Exploring reported genes of microglia RNA-sequencing data

van Wageningen, Thecla A.; Gerrits, Emma; Bonson, Sara Palacin; Huitinga, Inge; Eggen, Bart J. L.; van Dam, Anne-Marie

Published in:
Glia

DOI:
10.1002/glia.24078

Publication date:
2021

Received: 24 June 2021 | Revised: 6 August 2021 | Accepted: 11 August 2021

DOI: 10.1002/glia.24078

GLIA

WILEY

RESEARCH ARTICLE

Exploring reported genes of microglia RNA-sequencing data: Uses and considerations

Thecla A. van Wageningen[1] | Emma Gerrits[2] | Sara Palacin i Bonson[1] | Inge Huitinga[3] | Bart J. L. Eggen[2] | Anne-Marie van Dam[1]

[1]Department Anatomy & Neurosciences, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Department of Biomedical Sciences of Cells & Systems, section Molecular Neurobiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

[3]Neuroimmunology Research Group, Netherlands Institute for Neuroscience, Amsterdam, Netherlands

Correspondence
Thecla A. van Wageningen, Department Anatomy & Neurosciences, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.
Email: t.vanwageningen@amsterdamumc.nl

Abstract

The advent of RNA-sequencing techniques has made it possible to generate large, unbiased gene expression datasets of tissues and cell types. Several studies describing gene expression data of microglia from Alzheimer's disease or multiple sclerosis have been published, aiming to generate more insight into the role of microglia in these neurological diseases. Though the raw sequencing data are often deposited in open access databases, the most accessible source of data for scientists is what is reported in published manuscripts. We observed a relatively limited overlap in reported differentially expressed genes between various microglia RNA-sequencing studies from multiple sclerosis or Alzheimer's diseases. It was clear that differences in experimental set up influenced the number of overlapping reported genes. However, even when the experimental set up was very similar, we observed that overlap in reported genes could be low. We identified that papers reporting large numbers of differentially expressed microglial genes generally showed higher overlap with other papers. In addition, though the pathology present within the tissue used for sequencing can greatly influence microglia gene expression, often the pathology present in samples used for sequencing was underreported, leaving it difficult to assess the data. Whereas reanalyzing every raw dataset could reduce the variation that contributes to the observed limited overlap in reported genes, this is not feasible for labs without (access to) bioinformatic expertise. In this study, we thus provide an overview of data present in manuscripts and their supplementary files and how these data can be interpreted.

KEYWORDS
Alzheimers disease, bioinformatics, Jaccard index, microglia, multiple sclerosis, RNA-seq

1 | INTRODUCTION

With RNA-sequencing (RNA-seq), gene expression of cells and tissues can be profiled (Tang et al., 2009). In the field of neurobiology, RNA-seq has contributed to the identification of genes specific for various subsets of glial cells (Hammond et al., 2019) and genes discriminating, for example, microglia from brain infiltrating macrophages (Bennett et al., 2016; O. Butovsky et al., 2014; Hickman et al., 2013).

Furthermore, RNA-seq of cell populations offers the possibility to identify new molecular targets relevant for disease pathology or drug development. There is ample (genetic) evidence that microglia contribute to various neurological disorders, among which Alzheimer's disease (AD) and multiple sclerosis (MS) (Butovsky & Weiner, 2018; Voet et al., 2019). RNA-seq of microglia from these diseased conditions has identified genes involved in their pathophysiology, and the potential development of treatment for these neurological disorders. Yet, gene expression profiling of large amounts of (individual) cells leads to enormous datasets which can be challenging to untangle and interpret by inexperienced users (Koch et al., 2018). RNA-seq data are often available on open access platforms such as NCBI GEO and ArrayExpress, but they are not always directly accessible for use, first requiring bioinformatic processing and analyses. The most accessible source of data from RNA-seq studies is what is reported in published manuscripts and their supplementary materials. Thus, we set out to compare the reported differentially expressed (DE) microglia genes of interest for MS or for AD mouse models and human tissue compared to control tissue in addition to creating an overview of the experimental set up, techniques used and brain tissue used in these various studies. Using these data, the aim is to gain more insight into which genes are reported within current microglia RNA-seq manuscripts, (dis)similarities between reported genes in these manuscripts and how use by other researchers could be facilitated.

2 | METHODS

A literature search was performed using common platforms such as PubMed, Google Scholar and ScienceDirect using MESH TERMS tool for entries containing the following terms: "RNA," "RNA-sequencing," "MS," "Multiple Sclerosis," "AD," "Alzheimer's disease," and "microglia." Papers were filtered by year of publication, and papers published between 2015 and 2021 were included for analysis. Only papers reporting gene expression data from AD or MS mouse models compared to control mice, and from AD or MS human tissue compared to control tissue were included (see Tables 1 and 2).

Of note, in this manuscript the term "RNA-seq" refers to both bulk and single cell RNA-seq of microglia subjected to differential gene expression analysis.

2.1 | Construction of a core gene set from mouse model and human tissue microglia RNA-seq papers

Per paper (for AD or MS studies) a list of reported genes was derived from the main body of the paper or from the supplementary files. A subsequent "core gene set" was compiled of genes present in at least two gene lists of the same disease. This core gene set was then compared to the each gene list prepared per paper using the function COUNT.IF in Microsoft Excel to determine gene overlap and is represented in a binary matrix indicating the presence (1) or the absence of each gene (0) reported per paper (Tables S1–S4).

2.2 | Commonality assessment

From these binary matrices, Jaccard indices were calculated using the *jaccard score* function from the *sklearn_metrics* package for Python (Python Software Foundation, https://www.python.org/). The Jaccard index is a measure optimized to compare the overlap between two binary lists (Chung et al., 2019) and is calculated by the number of overlapping microglia genes present in two gene lists (i.e. from two papers) and are mentioned in the core gene set divided by the total number of genes present in the two combined gene lists. It gives an indication of the similarity between binary lists where 1 indicates a complete overlap between the gene lists and 0 indicates a complete lack of overlap between the gene lists. As the Jaccard index takes into account the length of the gene lists, comparison of papers with a considerable difference in number of reported genes may result in relatively low Jaccard indices. For example, gene list A of 100 genes present in the core gene set may show complete overlap with gene list B which features 1000 genes in the core gene set, yet the Jaccard index will be relatively low as the 100 shared genes are divided by the total amount of genes which in this case is 1100 genes, leading to a jaccard index of $100/1100 = 0.09$. The Jaccard index thus also takes into account the 900 genes reported in gene list B which were not reported in gene list A. A heatmap of the Jaccard indices was created in R (R Core Team, 2020) using the *heatmap.2* function from the *gplots* package (Warnes et al., 2009).

2.3 | Comparison of mouse model data to human data

In order to compare mouse and human gene expression data, a core gene set of human tissue related genes was constructed as described above (see Tables S3 and S4). Subsequently, we used the COUNT.IF function in Microsoft Excel to identify the presence or absence of a gene in the human core gene set in each mouse model paper. Homology between human and rodent genes was not assessed. Heatmaps of human genes reported in mouse model papers were created in R (R Core Team, 2020) using the *heatmap.2* function from the *gplots* package (Warnes et al., 2009).

2.4 | Pathway analysis of reported genes

Gene lists of reported genes derived from each paper were used for pathway analysis using the g:Profiler website (Raudvere et al., 2019). Commonality of the top 15 enriched GO biological pathways (GO BP) was assessed by using the COUNT.IF function in Microsoft Excel (Tables S1–S4).

