data. *Life Sci. Alliance* **2020**, *3*, e201900546. [[CrossRef](#)] [[PubMed](#)]
34. Mendoza-Parra, M.A.; Saleem, M.-A.M.; Blum, M.; Cholley, P.-E.; Gronemeyer, H. NGS-QC generator: A quality control system for ChIP-seq and related deep sequencing-generated datasets. In *Statistical Genomics: Methods and Protocols*; Mathé, E., Davis, S., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2016; pp. 243–265.
35. Liao, Y.; Smyth, G.K.; Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [[CrossRef](#)] [[PubMed](#)]
36. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma’Ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **2013**, *14*, 128. [[CrossRef](#)] [[PubMed](#)]
37. Marbach, D.; Lamparter, D.; Quon, G.; Kellis, M.; Kutalik, Z.; Bergmann, S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **2016**, *13*, 366–370. [[CrossRef](#)] [[PubMed](#)]
38. Cholley, P.-E.; Moehlin, J.; Rohmer, A.; Zilliox, V.; Nicaise, S.; Gronemeyer, H.; Mendoza-Parra, M.A. Modeling gene-regulatory networks to describe cell fate transitions and predict master regulators. *NPJ Syst. Biol. Appl.* **2018**, *4*, 29. [[CrossRef](#)]
39. Gosselin, D.; Skola, D.; Coufal, N.G.; Holtman, I.R.; Schlachetzki, J.C.M.; Sajti, E.; Jaeger, B.N.; O’Connor, C.; Fitzpatrick, C.; Pasillas, M.P.; et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **2017**, *356*, eaal3222. [[CrossRef](#)]
40. Banerjee, A.; Lu, Y.; Do, K.; Mize, T.; Wu, X.; Chen, X.; Chen, J. Validation of induced microglia-like cells (iMG Cells) for future studies of brain diseases. *Front. Cell. Neurosci.* **2021**, *15*, 85. [[CrossRef](#)]
41. Grubman, A.; Vandekolk, T.H.; Schroeder, J.; Sun, G.; Hatwell-Humble, J.; Chan, J.; Oksanen, M.; Lehtonen, S.; Hunt, C.; Koistinaho, J.E.; et al. A CX3CR1 reporter hESC line facilitates integrative analysis of in-vitro-derived microglia and improved microglia identity upon neuron-glia co-culture. *Stem Cell Rep.* **2020**, *14*, 1018–1032. [[CrossRef](#)]
42. Mendoza-Parra, M.-A.; Van Gool, W.; Saleem, M.A.M.; Ceschin, D.G.; Gronemeyer, H. A quality control system for profiles obtained by ChIP sequencing. *Nucleic Acids Res.* **2013**, *41*, e196. [[CrossRef](#)]
43. Masuda, T.; Tsuda, M.; Yoshinaga, R.; Tozaki-Saitoh, H.; Ozato, K.; Tamura, T.; Inoue, K. IRF8 is a critical transcription factor for transforming microglia into a reactive phenotype. *Cell Rep.* **2012**, *1*, 334–340. [[CrossRef](#)]
44. Crotti, A.; Ransohoff, R.M. Microglial physiology and pathophysiology: Insights from genome-wide transcriptional profiling. *Immunity* **2016**, *44*, 505–515. [[CrossRef](#)]
45. Dresselhaus, E.C.; Meffert, M.K. Cellular specificity of NF- $\kappa$ B Function in the nervous system. *Front. Immunol.* **2019**, *10*, 1043. [[CrossRef](#)]
46. Cortes-Canteli, M.; Luna-Medina, R.; Sanz-SanCristobal, M.; Alvarez-Barrientos, A.; Santos, A.; Perez-Castillo, A. CCAAT/enhancer binding protein  $\beta$  deficiency provides cerebral protection following excitotoxic injury. *J. Cell Sci.* **2008**, *121*, 1224–1234. [[CrossRef](#)] [[PubMed](#)]
47. Hock, H.; Shimamura, A. ETV6 in hematopoiesis and leukemia predisposition. *Semin. Hematol.* **2017**, *54*, 98–104. [[CrossRef](#)] [[PubMed](#)]
48. Stence, N.; Waite, M.; Dailey, M. Dynamics of microglial activation: A confocal time-lapse analysis in hippocampal slices. *Glia* **2001**, *33*, 256–266. [[CrossRef](#)]
49. Sievers, J.; Parwaresch, R.; Wottge, H.-U. Blood monocytes and spleen macrophages differentiate into microglia-like cells on monolayers of astrocytes: Morphology. *Glia* **1994**, *12*, 245–258. [[CrossRef](#)]
50. Hinze, A.; Stolzing, A. Microglia differentiation using a culture system for the expansion of mice non-adherent bone marrow stem cells. *J. Inflamm.* **2012**, *9*, 12. [[CrossRef](#)] [[PubMed](#)]
51. Perez-Pinera, P.; Kocak, D.D.; Vockley, C.M.; Adler, A.F.; Kabadi, A.M.; Polstein, L.R.; Thakore, P.I.; Glass, K.; Ousterout, D.G.; Leong, K.; et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **2013**, *10*, 973–976. [[CrossRef](#)]