



OPEN ACCESS

EDITED BY

Bo-Wei Zhao,
Zhejiang University, China

REVIEWED BY

Youtao Lu,
University of Pennsylvania, United States
Huifeng Zhao,
Zhejiang University, China

*CORRESPONDENCE

Matthew H. Bailey,
✉ matthew.bailey@byu.edu

[†]These authors share senior authorship

RECEIVED 11 August 2025

ACCEPTED 14 October 2025

PUBLISHED 04 November 2025

CITATION

Rahm KK, Kinghorn BS, Moody MJ, Stone BC,
Strong KC, Kim BS, Chang YJ, Sleight SN,
Nitz AA, Hansen DV and Bailey MH (2025)
Cellf-deception: human microglia clone 3
(HMC3) cells exhibit more astrocyte-like than
microglia-like gene expression.
Front. Bioinform. 5:1681811.
doi: 10.3389/fbinf.2025.1681811

COPYRIGHT

© 2025 Rahm, Kinghorn, Moody, Stone,
Strong, Kim, Chang, Sleight, Nitz, Hansen and
Bailey. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Cellf-deception: human microglia clone 3 (HMC3) cells exhibit more astrocyte-like than microglia-like gene expression

Kylee K. Rahm¹, Branden S. Kinghorn², Myanna J. Moody²,
Ben C. Stone², Kenton C. Strong³, Brian S. Kim²,
Yen Jou Chang⁴, Samantha N. Sleight⁴, Alyssa A. Nitz²,
David V. Hansen⁴ and Matthew H. Bailey^{2*†}

¹Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT, United States, ²Department of Biology, Brigham Young University, Provo, UT, United States, ³Computer Science Department, Brigham Young University, Provo, UT, United States, ⁴Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, United States

Introduction: Recent advances in Alzheimer's research suggest that the brain's immune system plays a critical role in the development and progression of this devastating disease. Microglial cells are vital as immune cells in the brain's defense system. Human Microglia Clone 3 (HMC3) is a cell line developed as a promising experimental model to understand the role of microglial cells in human diseases including Alzheimer's and other neurodegenerative diseases. The frequency of HMC3 cell usage has increased in recent years, with the idea that this cell line could serve as a convenient model for human microglial cell functions.

Methods: We utilized gene-pair ratios from bulk and single-cell RNA sequencing (scRNA-seq) expression data to create predictive models of cell-type origins.

Results: Our model reveals that the HMC3 cell line represents various cell types, with the highest cell similarity score relating to astrocytes, not microglia.

Discussion: These findings suggest that the HMC3 cell line is not a reliable human microglia model and that extreme caution should be taken when interpreting the results of studies using the HMC3 cell line.

KEYWORDS

HMC3 cells, iPSC-derived microglia (iMG), microglia, astrocytes, cell-type classification, Alzheimer's disease

Introduction

As Alzheimer's disease (AD) is a growing global health concern and a leading cause of death in the United States, it is essential to better understand this disease (Weuve et al., 2014; James et al., 2014; Author anonymous, 2024). AD causes neuronal breakdown and brain atrophy (Cedres et al., 2020). Recent studies suggest that the brain's immune system may play a key role in developing AD (Weiner, 2025). Microglial cells are the first line of the brain's innate immune defense. These cells maintain brain homeostasis and, when working properly, find cells that are diseased or injured (Bohlen et al., 2019; Condello et al., 2015; Vainchtein and Molofsky, 2020). Clustering and chronic activation of microglial cells around β -amyloid plaques have long hinted at potential roles for these cells in AD

progression (McGeer et al., 1993), and recent identification in genome-wide association studies of many AD risk genes with microglia-specific expression has underscored that microglia are a key cell type that governs AD pathogenesis (Hansen et al., 2018). Many genetic and environmental factors can alter the activity and responses of microglial cells, and the mechanisms by which microglia promote or restrain AD development and progression are not fully understood and require further investigation.

AD-relevant microglial activities such as chemotaxis, phagocytosis, lysosome function, and proteostasis can be modeled *in vitro* using cultured cells. Primary human microglial cells isolated from brain tissue are not commonly utilized since fresh brain tissue is not readily available. Human microglia-like cells known as iMG or iMGL (iPSC-derived microglia-like) cells can be differentiated from induced pluripotent stem cell (iPSC) lines and are increasingly recognized as the best available alternative to primary microglial cells (Wickstead, 2023; Abud et al., 2017). However, iPSC maintenance and iMG differentiation are expensive, laborious, and lengthy procedures. Another alternative that allows researchers to produce their (pseudo)cell type of interest in unlimited numbers is to convert primary cultured cells into “immortalized” cell lines with unlimited proliferative potential. Although immortalized cells are not the same as their primary cell progenitors, they may nonetheless serve as useful substitutes that are convenient and inexpensive to culture, and they are more amenable to genetic manipulation to study the functionality of the individual genes.

Currently, multiple immortalized lines of murine microglial cells are available, including the widely used BV-2 and N9 cell lines, both immortalized using retroviral oncogenes (Blasi et al., 1990; Righi et al., 1989; Timmerman, Burm, and Bajramovic, 2018). However, interspecific differences between human and mouse immune signaling argue for the use of human cells when possible. For example, the microglia-expressed AD risk gene and therapeutic target *CD33* has no murine ortholog (Bhattacharjee et al., 2019). Very few immortalized human microglia cell lines have been produced, but one cell line being increasingly utilized by researchers is the Human Microglia Clone 3 (HMC3) cell line (Figure 1A).

The HMC3 cell line was created by immortalizing human embryonic microglial cultures with SV40 antigen to further scientific investigation of how microglial cells impact human conditions including Alzheimer's and other neurodegenerative diseases (Janabi et al., 1995). Notably, HMC3 cells have also been distributed and used under different names including CHME-3, CHME-5, and C13-NJ cells (Russo et al., 2018). In 2016, it was discovered that ostensible CHME-5 cells being used by several labs at that time were in fact rat-derived cells (Garcia-Mesa et al., 2017), prompting the American Type Culture Collection (ATCC) to authenticate the human origin of the HMC3 cell line (product #CRL-3304).