3 | RESULTS

The included papers reported on RNA-seq data of microglia from control CNS compared to diseased CNS in the main body of the text or in

TABLE 1 Overview of the model, experimental set up, sequencing method and data availability of AD papers (mouse and human) included in this study

Study	Model	Tissue used and (number of cases = *n*)	Dissociation method	Microglia isolation method	Library prep kit	# microglia cells used for analysis[b]	Reported DE genes[c]	Data deposition	Data taken from
Frigerio et al. (2019) doi: 10.1016/j.celrep.2019.03.099	Female and male *C57BL/6J* WT and *App*$^{NL\text{-}G\text{-}F}$ mice	Cortex and hippocampus ($n = 2$ per condition)	E	Debris Removal Solution gradient followed FACS of CD11b$^+$ and DAPI$^-$	Modified SmartSeq2 scRNA-seq	10,801	25	GEO NCBI accession number: GSE127893	From main text and supplementary information figure s.5
Keren-Shaul et al. (2017) doi: 10.1016/j.cell.2017.05.018	*C57/BL6-SJL* WT and *5XFAD* transgenic mice	Whole brains ($n = 3$ per condition)	M	Percoll gradient followed by FACS of CD45$^+$	Massively parallel (MARS) scRNA-seq	8016	149	GEO NCBI accession number: GSE98971	Supplementary figures—table s.1 Average Gene Expression in Immune Cell Clusters from the Brain of AD and WT Mice, Related to figure 1 (ordered the gene list upon microglia subtype III, cut off at 150)
Friedman et al. (2018) doi: 10.1016/j.celrep.2017.12.066	*Cx3cr1*$^{GFP/+}$; *PS2APP*^{negative} mice and *Cx3cr1*$^{GFP/+}$; *PS2APP*^{homozygous} mice	Cortex ($n = 3$ per condition)	M	Percoll gradient followed by FACS for CD11b$^+$ and DAPI$^+$	Ovation RNA-Seq System V2 (NuGEN)	NA	198	GEO NCBI accession number: GSE89482	Supplementary information, Data s.4 mouse data. Adj. P values <0.05 for *PS2APP*$^+$ 13 mo samples.
Mathys et al. (2017) doi: 10.1016/j.celrep.2017.09.039	CK-p25 mice compared to control CK mice	Hippocampus ($n = 3$ per condition)	E	CD11b$^+$ microbeads followed by FACS for CD11b$^+$ and CD45$^+$	Modified SMART-Seq2	1685	40	GEO NCBI accession number: GSE103334	Main text, figure 3 D, E and F Supplementary table s.4 cluster 2 versus 3,
Holtman et al. (2015) doi: 10.1186/s40478-015-0203-5	Data of aging *DBA* and *BL6-SJL* mice and *APP-PS1* mice compared to intraperitoneal LPS injected mice	Whole brain (n = not indicated in the manuscript)	M	Percoll gradient followed by FACS for CD11b$^+$ and CD45$^+$	Illumina MouseRef8 bead-chip microarrays	NA	195	GEO NCBI accession number: GSE74615 GSE43366	Supplementary information—Additional file s3, mouse model specific genes
Srinivasan et al. (2016) doi: 10.1038/ncomms11295	Female PS2APP mice compared to WT mice	Cortex ($n = 5$ per condition)	M	Percoll gradient followed by FACS for CD11b$^+$ and DAPI$^+$	Ovation RNA-Seq System V2 (NuGEN)	NA	78	GEO NCBI accession number: GSE75431	Main text, figure 9.A 13 m.o Tg$^+$ microglia column
Krasemann et al. (2017) doi: 10.1016/j.immuni.2017.08.008	APP-PS1 mice compared to WT mice	Whole brain ($n = 2$ per condition)	M	Percoll gradient followed by FACS for FCRLS$^+$ and CD11b$^+$	Smart-Seq2	1000	405	GEO NCBI accession number: GSE101689	Supplementary information—table s.1. Microglia Signature during Disease Progression, Related to figure 1, MG in APPPS1 disease and aging
Zhou et al. (2020) doi: 10.1038/s41591-019-0695-9	WT (C57BL/6 J) and 5XFAD (Tg6799) mice	Whole brain ($n = 3$ per condition)	M	Isolation of frozen nuclei by sucrose	Droplet-based 3′ end massively parallel single-cell RNA sequencing	2251	1361	GEO NCBI accession number: GSE140511.	Supplementary table 1, cluster 0, p_value_adj <0.05

(Continues)

TABLE 1 (Continued)

Study	Model	Tissue used and (number of cases = *n*)	Dissociation method	Microglia isolation method	Library prep kit	# microglia cells used for analysis[b]	Reported DE genes[c]	Data deposition	Data taken from
Sobue et al. (2021) doi: 10.1186/s40478-020-01099-x	AppNL-G-F/NL-G-F compared to WT mice	Cerebral cortex ($n = 4$ per condition)	E	Removal of myelin using Myelin Removal Beads II, followed by incubation with magnetic beads for $CD16/CD32^+$, $CD11b^+$	TruSeq (Stranded) mRNA	NA	3318	NA	Supplementary table 1
Zhou et al. (2020)[a] doi: 10.1038/s41591-019-0695-9	AD patients and controls	Prefrontal cortical tissue ($n = 3$ per condition)	M	Isolation of frozen nuclei by sucrose	Droplet-based 3′ end massively parallel single-cell RNA sequencing	3986	565	AD Knowledge Portal (https://adknowledgeportal.org) under study snRNAseqAD_TREM2 and are also accessible through https://doi.org/10.7303/syn21125841.	Supplementary table 4, micro, p_value_adj <0.05
Grubman et al. (2019)[a] doi: 10.1038/s41593-019-0539-4	AD patients and non-demented controls	Entorhinal cortex ($n = 6$ per condition)	M	Purification of extracted nuclei by cell sorting using $DAPI^+$	NextSeq 500 (Illumina)	449	29	GEO NCBI accession number: GSE138852	Supplementary table 2
Gerrits et al. (2021)[a] doi: 10.1007/s00401-021-02263-w	AD patients and non-demented controls	Gray matter of the occipital cortex (OC) and occipitotemporal cortex (OTC) ($n = 10$ AD, $n = 8$ control)	E	Purification of extracted frozen nuclei by FACS for $DAPI^{pos}NEUN^{neg}OLIG2^{neg}$	3′ single cell RNAseq from 10x Genomics	150,000	2898	GEO NCBI accession number: GSE148822	Supplementary table 7, AD1 and AD2 vs. homeostatic, p_value_adj <.05.
Mathys et al. (2019)[a] doi: 10.1038/s41586-019-1195-2	AD patients with pathology and controls with limited to no pathology	Prefrontal cortex ($n = 24$ per condition)	M	Purification of extracted frozen nuclei	NextSeq 500/550 High Output v2	955	110	https://www.radc.rush.edu/docs/omics.htm (snRNA-seq PFC) or at Synapse (https://www.synapse.org/#!Synapse:syn18485175) under the doi 10.7303/syn18485175	Supplementary table 2, Sheet "Mic" comparison of no-pathology to pathology
Srinivasan et al. (2020)[a] doi: 10.1016/j.celrep.2020.107843	AD patients compared to control	Frozen superior frontal gyrus and fusiform gyrus tissue blocks ($n = 10$ AD, $n = 15$ control)	M	FACS sorting for $GFAP^+$, $NeuN^+$, $CD31^+$, $CD11b^+$ $DAPI^+$	Illumina HiSeq2500	NA	75	GEO NCBI accession number: GSE125050	Supplementary table 2, section GSE125050, diagnosis: AD—Control| celltype = myeloid, p value adj <0.05.

Abbreviations: APP, amyloid precursor protein; E, enzymatic; FACS, fluorescent activated cell sorting; M, mechanical; NA, not applicable; scRNA-seq, single cell RNA sequencing; WT, wild type.

[a]Human data.

[b]For single-cell sequencing studies, NA, bulk sequencing was employed.

[c]Total of enriched and depleted genes.

TABLE 2 Overview of the model, experimental set up, sequencing method and data availability of MS papers (mouse and human) included in this study

Study	Model	Tissue used and (number of cases = *n*)	Dissociation method	Microglia isolation method	Library prep kit	# microglia cells used for analysis[b]	Reported DE genes[b]	Data deposition	Data taken from
Mendiola et al. (2020) doi: 10.1038/s41590-020-0654-0	Healthy C57Bl/6 mice compared chronic EAE induced by MOG^{35-55} in C57Bl/6 mice	Spinal cord ($n = 2$ per condition)	M followed by myelin removal by magnetic beads	FACS sorting of $CD11b^+$	Chromium Single Cell 3′ v2 Reagent kit	8701	1041	GEO NCBI accession number: GSE146295	Supplementary information, Supp_table_4 selected genes from clusters MgIII, MgIV, MgV from EAE microglia.
Hammond et al. (2019) doi: 10.1016/j.immuni.2018.11.004	Control C57BL/6J mice compared to C57BL/6J mice injected with LPC	Microdissected white matter ($n = 3$ per condition)	M	Percoll gradient followed by FACS for $CD11b^+$, $CD45^+$ and $Cx3cr1^+$	Chromium single cell gene expression platform	76,000	196	GEO NCBI accession number: GSE121654	Supplementary information, table s.1 from cluster 9
Jordão et al. (2019) doi: 10.1126/science.aat7554	Control C57BL/6N mice compared to C57Bl/6N induced with EAE (MOG^{35-55} EAE)	Whole brain ($n = 5$–6 per condition)	M	Percoll gradient followed by FACS for $CD45^+$ and $CD11b^+$	mCEL-Seq2	3461	29	GEO NCBI accession number: GSE118948	Main text, from figure 3.F
Krasemann et al. (2017) doi: 10.1016/j.immuni.2017.08.008	Control SJL/J female mice compared to SJL/J female mice induced with EAE ($PLP^{139-151}$)	Spinal cord ($n = 2$ per condition)	M	Percoll gradient followed by FACS for $CD11b^+$	Smart-Seq2	1000	544	GEO NCBI accession number: GSE101689	Supplementary information—table s.1. Microglia Signature during Disease Progression, Related to figure 1, MG in EAE disease stages.
Schirmer et al. (2019)[a] doi: 10.1038/s41586-019-1404-z	MS patients and non-demented controls	Normal appearing and demyelinated subcortical WM and GM tissue ($n = 12$ MS, $n = 6$ control)	Isolated nuclei from snap-frozen tissue	No prior selection for microglia: microglial subpopulations were identified from overall scRNA-seq data based on known microglial markers	10x Genomics Single-Cell 3′ system	1683	25	Sequence Read Archive under accession number: PRJNA544731	Supplementary table 6, filtered by celltype microglia

(Continues)

TABLE 2 (Continued)

Study	Model	Tissue used and (number of cases = *n*)	Dissociation method	Microglia isolation method	Library prep kit	# microglia cells used for analysis[b]	Reported DE genes[b]	Data deposition	Data taken from
Van der Poel et al. (2019)[a] doi: 10.1038/s41467-019-08976-7	MS patients and non-demented controls	Normal appearing GM (occipital lobe) and WM (corpus callosum) ($n = 15$ MS, $n = 16$ control)	E	Percoll gradient followed by $CD11b^+$ magnetic beads and FACS for $CD45^+$, $CD11b^+$ and $CD15^-$	NEBNext Ultra Directional RNA Library Prep Kit from Illumina	NA	177	GEO NCBI accession number: GSE111972	Supplementary information, Supplementary table 2, pooled data from Excel sheets comparing DE MS WM vs. control WM up and down and DE MS GM vs control GM up and down.
Jäkel et al. (2019)[a] doi: 10.1038/s41586-019-0903-2	MS patients and non-demented controls	Macrodissected normal appearing white matter and white matter lesions ($n = 5$ for MS and for control)	M	Strained with a 30 μm strainer followed with debris removal solution	Chromium Single Cell 3′ Library and Gel Bead kit v2	428	163	European Genome-phenome Archive (EGA): EGAS00001003412	Supplementary table 3

Abbreviations: E, enzymatic; EAE, experimental autoimmune encephalitis; GM, gray matter; LPC, lysophosphatidylcholine; M, mechanical; MOG, myelin oligodendrocyte glycoprotein; PLP1, proteolipid protein; WM, white matter.

[a]Human studies.

[b]Both enriched and depleted.

the supplementary data. Of these, nine papers presented data obtained from mouse models of AD (Friedman et al., 2018; Frigerio et al., 2019; Holtman et al., 2015; Keren-Shaul et al., 2017; Krasemann et al., 2017; Mathys et al., 2017; Sobue et al., 2021; Srinivasan et al., 2016; Zhou et al., 2020) and four papers presented data on mouse models of MS (Hammond et al., 2019; Jordão et al., 2019; Krasemann et al., 2017; Mendiola et al., 2020). in addition, five papers reported on microglia RNA-seq data from human AD tissue compared to control tissue (Gerrits et al., 2021; Grubman et al., 2019; Mathys et al., 2019; Srinivasan et al., 2020; Zhou et al., 2020) and three papers presented RNA-seq data from human MS tissue compared to control tissue (Jäkel et al., 2019; Schirmer et al., 2019; van der Poel et al., 2019). An overview of the selected papers and used experimental procedures therein can be found in Tables 1 and 2.

3.1 | Commonality of reported genes in mouse models of AD compared to control

Commonality of reported genes was assessed by calculating the Jaccard index (Chung et al., 2019) of genes reported in two or more papers (Table S1). If papers showed a complete overlap in reported genes, a value of 1 was given, if there was no overlap, a value of 0 was given. We observed very limited overlap in reported microglial DE genes between all AD mouse model papers (Figure 1(a)), with a complete lack of overlap between i.e. Mathys et al. (2017) and Zhou et al. (2020). The largest overlap was observed between Sobue et al. (2021) and Zhou et al. (2020) (Jaccard index of 0.79, Figure 1(a)) which was surprising as these two studies used different AD mouse models, had different experimental procedures and different sequencing methods (Table 1). Interestingly, though Srinivasan et al. (2016) and Friedman et al. (2018) reported on similar mouse models, brain areas studied, dissociation method and microglia isolation methods (Table 1), they only showed a Jaccard index of 0.2 (Figure 1(a)) which can be considered relatively low. Even though Zhou et al. (2020) and Sobue et al. (2021) showed a high Jaccard index between them, their Jaccard indices with other papers were relatively low (Jaccard index ranging from 0 to 0.12, Figure 1(a)). Many genes reported by, that is, Sobue et al. (2021) and Keren-Shaul et al. (2017) were represented in the top 25 most reported microglia DE genes (Figure 1(b)) even though their Jaccard Index was only 0.1 (Figure 1(a)). Thus, whereas Sobue et al. (2021) and Keren-Shaul et al. (2017) reported on the same DE genes, Sobue et al. (2021) reported on many other genes not reported in Keren-Shaul et al. (2017) lowering the Jaccard index (Table S1). The top 25 most reported genes included well known genes related to AD such as *Clec7a*, *Trem2*, *Apoe*, and *Itgax*. Lastly, most papers included in this study used single-cell sequencing methods and used about 2–3 animals per condition. Still we observed considerable variation in the number of microglia cells sequenced (ranging from ~1000 to 10,801 cells, Table 1), but we observed no direct relation between the number of cells sequenced and the number of DE microglia genes reported in a paper. Enriched pathways present within gene lists taken from all AD mouse model studies also showed considerable differences between papers. Overall, only two GO-pathways ("*response to external stimulus*" and "*immune system process*") were enriched in 4 out of the 9 studies included within this study whereas 79 of the total of 104 unique pathways were found enriched in one study only (Table S1). Together, the data indicate that there is little overlap in reported DE microglia genes between various AD mouse model studies which is at least related to differential experimental set ups and the number of reported microglia genes per paper. In addition, lack of overlap in enriched pathways corroborate that the lack of overlap in reported genes within each paper may influence conclusions about microglia function drawn from each paper.