Motivated by interest in using HMC3 cells as potential research tools, we spot-checked some existing HMC3 RNA-seq datasets from Gene Expression Omnibus to validate the expression of common myeloid cell markers including *ITGAM* (CD11b), *PTPRC* (CD45), *SYK*, *TYROBP* (DAP12), *CX3CR1*, and Fc receptors. To our surprise, these markers were not expressed, prompting us to examine the research literature for insights as to whether HMC3 cells should be classified as microglia-derived. Indeed, some researchers have noted the lack of expected marker expression in HMC3 cells (Rawat and Spector, 2017; Rai et al., 2020), while others reported that HMC3

cells' transcriptional profile more closely resembled that of U87 cells (Quiroga et al., 2022)—a glioblastoma-derived cell line commonly used to represent astrocyte biology—than profiles of myeloid lineage cells including iMG cells, the monocytic leukemia-derived THP-1 cell line, or primary cultures of microglia, macrophages, or monocytes (see our similar principal component analysis in Figure 1B). If the research community increasingly uses the HMC3 cell line to study the role of microglial cells in neurodegenerative diseases, we must present a definitive classification of what type of brain cell it best represents.

The purpose of this paper is to computationally assess the myeloid nature of the HMC3 cell line, specifically distinguishing between microglia and astrocytes or other CNS cell types. Although astrocytes and microglia are physically different and easy to distinguish while examining morphology *in vivo* (Vainchtein and Molofsky, 2020), transformed or immortalized cells cultured *in vitro* have less distinct morphologies and can have altered traits due to the immortalization process (Kaur and Dufour, 2012). Single-cell and bulk RNA sequencing are powerful tools for distinguishing cell types by measuring transcriptome-wide differences in gene expression levels, providing a robust alternative to morphology-based observations (Haque et al., 2017; Mukamel and Ngai, 2019; Ofengeim et al., 2017).

Our strategy for HMC3 cell classification used publicly available RNA-seq expression data from multiple studies to determine the proper placement of the HMC3 cell line amongst different lineages of cells within the brain. Microglial cells and other CNS cell types have been sequenced many times for study and classification (Spurgat, 2022), and the resulting datasets are accessible in databases such as Gene Expression Omnibus (GEO) and Genotype-Tissue Expression Portal (GTEx) (Gerrits et al., 2020; Keil et al., 2018).

Our primary goal was to understand if the HMC3 line could be confidently characterized as microglia-derived cells and subsequently used in researching the mechanisms of neurodegenerative diseases such as Alzheimer's. Given that gene expression directly influences phenotypic features of a cell, including its behavior, we developed two independent Random Forest classifiers to investigate the cellular identity of the HMC3 cell line using gene-ratio comparisons. The first classifier was trained on primary human cells from multiple studies to establish gene-pair rules for distinguishing among cell types and to generate prediction scores reflecting cell-type similarity. When the HMC3 cell line was analyzed with this classifier, it exhibited variability in predicted cell-type scores, with the highest similarity observed for astrocytes. The second classifier was trained on the extensive cohort of cell lines collected by the DepMap project (Tsherniak et al., 2017). Using this classifier, HMC3 cells were predicted to have greater similarity with cells of the central nervous system as opposed to the expected myeloid lineage.

Materials and Methods

PubMed trend analysis of “HMC3 and microglia” publications

We conducted a comprehensive search on the PubMed (pubmed.ncbi.nlm.nih.gov/) research database to identify all articles published between 2009 and 12 July 2025 that included both “HMC3

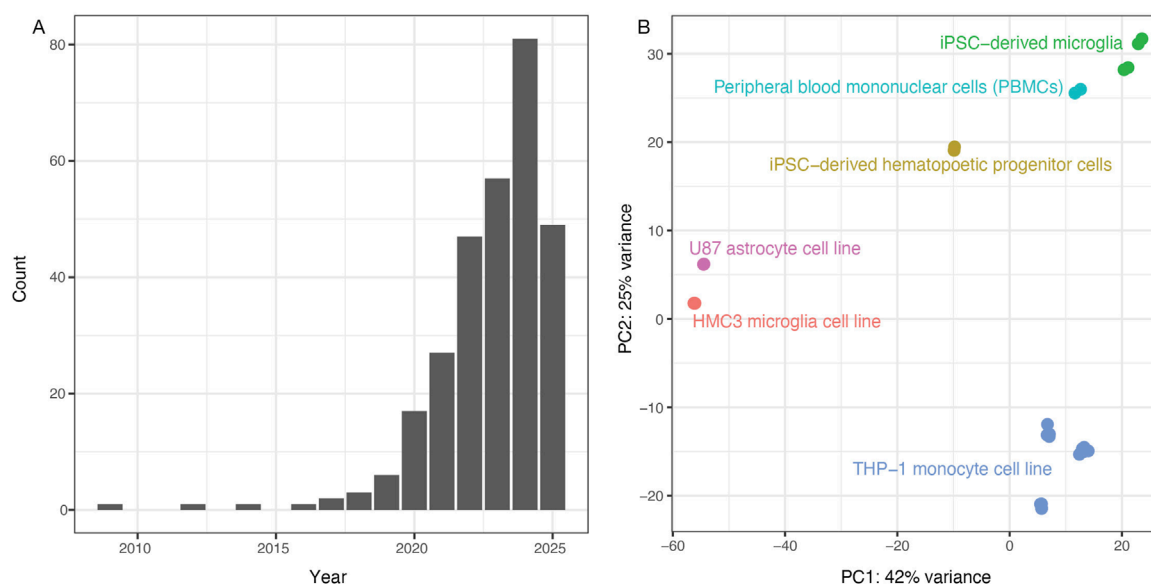


FIGURE 1

Analysis of previous literature justifies the study. (A) PubMed search results for “(HMC3) AND (microglia)” show sharply increasing numbers of publications since 2020, with 57 articles in 2023, 81 in 2024 and 46 so far in 2025 as of July 12. (B) Principal components were recalculated using brain cell types from Quiroga et al., 2022 (GSE181153). We observed an unexpected clustering of the HMC3 microglial cell line with the U87 astrocyte cell line (red and pink clusters). Note: Adapted from *Synthetic amyloid beta does not induce a robust transcriptional response in innate immune cell culture systems*, (I) Y. Quiroga et al., 2022, <http://creativecommons.org/licenses/by/4.0/> (Quiroga et al., 2022).

AND Microglia” as search terms. After downloading the search results, we compiled the number of relevant publications per year and created a table summarizing the annual publication counts.

Data access and cleaning

RNA sequencing datasets derived from healthy human tissues, a variety of brain cell types, and various cell lines were collected from GEO, DepMap, and several papers (Supplementary Table S1). Control and diseased samples from each dataset were used as part of our analysis. All datasets used the HUGO gene symbols, and any datasets that used the human Ensembl gene nomenclature were converted through the R IDConverter package to enable gene-pair comparisons among different datasets (Wang et al., 2021). Additionally, we restricted genes within the datasets to those where 80% or higher of the samples had a non-zero observation in the specified gene. This restricted our training datasets to 8,723 genes.