3.2 | Commonality of reported genes in mouse models of MS compared to control

Similar to data from AD mouse models, we observed a limited overlap in reported microglia DE genes in MS mouse model RNA-seq papers (Figure 1(c), Table S2). Of note is that compared to AD, we found less papers describing microglia RNA-seq data of MS mouse models making it more difficult to assess the possible cause of little overlap in reported genes. A high Jaccard index (0.69) was found between the data of Hammond et al. (2019) and Mendiola et al. (2020), even though these studies used different MS mouse models and tissues from which microglia were isolated (Table 2). Conversely, a very low Jaccard index (0.09, Figure 1(c)) was found between Mendiola et al. (2020) and Jordão et al. (2019), even though these papers used the same MS mouse model and microglia isolation methods, though they did use different tissues (Table 2). When looking at the top 25 reported microglia DE genes (including *Apoe*, *Cxcl10*, *Cst7*, *Ccl5*, *Ccl4*, *Ccl2*) in papers, we observed more overlap in often reported top 25 DE genes between Krasemann et al. (2017) and Mendiola et al. (2020) while Jordão et al. (2019) reported the least genes present in the top 25 DE genes (Figure 1(d)). The lack of overlap in DE genes reported by Jordão et al. (2019) with the other studies may partly be related to the relatively low number of DE genes reported by Jordão et al. (2019) (Table 2). Analysis of enriched GO-pathways of gene lists showed that 2 pathways were enriched in 3 out of 4 studies ("*cellular response to chemical stimulus*" and "*response to external stimulus*") and 9 out of 47 unique pathways found were present in 2 out of 4 studies (Table S2). Similarly to what is observed in AD mouse model papers, there is little overlap in reported DE genes and enriched pathways between MS mouse model studies which could be due to experimental set up (such as i.e. brain region studied) but is likely also influenced by differences between the papers in the numbers of reported microglia DE genes.

3.3 | Commonality of reported genes data from human tissue of AD to control and to mouse model data

There was slightly more overlap in reported DE genes in human AD RNA-seq papers than in papers on mouse models of AD (Table S3).

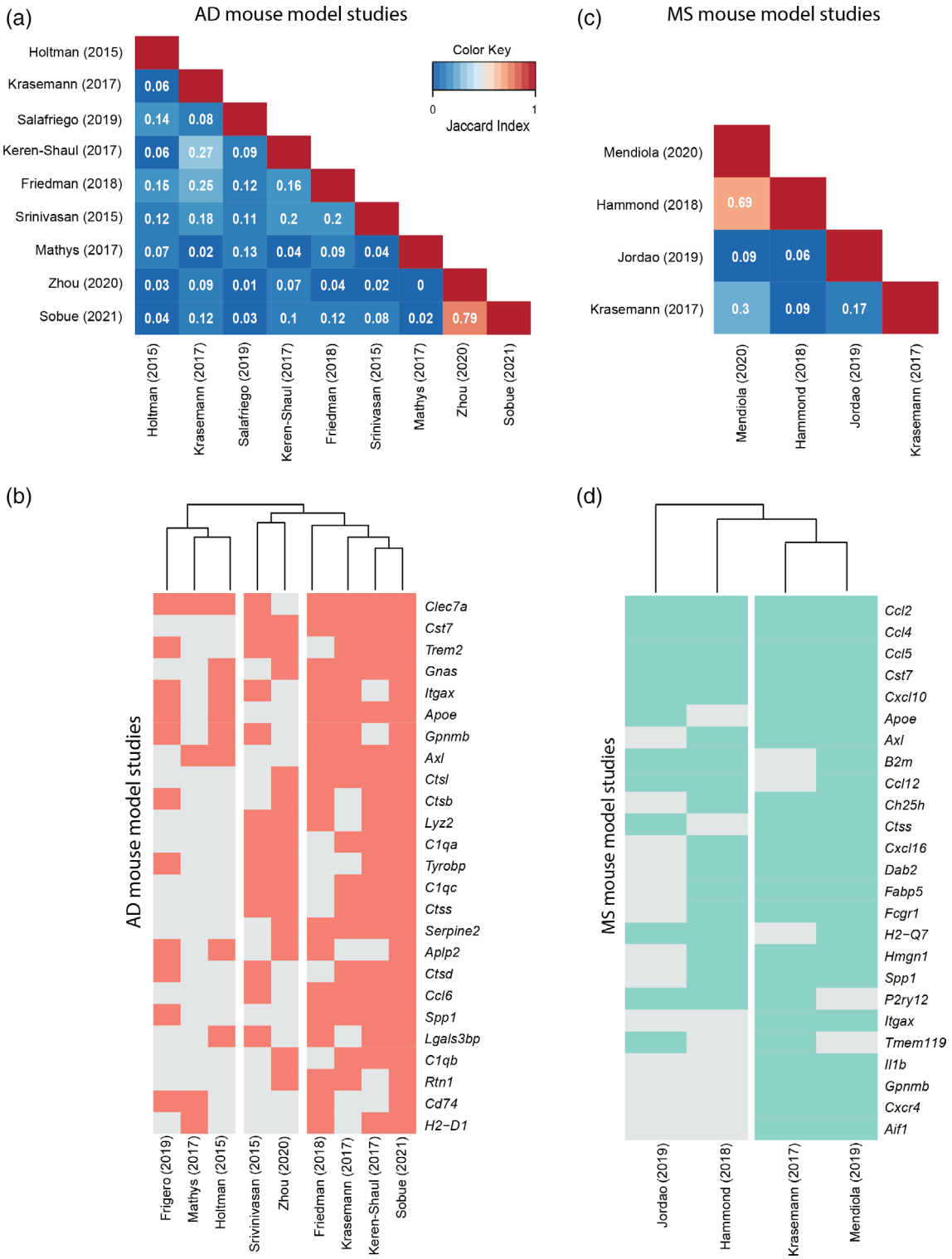

FIGURE 1 Overview of the overlap in microglia DE genes reported in papers investigating mouse models for AD and MS. (a) Jaccard indices indicating overlap of reported genes of 9 included AD mouse model. (b) Heatmap of the top 25 genes reported in at least two AD mouse model papers. Red boxes indicate that the gene was reported in the paper, gray boxes indicate an absence in reporting of the gene. (c) Jaccard indices indicating the overlap of reported genes of 4 included MS mouse model. (d) Heatmap of the top 25 genes reported in at least two MS mouse model papers. Green boxes indicate that the gene was reported in the paper, gray boxes indicate an absence in reporting of the gene

Based on the human core gene set, we found that Gerrits et al. (2021) and Zhou et al. (2020) reported almost the exact same genes (Jaccard index = 0.97, Figure 2(a)). Additionally, there was a large overlap in genes reported by Gerrits et al. (2021) and Mathys et al. (2019) (Jaccard index = 0.7, Figure 2(a)). These three studies all reported single cell RNA-seq data from frozen microglia nuclei though they used