Principal component analysis

Figure 1B in this manuscript mimics Figure 1 from Quiroga et al. (GSE181153) but with fewer cell type comparisons (Quiroga et al., 2022). First, the gene-counts table was converted to transcripts per million. As performed by Quiroga et al., only genes with greater than 100 transcripts per million were kept in the analysis. Samples were then restricted to only include control non-treated samples, i.e., only “NONE_NONE” labels were kept. Additionally, To recreate the Quiroga PCA, iPSCs were not included in the plot despite the raw data including iPSCs. Of note, when

the iPSC lines were included, they cluster with U87a and HMC3 lines (Supplementary Figure S1). Once samples and genes were removed, gene count data were processed using DES (Love et al., 2014), followed by variant stabilization (‘vst()’ function). Following the variant stabilization step, the ‘plotPCA()’, a native function in DES, was used to generate the Principal Component plots.

The primary cell Random Forest classifier

A primary cell Random Forest classifier was developed using the multiclassPairs R package (v0.4.3), which implements a rule-based classification framework with gene-pair comparisons (Marzouka and Eriksson, 2021). The classifier was built from five publicly available datasets, comprising a total of 258 healthy and diseased human samples. From the Galatro et al. dataset (GSE99074), 65 microglia samples were included. The Gosselin et al. dataset (phs001373. v1. p1.) contributed 46 microglia and 13 monocyte samples. The Srinivasan et al. dataset (GSE125050) provided 19 adult astrocytes, 27 endothelial, 25 microglia, and 42 neuron samples. From the Zhang et al. dataset (GSE73721), 9 adult astrocyte, 4 microglia, and 1 neuron sample were included. Lastly, the Costa-Verdera et al. dataset (GSE253820) contributed 4 adult astrocyte and 3 neuron samples. After preprocessing, each sample was annotated with its respective cell type and dataset of origin. All data were merged into a single gene expression matrix, with genes as rows and samples as columns, for input into the multiclassPairs workflow. It should be noted that data normalization and batch effect correction was not performed on the combined dataset as the only comparisons of gene expression values occurred within samples, and never across samples, in all further analyses using this dataset.

The ‘ReadData’ function from the multiclassPairs R package was implemented to structure the combined matrix into an appropriate format with correct labeling for downstream analysis. The data were partitioned into training (60%) and testing (40%) sets, ensuring no sample overlap between sets.

To reduce dimensionality and improve classifier performance, gene selection was performed using the ‘sort_genes_RF’ function and the ‘rank_data’ parameter set to TRUE. This function ranks genes within samples based on their importance in differentiating cell types, employing a Random Forest-based feature ranking strategy of importance scores computed by the embedded ranger package. Two methods are employed in the function to rank genes: an “altogether” method and a “one-vs-rest” method. In the “altogether” method, genes are ranked based on their ability to differentiate all cell types from each other. In the “one-vs-rest” method, for each cell type, genes are ranked based on their ability to differentiate that cell type from the others. A total of 2,000 trees were used in the gene sorting step. The top-ranked genes were selected for further analysis using the ‘summary_genes_RF’ function, which indicated that 85 genes from the “altogether” category and 100 genes per class from the “one-vs-rest” category provided optimal rule coverage for downstream training.

From the unique set of these genes, binary classification rules were generated by forming all gene-pair combinations, where a rule was defined as “Gene A < Gene B” and evaluated by comparing the gene expression values within each sample. These rules were sorted based on their discriminative power using the ‘sort_rules_RF’ function. To ensure robustness, rules were ranked selection was performed using both the above-described “one-vs-rest” and “all-vs-altogether” ranking approaches. Top-ranked rules from each method were selected for downstream analysis, allowing for differential weights of all cell types to reduce bias due to sample imbalances.

The final classifier was then trained on binary input generated by evaluating the selected rules using the ‘train_RF’ function, with model parameters optimized through the ‘optimize_RF’ function. Combinations of the following parameters were tested to find the optimal parameter set: maximum number of times a gene can be repeated among the selected rule set, number of rules derived from the “one-vs-rest” method used in the model, number of rules per class derived from the “altogether” method used in the model, whether or not to remove uninformative rules from the selected rule set (boruta-based feature selection), and number of trees generated in the model. Based on the results from the parameter optimization, the final model was then constructed of 1,000 trees, with a gene repetition limit of one to ensure rule diversity. A total of 100 rules derived from the one-vs-rest scheme and 85 rules from the altogether scheme were selected for training. A total of 100 top-ranking rules derived from the “altogether” method and 100 top-ranking rules per class derived from the “one-vs-rest” method were initially selected for training. Boruta-based feature selection was enabled to exclude non-informative rules (Kursa et al., 2010). Additionally, probability estimation was activated, allowing the model to output class scores instead of categorical predictions. The resulting model was composed of 452 binary rules (gene-pairs) across all classes (Supplementary Figure S2).

Primary cell-line RF classifier

Concurrently, we generated a cell-line-based Random Forest classifier as above but made the training set using cell line expression profiles. Specifically, we used data from DepMap (Tsherniak et al., 2017; Cancer Cell Line Encyclopedia Consortium and Genomics of Drug Sensitivity in Cancer Consortium, 2015), including 174 cell lines collected from the DepMap Download portal at https://depmap.org/portal/data_page/?tab=currentRelease version 24Q4 under the expression tab. We downloaded and integrated data from “Model.csv” (the collection of metadata used to describe the cell types and their origins, including sex and tissue of origin) and “OmicsExpressionProteinCodingGenesTPMLogp1.csv” to get the expression data. Entrez gene IDs were removed, and genes were subset as above (see Data access and cleaning). Additionally, we subset our data to cell types that originated in the brain/central nervous system, or blood cancers, in order to classify cells as neural or myeloid in nature. 22 gene names were missing from the DepMap data and were not considered in the training: “RBM14_RBM4”, “FPGT_TNNI3K”, “BCL2L2_PABPN1”, “TEN1_CDK3”, “PPT2_EGFL8”, “RTEL1_TNFRSF6B”, “SENP3_EIF4A1”, “P2RX5_TAX1BP3”, “STX16_NPEPL1”, “DLEU1”, “ERV3_1”, “HLA_A”, “HLA_C”, “HLA_DMA”, “HLA_DQB1”, “HLA_DRB1”, “HLA_F”, “CHKB_CPT1B”, “ST20”, “ANKHD1_EIF4EBP3”, “ZNF286B”, and “JMJD7_PLA2G4B”. This left 8,702 genes for training the tissue of origin data.