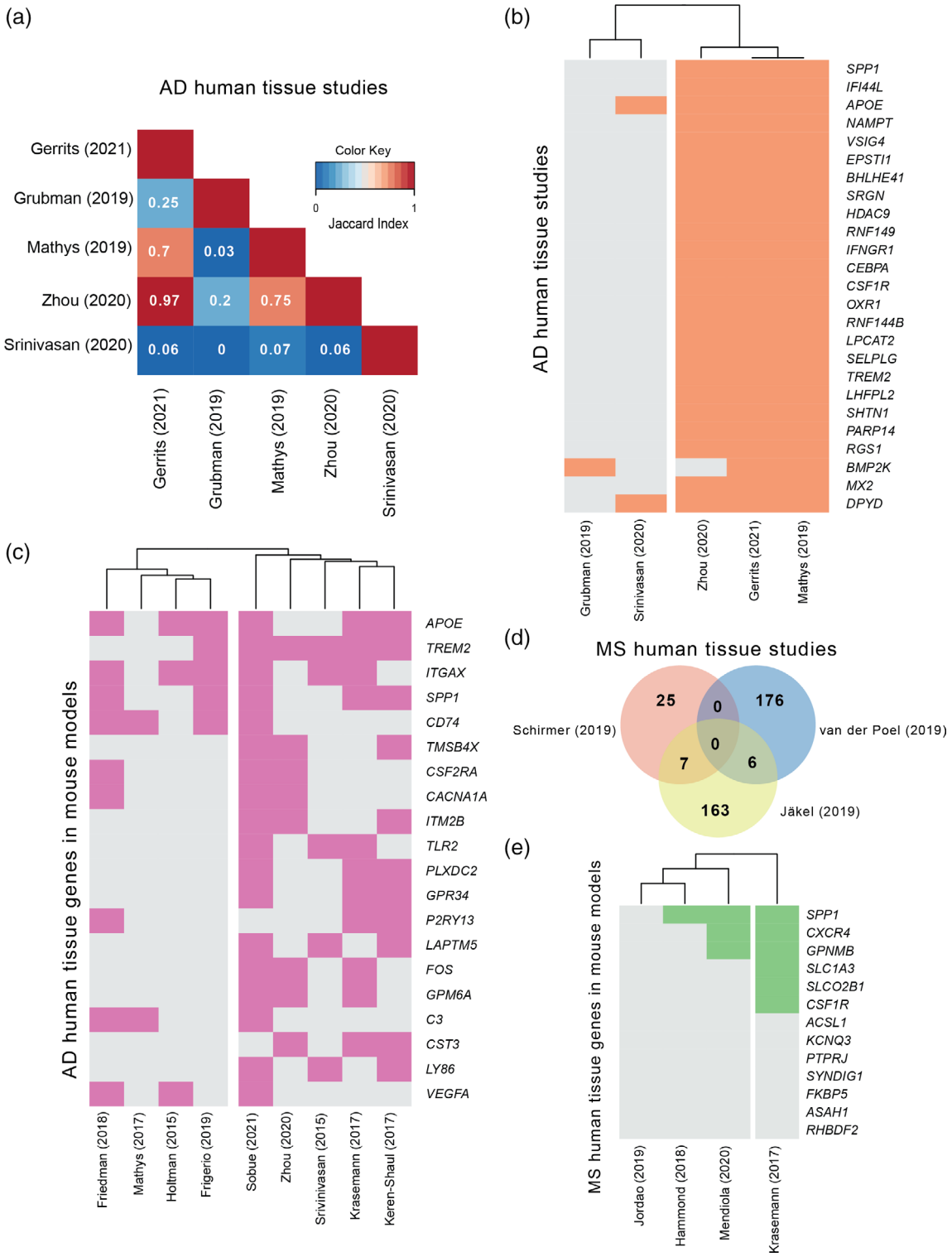

FIGURE 2 Overview of microglia DE genes reported in studies using AD and MS human tissue and the comparison of mouse model data to human data. (a) Jaccard indices indicating overlap in reported genes of 5 studies included reporting data from AD human tissue. (b) Heatmap of the top 25 genes reported in at least 2 human AD tissue studies. Orange boxes indicate that the gene was reported in the paper, gray boxes indicate an absence in reporting of the gene. (c) Heatmap genes reported in at least 2 AD mouse model papers and 2 AD human tissue papers. Pink boxes indicate that the gene was reported in the paper, gray boxes indicate an absence in reporting of the gene. (d) Venn diagram of the number of genes reported by papers showing data of human MS tissue. (e) Heatmap of genes reported in at least two MS mouse model papers and at least 2 human MS tissue papers

different tissue dissociation methods (Table 1). Concurrently, Srinivasan et al. (2020) used a different experimental set-up and showed almost no overlap in reported DE genes by other papers (Figure 2(a); Table 1). Despite using similar research methods as Gerrits et al. (2021) and Zhou et al. (2020), Grubman et al. (2019) showed little overlap in reported DE genes with other papers (Jaccard index = 0.2, Figure 2(a)). Possibly, this could be related to the lower number of cells sequenced by Grubman et al. (2019) (449 compared to 3986 and 150,000 cells for Zhou et al. (2020) and Gerrits et al. (2021), respectively). The top 25 most reported microglia DE genes (Figure 2(b)) were dominated by the overlap in reported genes between Mathys et al. (2019), Gerrits et al. (2021) and Zhou et al. (2020). Similar to the AD mouse model data, most studies utilized single-cell sequencing but reported a large range in the number of cells/nuclei sequenced (449–150,000, Table 1). Taken together, these data indicate slightly better overlap in reported microglia DE genes between AD human tissue studies, most probably due to the fact that several studies used very similar experimental set ups. Furthermore, GO-pathway analysis showed that there was no pathway enriched in a majority of the studies. This could be due to the absence of pathways enriched for the gene list derived from Srinivasan et al. (2020). In total we found 43 unique enriched pathways of which 5 were present in 2 out of 5 included studies (Table S3).

To assess the overlap of reported genes in AD mouse models with AD human tissue data, we compared the microglia DE genes reported by at least 2 mouse model papers with the DE genes reported in at least 2 human studies generating a list of genes mentioned in both mouse model and human tissue studies. Generally, reported genes of mouse models overlapped poorly with human tissue data. For example, the microglia DE gene most reported in mouse model papers (*Clec7a*) was not mentioned in at least two human RNA-seq studies (Figures 1(c) and 2(c)). The AD mouse model paper showing the most overlap in reported microglia DE genes with AD human tissue was described by Sobue et al. (2021) (Figure 2(c)).

3.4 | Commonality of reported genes data from human tissue of MS to control and to mouse model data

The three papers reporting sequencing data from human MS tissue microglia compared to control tissue showed very little overlap in reported DE genes. With only these three papers, we did not calculate Jaccard indices, instead we plotted a Venn-diagram. Schirmer et al. (2019) and van der Poel et al. (2019) showed no overlap in reported genes (Figure 2(d) and Table S4). The reported genes found overlapping between Jäkel et al. (2019) and Schirmer et al. (2019) were *ACSL1*, *SLC1A3*, *KCNQ3*, *PTPRJ*, *SYNDIG1*, *FKBP5*, and *ASAH1* (Table S4) and the reported genes overlapping between Jäkel et al. (2019) and van der Poel et al. (2019) were *CXCR4*, *GPNMB*, *SPP1*, *SLCO2B1*, *CSF1R*, and *RHBDF2* (Table S4). The lack of overlap could be due to the very different experimental set ups such as microglia isolation and sequencing method. In addition, different brain areas were used to isolate microglia from as well as the pathological characterization of the MS tissue was unclear (Table 2). We also observed a limited number of enriched GO-pathways with only one pathway found in two studies ("*cell activation*," Table S4). It must be noted that though Schirmer et al. (2019) and Jäkel et al. (2019) feature microglia RNA-seq data in their supplementary files, microglial gene expression was not a primary outcome of their study. Thus, it could be that their methods were not optimized to detect microglial gene expression. When comparing the genes mentioned by either van der Poel et al. (2019), Schirmer et al. (2019) or Jäkel et al. (2019) to the genes reported in at least two mouse model studies, we found only a limited number of overlapping microglia DE genes in papers reporting on mouse model data, with Jordão et al. (2019) reporting no genes reported in MS human tissue studies (Figure 2(e)).

4 | DISCUSSION

In the present study we provide an overview of reported microglia DE genes within manuscripts and their supplementary files of current microglia RNA-seq studies in MS or AD rodent or human tissue. We created this overview with the aim to assess which RNA-seq data are readily available from the manuscript, how comparable they are between manuscripts and what researchers need to be aware of when using reported RNA-seq data. Overall, we observed very limited overlap in reported microglia DE genes among various microglia RNA-seq studies on AD or on MS. Moreover, limited overlap in enriched GO-pathways in the reported DE genes per study indicate that different gene lists could also lead to different inferences of microglia function per study. This was especially the case for AD mouse-model studies and for human MS tissue studies. Factors that may influence the overlap in reported genes between studies will be discussed below.