Note “ACH-000075” or the U87 cell line is part of the DepMap dataset, but it was withheld from training and used only for testing. This also provides some justification for not performing cross-validation on our data, where U87 might eventually be included to optimize the model.

After following the same methodology above, 41 rules were used to build our classifier (Supplementary Figure S3).

Model testing and evaluation

Upon the creation of two classifiers, a primary cell classifier and cell-line classifier, we tested RNA-seq data generated from many different studies (Gupta et al., 2024; Schirmer et al., 2019; Masuda et al., 2019; Quiroga et al., 2022; Armanville et al., 2025; Baek and Yoo, 2021; Baek et al., 2022; Abud et al., 2017) to evaluate the performance of our model and assess its validity. The ‘CaretconfusionMatrix’ function was utilized to generate model performance metrics for both the training and test sets (Supplementary Figure S2, S3). Because our aim was to interrogate HMC3 lineage rather than develop a broadly generalizable tool, we did not perform k-fold cross-validation. Instead, model reliability was assessed through independent validation using datasets from Masuda et al. (2019) and Schirmer et al. (2019), which provided stringent external testing and demonstrated strong classifier performance. These single-nuclei and single-cells studies, respectively, were analyzed using pseudo-bulk strategies by aggregating expression data according to the cell-type annotation of the original publications.

Testing for rat sequencing reads

FASTQ files were downloaded to a Google bucket from the Sequence Read Archive (SRA) for all HMC3 samples tested (GSE181153, GSE275256, GSE155408, specifically, SRR12347826, SRR12347827, SRR12347828, SRR15301012, SRR15301013, SRR15301082, SRR15301083, SRR30311414, SRR30311415, SRR30311416, SRR30311417, SRR30311418, SRR30311419, SRR30311420, SRR30311421, and SRR30311422). To test whether samples had more RNA sequencing reads that aligned to Rat (*Rattus norvegicus*) or Human (*Homo sapiens*) reads, we used xengsort (Zentgraf and Rahmann, 2021) <https://gitlab.com/genomeinformatics/xengsort> to characterize reads according to their alignment preference. We used Ensembl GRCh38 release 114 cDNA and DNA FASTA files for the human reference and GRCr8 release 114 cDNA and DNA sequences for the rat genome reference. Briefly, the xengsort pipeline uses a memory-intensive step to index the reference genomes and then uses Cuckoo hashing for rapid assessment of k-mer fidelity in order to classify raw reads to the different references for each sample.

Testing for rRNA in shared reads

Due to the high sequence synteny of rRNA fragments (even across species), we evaluated the fraction of reads that aligned to both human and rat sequences. Note, rRNA, is the most abundant molecule in the cell, can comprise up to 80% of the RNA content in a cell (O'Neil et al., 2013). To do this, we first collected a compilation of rRNA reference data from SILVA: <https://www.arb-silva.de/>, a database dedicated to building “A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data.” There, we downloaded a multi-species rRNA reference file, “smr_v4.3_fast_db.fasta” to capture the Large Subunit (LSU) and Small Subunit (SSU) of rRNA (Glöckner et al., 2017). We used SortMeRNA (to rapidly quantify the fraction of rRNA reads found to be shared between human and rat—as calculated by xengsort (Kopylova et al., 2012). We applied this methodology to the Baek et al. data, which indicated a high fraction of reads shared between rat and human.

Code availability

The code used to clean and analyze these data is available at https://github.com/MHBailey/Cellf_deception.

Results

Data collection

Data were collected from fourteen different studies involving brain cells and cell lines (Srinivasan et al., 2020; Galatro et al., 2017; Zhang et al., 2016; Gosselin et al., 2017; Costa-Verdera et al., 2025; Tsherniak et al., 2017; Gupta et al., 2024; Schirmer et al., 2019; Masuda et al., 2019; Quiroga et al., 2022; Armanville et al., 2025; Baek and Yoo, 2021; Baek et al., 2022;

Abud et al., 2017) (Supplementary Table S1). In total, 432 samples from six of these studies were used to train two different classifiers—a classifier for cells collected from human tissues and another classifier from cell lines (Materials and Methods).

Building a primary tissue-type Random Forest classifier

A statistical R package called multiclassPairs was used to build these classifiers (Marzouka and Eriksson, 2021) (Materials and Methods). Briefly, multiclassPairs builds Random Forest classification models based on the relationship between two genes instead of single gene quantities. Gene-pair ratios are less subject to batch effects (Ellrott et al., 2025) and thus more comparable when including many studies. Specifically, we leveraged both bulk-sorted and single-cell RNA-seq (scRNA-seq) data from multiple studies to create our classifiers.

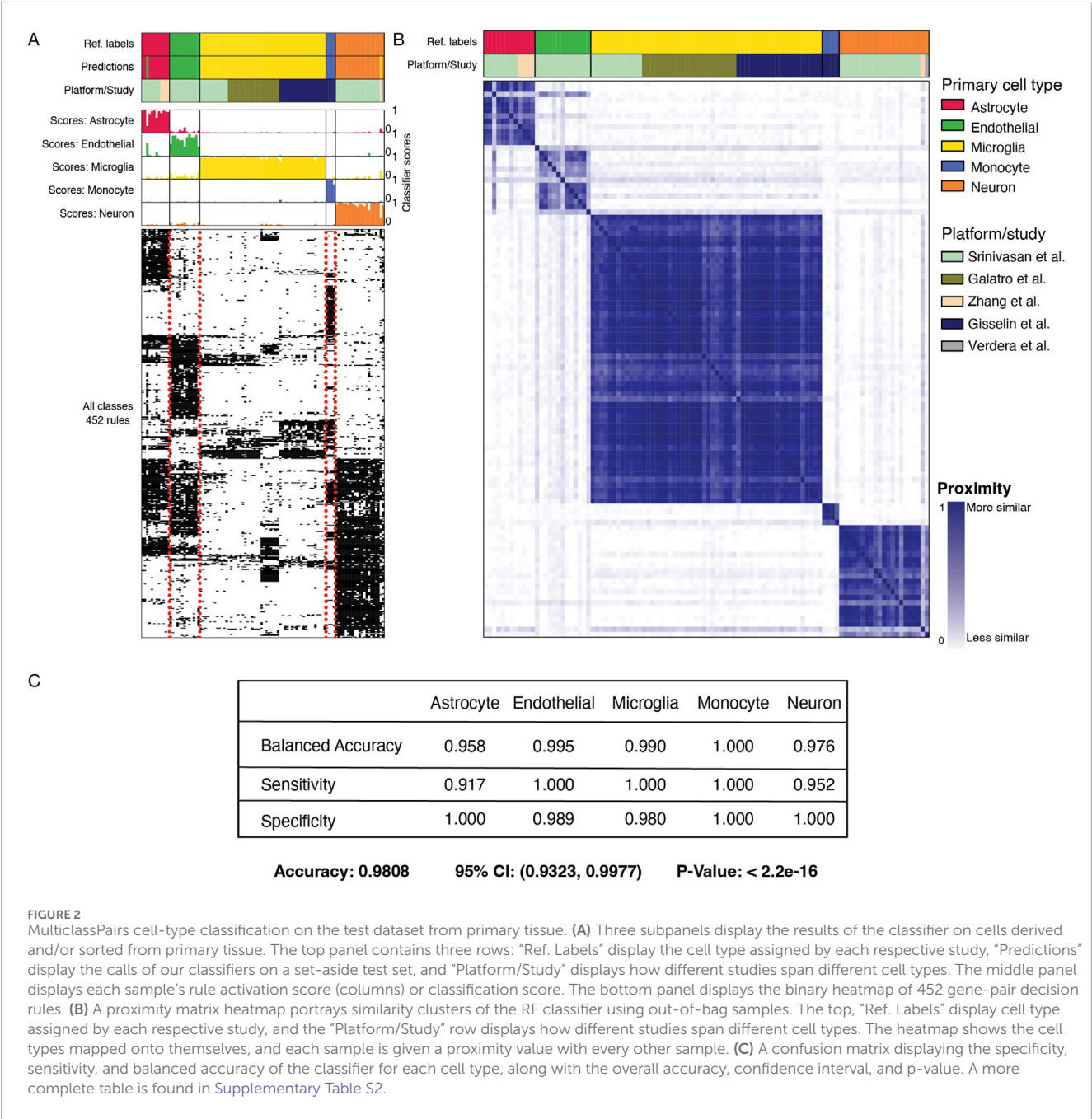
Primary tissue classifier training

258 samples were derived from tissue material (Supplementary Figure S2). Using RNA-seq information from these samples (bulk and single-cell transcriptomics), we built a Random Forest using multiclassPairs. We used a 60:40 training-to-test ratio. In the training data, only one labeled astrocyte sample was predicted as an endothelial cell, and one neuron was misclassified to be microglial in origin (Supplementary Table S2). The testing confusion matrix showed promising data with an overall accuracy of 98.08% (CI 0.9323–0.9977) and high balanced accuracies per cell type (>95.83% for each cell type), showing that the majority of samples were classified as their correct cell type.

The primary classifier used 452 gene-pair rules to classify labeled cell types (Figure 2A). Using these gene-pair rules, most cell types have distinguishable groups of gene rules that lead to their prediction; notably, gene-pairs-features are less sensitive to the data platform (Figure 2B). Overall balanced accuracy, sensitivity, and specificity provided strong evidence that our model could correctly classify the cell-type of origin for data derived from specific cells (Figure 2C; Supplementary Table S2). For our purposes, we wished to use this classifier to determine the fidelity of established microglial cell lines to the human-derived cellular material.

Validate primary tissue classifier with additional datasets

To further assess the validity and accuracy of the RF classifiers, we tested data from two unrelated studies as independent test sets using single-cell and single-nuclei sequencing (Figure 3). The first dataset from Masuda et al. contains gene expression data for seven samples of human microglia cells from both non-MS control brain tissue and diseased MS brain tissue (Masuda et al., 2019). The second dataset from Schirmer et al. contains gene expression data for nine samples of human astrocyte cells from non-MS control brain tissue (Schirmer et al., 2019). Despite higher imputation levels (291



missing genes for Schirmer et al. and 292 missing genes for Masuda et al., i.e., more than half of the rules missing, see Materials and Methods), all samples from both studies were correctly identified as their known cell type of origin (Figure 3). In the Masuda et al. study, the microglial cell prediction score for each sample was above 95% (Figure 3A). These consistent results for every sample demonstrate the high efficiency of the RF classifier to make cell-type-specific predictions even with missing data. Similarly, 8 of 9 astrocyte samples from the Schirmer et al. study were correctly classified by a cell prediction score of at least 0.87 (Schirmer et al., 2019). The highest astrocyte cell prediction score was 0.957 (Figure 3B). Interestingly, samples with the most extreme predictive score, C7 Astrocyte, C9 Astrocyte, C2 Astrocyte, and C8 Astrocyte, had the

fewest astrocyte-labeled cells from their single-cell experiments (average number of labels astrocytes were 31.5 cells compared to 289 cells in the rest of the samples: C4, C5, C1, C6, C3; averages generated from Schirmer et al.'s [Supplementary Table S4](#)). These lower cell counts may have led to their less consistent predictions. Overall, these two validation sets provide strong evidence for the strength of our classification tool.

Classification of the HMC3 cell line

We next employed our primary cell-based RF classifier on HMC3 cell line expression datasets to determine how our model