4.1 | Experimental set up before sequencing

The differences in reported genes between the various currently available microglia RNA-seq studies could for a large part be attributed to differences in experimental set-up. Considerable overlap was observed between papers reporting DE gene data on human AD tissue with very similar experimental set up, that is, all reported single-cell sequencing data from extracted frozen nuclei of human cortical brain areas using the same sequencing method (Gerrits et al., 2021; Grubman et al., 2019; Zhou et al., 2020). In contrast, the only study reporting bulk-sequencing of FACS sorted microglia from human AD tissue (Srinivasan et al., 2020) showed considerably less overlap in reported genes. This indicates that gene expression data generated by RNA-seq is likely influenced by the experimental set up being used, with alike research designs leading to more similar gene expression results (Stark et al., 2019). For mouse model studies, we observed considerably more variation in the experimental set up. For example, AD mouse model studies used either different APP-mutation models

(i.e., *APP-PS1* or *AppNL-G-F*) or used various mice models of neurodegeneration (i.e., *CK-p25* or *5XFAD*). Additionally, the experimental procedures before sequencing differed between studies. These included, that is, the brain region dissected, method of microglia isolation or the library preparation kit used prior to sequencing. However strikingly, two studies reporting almost the exact same experimental set up (Friedman et al., 2018; Srinivasan et al., 2016) still showed minimal overlap in reported genes. In a similar vein, we observed that the two MS mouse model studies showing the most overlap in reported genes isolated microglia from completely different mouse models (EAE vs. LPC induced demyelination) and tissue (spinal cord vs. dissected white matter. Taken together, this suggests that whereas it is important to minimize set-up and analysis variation in studies investigating microglia gene expression in the context of AD or MS, other factors may also be at play, contributing to the lack of overlap in reported genes.

4.2 | Different requirements for reporting a gene

The large heterogeneity in reported DE genes may make it difficult to assess the use of current reported data within manuscripts for other researchers. Our results were partially influenced by the length of gene lists reported by papers. Papers reporting many different DE genes, for example, within an Excel file in the supplementary materials often showed more overlap with each other. Thus, if a paper reports only highly significantly regulated genes or genes with a high fold change, this could decrease its overlap in reported genes with other papers. This does not necessarily mean that papers reporting a lower number of genes under-report data: What determines when a gene is reported or not could be influenced by factors such as, that is, the specific focus or aims of the paper, the research method used (i.e., bulk RNA-seq compared to scRNA-seq) or restrictions on the length of supplementary materials set by the journal. The selection criteria used to report genes could differ between papers. For example, only genes with a log fold change of (−) 2 or higher, only genes that showed high statistical significant DE compared to control, or various combinations of the two could be reported. To increase usability of data for other researchers, it would be advisable to include data in i.e. the supplementary files without too many restrictions. Reporting all data, already within the manuscript itself, allows other researchers to set their own filters and focus on their specific gene(s) of interest.

4.3 | Pathology of animal models and human tissue

We also observed limited commonality in reported genes between papers describing mouse-model data and data from human tissue. For example, disease associated microglia (DAM) genes were first described in an impactful paper by Keren-Shaul et al. (2017) using a mouse model of AD. Microglia expressing high levels of DAM genes were identified as phagocytic cells responding to amyloid beta plaques, which were later also shown to be present, that is, other neurological diseases (Böttcher et al., 2019) or during neurodegeneration (Anderson et al., 2019). Since then, we have observed many references to this gene set in various (AD) microglia RNA-seq papers. However, microglia RNA-seq data from fresh human AD tissue did not show differential expression of these DAM genes compared to control (Alsema et al., 2020). Various explanations may be possible for this, including that post-mortem human AD tissue is often taken from end-stages of the diseases whereas mouse models often mimic relatively early pathological events. However, as postulated by Gerrits et al. (2021), it could also be that the isolation of amyloid plaque associated microglia from fresh human tissue may be difficult due to the location of microglia within or between the plaques affecting their isolation. Thus, whereas we did find some overlap in reporting of specific DAM related genes in human and mouse, in our study we found that there is a wealth of genes which are found in mouse models of disease but not in the human counterpart. It remains to be elucidated whether these genes indicate pathological differences between human or mouse or not. Until then, mouse model data should be considered as genes regulated in the brain featuring specific pathology, such as the formation of amyloid beta plaques inducing, that is, phagocytic DAM. Concurrently, RNA-seq data from human AD tissue is of interest to identify microglia phenotypes present at late stages of disease, characterized by amyloid beta plaques, hyperphosphorylated tau and substantial neurodegeneration (Spires-Jones & Hyman, 2014) which may induce a different microglia phenotype (Gerrits et al., 2021). To determine the use of the microglia RNA-seq data presented within a published manuscript, it is therefore important that the pathology present within the tissue used is extensively described such that it is clear which microglia function or response may be reflected in the gene expression data. We observed that pathological data of especially animal models is often not extensively reported. This is perhaps most clear in the limited amount of MS microglia sequencing studies currently published. For example, it is known that the EAE mouse model of MS, described in two of the four manuscripts (Jordão et al., 2019; Mendiola et al., 2020) used in this study leads to some demyelination in white matter areas such as the corpus callosum and spinal cord white matter, but much less in gray matter areas. Yet, both studies describing data from the EAE mouse model used either the whole brain or the whole spinal cord. In addition, the inflammatory status of the demyelinated areas is often not reported. Therefore, it is important to realize that these data are indicative of the total microglia response possibly present in the EAE brain, but not of demyelinated brain areas, in which the microglia response may be different. This is substantiated by the observation that though most mouse model studies showed a lack of overlap in reported DE genes in human MS tissue studies, the genes that were overlapping were present in normal-appearing (i.e., non-demyelinated) brain areas of patients with MS (van der Poel et al., 2019).

Surprisingly, though a role for microglia in MS pathology is quite clear (O'Loughlin et al., 2018) there are less studies reporting microglia RNA-seq data of MS tissue compared to AD. Especially for the human tissue data, van der Poel et al. (2019) is the only study specifically

focusing on microglia, Schirmer et al. (2019) and Jäkel et al. (2019) both focus on different cell types but present microglia data in their supplementary files, albeit relatively low numbers. Moreover, all three studies feature data from different brain areas. As microglia function or states can differ depending on the local environment, for example, gray versus white matter or demyelinated versus myelinated tissue (Bø et al., 2003; Geurts & Barkhof, 2008; Prins et al., 2015; Van Wageningen et al., 2019; Zrzavy et al., 2017), the lack of overlap between the three human MS RNA-seq studies is perhaps therefore not surprising. To disentangle the role of microglia in specific aspects of MS pathology such as demyelination, inflammation or neurodegeneration, more studies are needed with similar experimental set up and reporting, as extensively as possible, the pathology present within samples used for sequencing.

4.4 | Considerations

It is important to note that from our analysis we cannot make any statements about the overlap of differentially expressed genes between the entire generated datasets of several studies as this would require a bioinformatic reanalysis of all current deposited datasets. However, we can make inferences about the reported datasets within the manuscript and within supplementary data files, which is what most researchers without access to extensive bioinformatics analysis will use. Based on our observations, there is currently a large variation in which genes are reported within manuscripts, even when very similar research designs, animal models or tissues are used. Yet, we also observe that the lack of overlap in reported microglia DE genes can possibly be mitigated by standardizing the way genes are reported within the manuscript: Overall, overlap in reported genes between manuscripts is increased when many genes are reported. This can be achieved by reducing the selection criteria for genes that are reported within the manuscript or in the supplementary files. For example: the complete gene set found to be differentially expressed should be included, without extra restrictions on e.g. the minimal fold change. If this is not possible, it should be mentioned in the manuscript or supplementary files which exact threshold was used to select genes (i.e., an FDR or adjusted p value <.05 and log fold change >2). If no supplementary files are allowed, these analyzed gene sets could be added to the unanalyzed (raw) data deposited in databases such as NCBI GEO. Perhaps more importantly, as we observed that microglia RNA-seq data are reflective of the specific pathology present within the tissue used, more extensive reporting on the specific pathology present within tissue used for sequencing (or if not possible, adjacent tissue) within the manuscript would facilitate interpretation of the specific gene sets created. This can include i.e. amount of amyloid beta plaques and hyperphosphorylated tau tangles for AD or presence or amount of inflammation and demyelination or remyelination in MS tissue. For researchers wanting to use microglia gene expression data, the above-mentioned points are important to keep in mind when selecting genes as outcome measurements for further experiments.