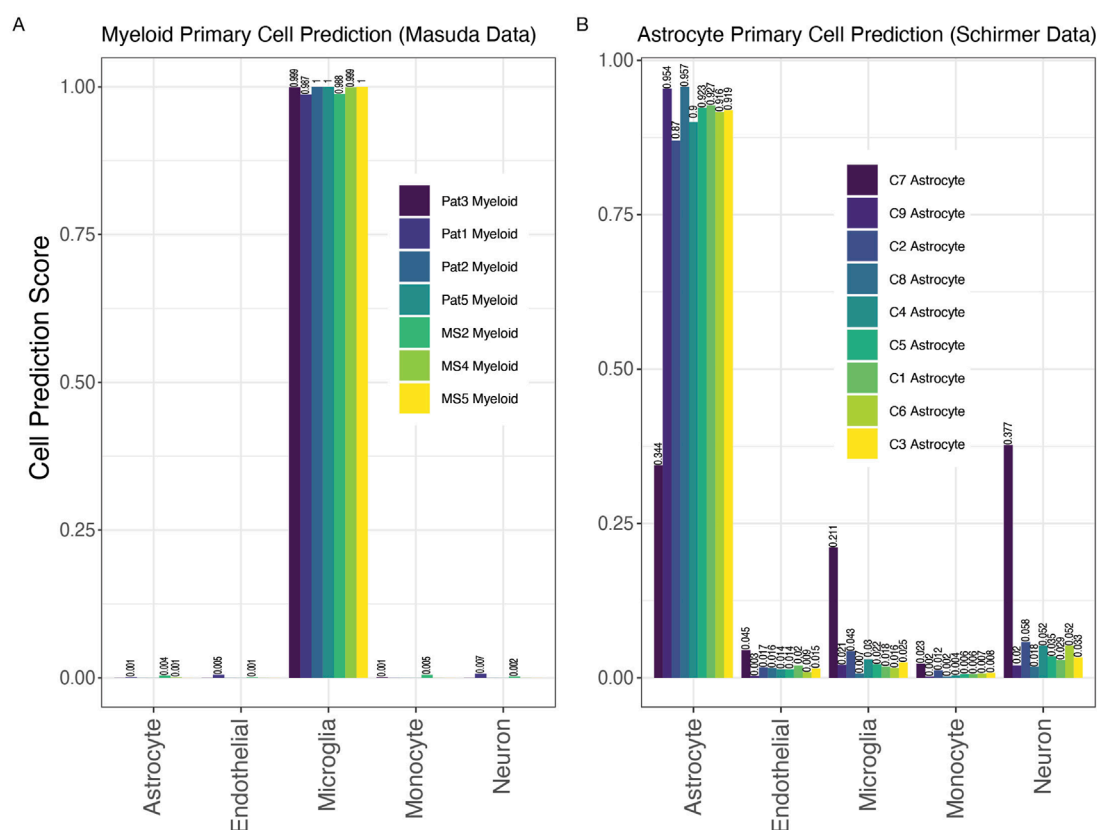


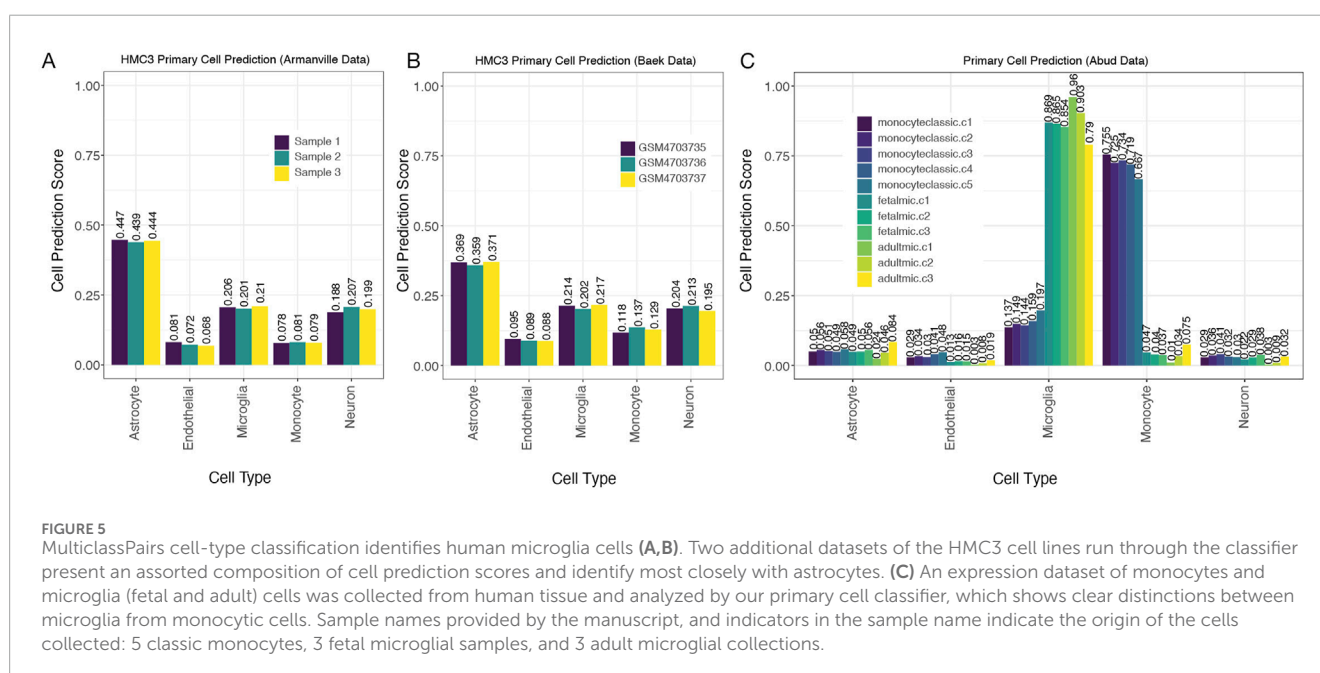
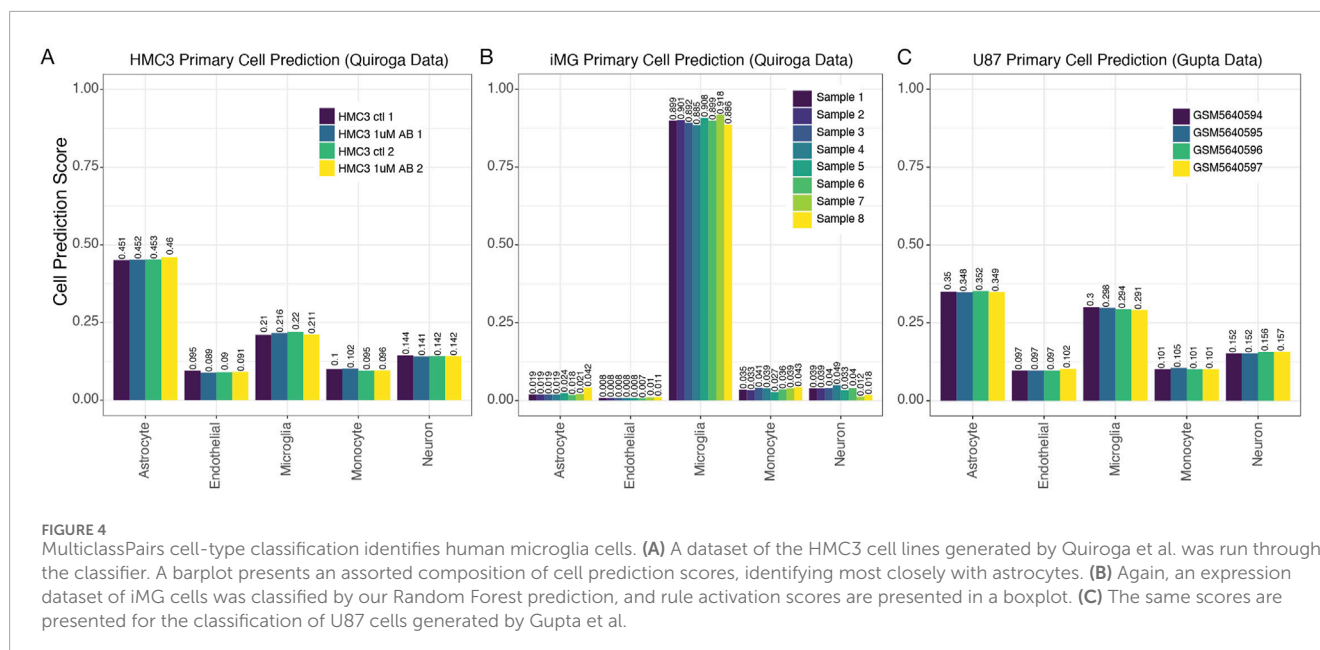
FIGURE 3