In conclusion, one of the main advantages of using RNA-seq is the generation of large unbiased datasets which can be used either to elucidate novel genes regulated in specific pathological conditions or as a reference for further experiments and thus, these data can be very powerful for scientists. Through various database initiatives, the unanalyzed (raw) data in the form of count matrices or FASTA files is often available to those with knowledge of bioinformatics. However, this is not the case for all research groups. In addition, reanalyzing all datasets can be time-consuming even for a dedicated bioinformatician. Thus, in order to facilitate and spread the knowledge generated by RNA-seq datasets, in this study we provide insight into how current microglia gene expression data from MS and AD tissue is reported and some considerations which may be adopted to increase use of reported gene expression data for other researchers with or without bioinformatics experience.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

All data used is available in article supplementary material or within the manuscript.

ORCID

Thecla A. van Wageningen https://orcid.org/0000-0002-5304-9487
Bart J. L. Eggen https://orcid.org/0000-0001-8941-0353

REFERENCES

Alsema, A. M., Jiang, Q., Kracht, L., Gerrits, E., Dubbelaar, M. L., Miedema, A., … Eggen, B. J. L. (2020). Profiling microglia from Alzheimer's disease donors and non-demented elderly in acute human postmortem cortical tissue. *Frontiers in Molecular Neuroscience*, *13*, 134. https://doi.org/10.3389/fnmol.2020.00134

Anderson, S. R., Roberts, J. M., Zhang, J., Steele, M. R., Romero, C. O., Bosco, A., & Vetter, M. L. (2019). Developmental apoptosis promotes a disease-related gene signature and independence from CSF1R signaling in retinal microglia. *Cell Reports*, *27*(7), 2002–2013.e2005.

Bennett, M. L., Bennett, F. C., Liddelow, S. A., Ajami, B., Zamanian, J. L., Fernhoff, N. B., Mulinyawe, S. B., Bohlen, C. J., Adil, A., Tucker, A., Weissman, I. L., Chang, E. F., Li, G., Grant, G. A., Hayden Gephart, M. G., & Barres, B. A. (2016). New tools for studying microglia in the mouse and human CNS. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(12), E1738–E1746. http://www.pnas.org/content/113/12/E1738.full.pdf

Bø, L., Vedeler, C. A., Nyland, H., Trapp, B. D., & Mørk, S. J. (2003). Intracortical multiple sclerosis lesions are not associated with increased lymphocyte infiltration. *Multiple Sclerosis Journal*, *9*(4), 323–331.

Böttcher, C., Schlickeiser, S., Sneeboer, M. A., Kunkel, D., Knop, A., Paza, E., Fidzinski, P., Kraus, L., Snijders, G. J. L., Kahn, R. S., Schulz, A. R., Mei, H. E., NBB-Psy, Hol, E. M., Siegmund, B., Glauben, R., Spruth, E. J., de Witte, L. D., & Priller, J. (2019). Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. *Nature Neuroscience*, *22*(1), 78–90.

Butovsky, O., & Weiner, H. L. (2018). Microglia signatures and their role in health and disease. *Nature Reviews Neuroscience*, *19*(10), 622–635.

Butovsky, O., Jedrychowski, M. P., Moore, C. S., Cialic, R., Lanser, A. J., Gabriely, G., Koeglsperger, T., Dake, B., Wu, P. M., Doykan, C. E., Fanek, Z., Liu, L. P., Chen, Z., Rothstein, J. D., Ransohoff, R. M., Gygi, S. P., Antel, J. P., & Weiner, H. L. (2014). Identification of a unique TGF-beta-dependent molecular and functional signature in microglia. *Nature Neuroscience*, *17*(1), 131–143. http://www.ncbi.nlm.nih.gov/pubmed/24316888, https://doi.org/10.1038/nn.3599

Chung, N. C., Miasojedow, B., Startek, M., & Gambin, A. (2019). Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. *BMC Bioinformatics*, *20*(15), 644. https://doi.org/10.1186/s12859-019-3118-5

Friedman, B. A., Srinivasan, K., Ayalon, G., Meilandt, W. J., Lin, H., Huntley, M. A., Cao, Y., Lee, S. H., Haddick, P. C. G., Ngu, H., Modrusan, Z., Larson, J. L., Kaminker, J. S., van der Brug, M., & Hansen, D. V. (2018). Diverse brain myeloid expression profiles reveal distinct microglia activation states and aspects of Alzheimer's disease not evident in mouse models. *Cell Reports*, *22*(3), 832–847.

Frigerio, C. S., Wolfs, L., Fattorelli, N., Thrupp, N., Voytyuk, I., Schmidt, I., Mancuso, R., Chen, W.-T., Woodbury, M. E., Srivastava, G., Möller, T., Hudry, E., Das, S., Saido, T., Karran, E., Hyman, B., Perry, V. H., Fiers, M., & De Strooper, B. (2019). The major risk factors for Alzheimer's disease: Age, sex, and genes modulate the microglia response to Aβ plaques. *Cell Reports*, *27*(4), 1293–1306.e1296.

Gerrits, E., Brouwer, N., Kooistra, S. M., Woodbury, M. E., Vermeiren, Y., Lambourne, M., Mulder, J., Kummer, M., Möller, T., Biber, K., Dunnen, W. F. A., de Deyn, P. P., Eggen, B. J. L., & Boddeke, E. W. G. M. (2021). Distinct amyloid-β and tau-associated microglia profiles in Alzheimer's disease. *Acta Neuropathologica*, *141*, 681–696. https://doi.org/10.1007/s00401-021-02263-w

Geurts, J. J., & Barkhof, F. (2008). Grey matter pathology in multiple sclerosis. *The Lancet Neurology*, *7*(9), 841–851.

Grubman, A., Chew, G., Ouyang, J. F., Sun, G., Choo, X. Y., McLean, C., Simmons, R. K., Buckberry, S., Vargas-Landin, D. B., Poppe, D., Pflueger, J., Lister, R., Rackham, O. J. L., Petretto, E., & Polo, J. M. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience*, *22*(12), 2087–2097.

Hammond, T. R., Dufort, C., Dissing-Olesen, L., Giera, S., Young, A., Wysoker, A., Walker, A. J., Gergits, F., Segel, M., Nemesh, J., Marsh, S. E., Saunders, A., Macosko, E., Ginhoux, F., Chen, J., Franklin, R. J. M., Piao, X., McCarroll, S., & Stevens, B. (2019). Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. *Immunity*, *50*(1), 253–271.e256.

Hickman, S. E., Kingery, N. D., Ohsumi, T. K., Borowsky, M. L., Wang, L. C., Means, T. K., & El Khoury, J. (2013). The microglia sensome revealed by direct RNA sequencing. *Nature Neuroscience*, *16*(12), 1896–1905. http://www.ncbi.nlm.nih.gov/pubmed/24162652, https://doi.org/10.1038/nn.3554

Holtman, I. R., Raj, D. D., Miller, J. A., Schaafsma, W., Yin, Z., Brouwer, N., Wes, P. D., Möller, T., Orre, M., Kamphuis, W., Hol, E. M., Boddeke, E. W. G. M., & Eggen, B. J. L. (2015). Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: A co-expression meta-analysis. *Acta Neuropathologica Communications*, *3*(1), 31. https://doi.org/10.1186/s40478-015-0203-5

Jäkel, S., Agirre, E., Mendanha Falcão, A., van Bruggen, D., Lee, K. W., Knuesel, I., Malhotra, D., Ffrench-Constant, C., Williams, A., & Castelo-Branco, G. (2019). Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature*, *566*(7745), 543–547.