MulticlassPairs cell-type classification on independent validation sets of primary tissue. (A) Seven human microglia samples from Masuda et al. were run through the classifier. Cell prediction scores for microglia range from 0.9875 to 1.000. (B) Nine human astrocyte samples from Schirmer et al. were run through the classifier. Cell prediction scores for astrocytes range from 0.344 to 0.957. Values closer to 1 indicate stronger cell assignment congruity.

classifies this line. We began with four HMC3 samples from GSE181153 by Quiroga et al. (2022). Our model's predictions for HMC3 cells based on gene-pair expression patterns were less conclusive than for our test cases of primary cell isolates, but the highest cell prediction scores suggested that HMC3 cells were astrocyte-derived, with microglial derivation receiving the second highest scores (Figure 4A). The cell prediction scores for HMC3 cells representing microglia-derived cells ranged from 0.21 to 0.22 while the range of scores for HMC3 cells representing astrocyte-derived cells was 0.45–0.46 (Figure 4A). Thus, we can see that the HMC3 cell line exhibits a heterogeneous profile, and it may be most appropriately classified as an astrocyte-derived line according to its gene expression profile (Voloshin et al., 2023). When we utilized the primary cell-based RF classifier to analyze iMG samples from the same study (Quiroga et al., 2022) (withheld from the original training data), iMG cells were confidently classified as microglia, with predictions ranging from 0.884 to 0.918 (Figure 4B). It is evident that iMG cells, though tedious to produce, provide a much more accurate experimental model of microglia than the HMC3 cell line. Notably, the HMC3 cell line, while previously considered a microglia model, does not effectively capture microglial characteristics and instead presents a transcriptionally diverse profile more closely aligning with astrocytes.

To further explore the original finding from Quiroga et al. that HMC3 cell profiles clustered more closely with U87 glioblastoma-derived cells than with myeloid cells, we also tested U87 cell line profiles in our primary cell-based RF classifier (Figure 4C). The U87 samples were also given an ambiguous classification by our tool, with astrocyte scores at ~0.35 and microglia scores at 0.29–0.3 (Figure 4C). Interpreting these results, we hypothesized scenarios that might explain the data. First, the immortalization process could put U87 and HMC3 cell lines onto a similar transcriptomic trajectory. This could be supported by our PCA analysis when including iPSCs, which showed iPSCs clustering with U87 and HMC3 cells (Supplementary Figure S1). An alternative, albeit speculative, explanation is inadvertent mixing of U87 and HMC3 cells during sample handling.

To address the latter possibility, we extended our analysis to two additional datasets that sequenced HMC3 cell lines and deposited their data (Armanville et al., 2025; Baek and Yoo, 2021; Baek et al., 2022). Similar to the HMC3 samples from Quiroga et al., HMC3 samples from these two datasets suggested that HMC3 cells reflect astrocyte expression more than microglial expression, with predictive scores for astrocyte derivation again being ~2-fold higher than for microglial derivation (Figures 5A,B). In fact, predictive scores for neuron derivation were as high as the scores for microglial derivation in these HMC3 samples.



To further explore the specificity of our model in accurately classifying monocytes and microglia cells from varying age groups (fetal and adult), we incorporated an additional test using the Abud et al. dataset (2017). Their study included gene expression profiles of monocytes as well as induced microglial (iMG) lines derived from induced pluripotent stem cells. Abud et al. demonstrated that these iMGs closely resembled primary human microglia in their transcriptional profiles (Abud et al., 2017). Due to the thoroughness of their sequencing across many facets, we found that both fetal and adult microglial populations were properly identified by our classifier (Figure 5C). These findings further support the robustness of our gene-pair classifier across data platforms and its relative immunity to batch effects while also

confirming its ability to accurately identify microglia regardless of developmental stage.

Classification of a cell-type classifier using DepMap cell lines

To address the possibility that HMC3 cells and U87 cells clustered together mainly because they are transformed/immortalized cell lines and not because they have similar cell origins, we built another RF classifier using only transformed cell lines derived from human cancers of myeloid cell lineages or of the brain (neural, astroglial, or oligodendroglial

cancers). To build our transformed cell line-based RF classifier, we used cell lines studied in the DepMap cell line project (Tsherniak et al., 2017; Cancer Cell Line Encyclopedia Consortium and Genomics of Drug Sensitivity in Cancer Consortium, 2015). Similar to above, we used multi-class pairs to build a gene-pair classifier, allowing us to compare the DepMap sample to the datasets we collected above (Materials and Methods). Of the 174 different cell lines from CNS/Brain or Myeloid in the DepMap dataset, we used 56 CNS/Brain labeled cell lines and 48 Myeloid cell lines to train the gene-pair Random Forest classifier (Supplementary Figure S3A, S3B). The test set revealed an F1 statistic of 0.978, an accuracy of 97.1%, a sensitivity score of 1.00, and a specificity of 0.923 when using the remaining cells as a test set (Supplementary Figure S3). Of note, the 2 cell lines that were misclassified in the test set were the HAP1 line (ACH-002475) and HDMYZ (ACH-000190, Supplementary Figure S3C, S3D). Both lines have lineages labeled by DepMap to be myeloid in nature, but were better classified with the Brain/CNS labels in our classifier. Interestingly, HAP1 is a cell line derived from the chronic myeloid leukemia line, KBM-7 (Kotecki et al., 1999), and is a near-haploid cell line, making it easier to introduce mutations into a model system. Interestingly, the HAP1 arose from the failed attempt to induce pluripotency in KBM-7 by the overexpression of OCT4, SOX2, MYC, and KLF4. However, the resulting cell lost its hematopoietic markers (Carette et al., 2011), which agrees with our misclassification. The HD-MY-Z cell line, originally derived from a pleural effusion taken from a 29-year-old Hodgkin's patient (Bargou et al., 1993), has also undergone scrutiny, and is now considered to be more AML in origin (Drexler et al., 2018). Drexler et al.'s commentary suggests these cells are more typical of myelomonocytic cells. Our classifier suggests a different origin that warrants further investigation. These results suggest that our classifier can accurately distinguish between cells and can capture known discrepancies in the field.