Jordão, M. J. C., Sankowski, R., Brendecke, S. M., Locatelli, G., Tai, Y.-H., Tay, T. L., Schramm, E., Armbruster, S., Hagemeyer, N., Groß, O., Mai, D., Çiçek, Ö., Falk, T., Kerschensteiner, M., Grün, D., & Prinz, M. (2019). Single-cell profiling identifies myeloid cell subsets with distinct fates during neuroinflammation. *Science*, *363*(6425), eaat7554.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T. K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., Itzkovitz, S., Colonna, M., Schwartz, M., & Amit, I. (2017). A unique microglia type associated with restricting development of Alzheimer's disease. *Cell*, *169*(7), 1276–1290.e1217.

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., & Winter, D. R. (2018). A beginner's guide to analysis of RNA sequencing data. *American Journal of Respiratory Cell and Molecular Biology*, *59*(2), 145–157.

Krasemann, S., Madore, C., Cialic, R., Baufeld, C., Calcagno, N., El Fatimy, R., Beckers, L., O'Loughlin, E., Xu, Y., Fanek, Z., Greco, D. J., Smith, S. T., Tweet, G., Humulock, Z., Zrzavy, T., Conde-Sanroman, P., Gacias, M., Weng, Z., Chen, H., … Butovsky, O. (2017). The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity*, *47*(3), 566–581.e569.

Mathys, H., Adaikkan, C., Gao, F., Young, J. Z., Manet, E., Hemberg, M., de Jager, P. L., Ransohoff, R. M., Regev, A., & Tsai, L.-H. (2017). Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. *Cell Reports*, *21*(2), 366–380.

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A. J., Ransohoff, R. M., Hafler, B. P., Bennett, D. A., Kellis, M., & Tsai, L.-H. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, *570*(7761), 332–337. https://doi.org/10.1038/s41586-019-1195-2

Mendiola, A. S., Ryu, J. K., Bardehle, S., Meyer-Franke, A., Ang, K. K.-H., Wilson, C., Baeten, K. M., Hanspers, K., Merlini, M., Thomas, S., Petersen, M. A., Williams, A., Thomas, R., Rafalski, V. A., Meza-Acevedo, R., Tognatta, R., Yan, Z., Pfaff, S. J., Machado, M. R., … Akassoglou, K. (2020). Transcriptional profiling and therapeutic targeting of oxidative stress in neuroinflammation. *Nature Immunology*, *21*(5), 513–524. https://doi.org/10.1038/s41590-020-0654-0

O'Loughlin, E., Madore, C., Lassmann, H., & Butovsky, O. (2018). Microglia phenotypes and functions in multiple sclerosis. *Cold Spring Harbor Perspectives in Medicine*, *8*(2), a028993.

Prins, M., Schul, E., Geurts, J., van der Valk, P., Drukarch, B., & van Dam, A. M. (2015). Pathological differences between white and grey matter multiple sclerosis lesions. *Annals of the New York Academy of Sciences*, *1351*, 99–113. http://www.ncbi.nlm.nih.gov/pubmed/26200258, https://doi.org/10.1111/nyas.12841

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). G: Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198.

Schirmer, L., Velmeshev, D., Holmqvist, S., Kaufmann, M., Werneburg, S., Jung, D., Vistnes, S., Stockley, J. H., Young, A., Steindel, M., Tung, B., Goyal, N., Bhaduri, A., Mayer, S., Engler, J. B., Bayraktar, O. A., Franklin, R. J. M., Haeussler, M., Reynolds, R., … Rowitch, D. H. (2019). Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature*, *573*(7772), 75–82. https://www.ncbi.nlm.nih.gov/pubmed/31316211, https://doi.org/10.1038/s41586-019-1404-z

Sobue, A., Komine, O., Hara, Y., Endo, F., Mizoguchi, H., Watanabe, S., Murayama, S., Saito, T., Saido, T. C., Sahara, N., Higuchi, M., Ogi, T., & Yamanaka, K. (2021). Microglia gene signature reveals loss of homeostatic microglia associated with neurodegeneration of Alzheimer's disease. *Acta Neuropathologica Communications*, *9*(1), 1–17.

Spires-Jones, T. L., & Hyman, B. T. (2014). The intersection of amyloid beta and tau at synapses in Alzheimer's disease. *Neuron*, *82*(4), 756–771.

Srinivasan, K., Friedman, B. A., Etxeberria, A., Huntley, M. A., van der Brug, M., Foreman, O., Paw, J. S., Modrusan, Z., Beach, T. G., Serrano, G. E., & Hansen, D. V. (2020). Alzheimer's patient microglia exhibit enhanced aging and unique transcriptional activation. *Cell Reports*, *31*(13), 107843.

Srinivasan, K., Friedman, B. A., Larson, J. L., Lauffer, B. E., Goldstein, L. D., Appling, L. L., Borneo, J., Poon, C., Ho, T., Cai, F., Steiner, P., van der Brug, M. P., Modrusan, Z., Kaminker, J. S., & Hansen, D. V. (2016). Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. *Nature Communications*, *7*(1), 11295. https://doi.org/10.1038/ncomms11295

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics*, *20*(11), 631–656.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377–382. https://doi.org/10.1038/nmeth.1315

van der Poel, M., Ulas, T., Mizee, M. R., Hsiao, C.-C., Miedema, S. S., Adelia, Schuurman, K. G., Helder, B., Tas, S. W., Schultze, J. L., Hamann, J., & Huitinga, I. (2019). Transcriptional profiling of human microglia reveals grey–white matter heterogeneity and multiple sclerosis-associated changes. *Nature Communications*, *10*(1), 1139.

Van Wageningen, T. A., Vlaar, E., Kooij, G., Jongenelen, C. A., Geurts, J. J., & Van Dam, A.-M. (2019). Regulation of microglia TMEM119 and P2RY12 immunoreactivity in multiple sclerosis white and grey matter lesions is dependent on their inflammatory environment. *Acta Neuropathologica Communications*, *7*(1), 206.

Voet, S., Prinz, M., & van Loo, G. (2019). Microglia in central nervous system inflammation and multiple sclerosis pathology. *Trends in Molecular Medicine*, *25*(2), 112–123.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B., & Galili, T. (2009). *gplots: Various R programming tools for plotting data*. *R package version*, *2*(4), 1.

Zhou, Y., Song, W. M., Andhey, P. S., Swain, A., Levy, T., Miller, K. R., Poliani, P. L., Cominelli, M., Grover, S., Gilfillan, S., Cella, M., Ulland, T. K., Zaitsev, K., Miyashita, A., Ikeuchi, T., Sainouchi, M., Kakita, A., Bennett, D. A., Schneider, J. A., … Colonna, M. (2020). Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nature Medicine*, *26*(1), 131–142. https://doi.org/10.1038/s41591-019-0695-9

Zrzavy, T., Hametner, S., Wimmer, I., Butovsky, O., Weiner, H. L., & Lassmann, H. (2017). Loss of 'homeostatic' microglia and patterns of their activation in active multiple sclerosis. *Brain*, *140*(7), 1900–1913. https://www.ncbi.nlm.nih.gov/pubmed/28541408, https://doi.org/10.1093/brain/awx113

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: van Wageningen, T. A., Gerrits, E., Palacin i Bonson, S., Huitinga, I., Eggen, B. J. L., & van Dam, A.-M. (2021). Exploring reported genes of microglia RNA-sequencing data: Uses and considerations. *Glia*, *69*(12), 2933–2946. https://doi.org/10.1002/glia.24078