When we tested HMC3 lines datasets (none of which are in the DepMap data), we observed close alignment with CNS/Brain cancer-derived cell line profiles, a result that is counter to the non-CNS origin of microglial cells (Figures 6A,D,E). Similarly, we also observed that U87 cells were predicted to originate from a CNS/Brain lineage, as expected (Figure 6B). U87 lines are part of the DepMap project but were withheld from our original training data to be used as a test case. The iMG cell profiles reported by Quiroga et al. or by Abud et al. aligned much more strongly with myeloid lineage cancers than CNS/Brain cancer cell lines, a result in line with expectation (Figures 6C,F). Finally, from tissue collected by Abud et al., we show that monocytes, fetal microglia, and adult microglia all classify with myeloid cell line profiles, according to their natural origins (Figure 6F). Overall, the transformed cell line-based classifier provides strong additional evidence that HMC3 cells lack the microglial-like identity or transcriptomic signatures sufficient to warrant their classification as a useful cell line for microglial research.

Classification of reads in HMC3 cell lines

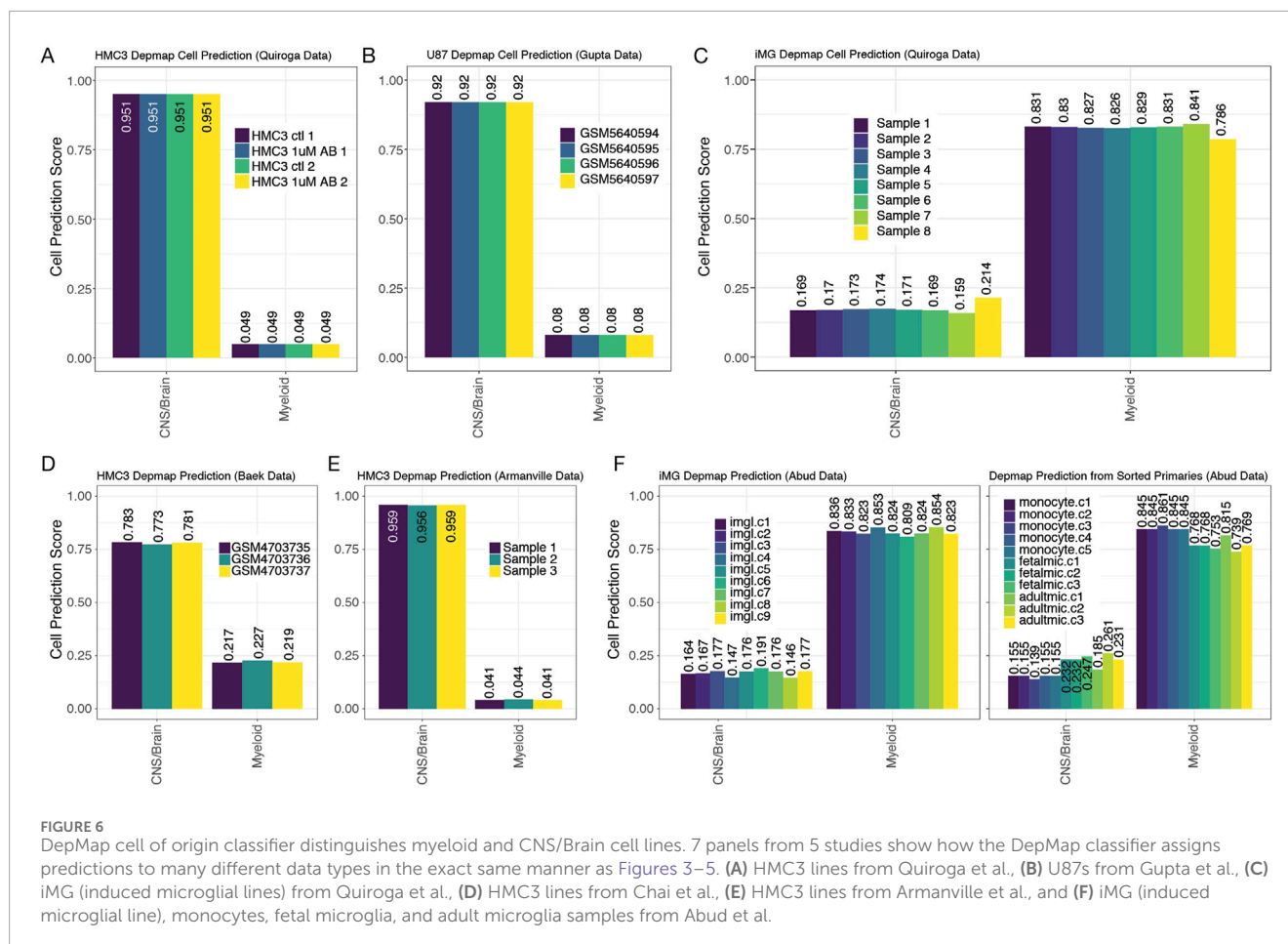
Finally, one important aspect that needs to be addressed is the possibility that all three HMC3 datasets analyzed in this manuscript were actually derived from rat glioma-derived cells, a widespread

problem previously described for some neural cell lines (Garcia-Mesa et al., 2017). A number of sequencing reads from rat cells will align with the human genome. While ATCC consistently provides high-quality cell lines for labs, we performed a quick assessment of RNA read quality and checked alignments for quality control purposes. Because of the previous challenges with deriving cell lines from *R. norvegicus* (rat), we assessed read origins based on the k-mer fidelity. To accomplish this task, we used a well-designed tool called xengsort (Zentgraf and Rahmann, 2021). Briefly, xengsort uses a large alignment-free k-mer store, i.e., a large key-value hash table, to assess whether the DNA or RNA reads best align to one, both, or neither of two references provided. Because of historical problems, it must be addressed explicitly by aligning the HMC3 sequencing reads with human vs. rat genomes and analyzing mismatches to ensure that the HMC3 datasets are in fact, human-derived cells and not rat-derived cells. Results from this analysis clearly show that data from the Armanville et al. and Quiroga et al. studies are primarily from the *Homo sapiens* genome (mean of 98.53% and 98.07%, respectively, Figure 7A). Surprisingly, the Baek et al. study did show many shared k-mers with the Rat reference (mean 30.69% shared between human and rat); but only 0.13% of the k-mers were rat-specific. Figure 6D.

To further investigate, we traced the cell line's reported origin in the Baek et al. study. To better understand the origin of the reads that mapped to both the human and rat genomes, we consulted the developers of xengsort. They helped identify that most of these "both" reads originated from ribosomal RNA (rRNA), which is highly conserved across species. When we applied this theory using SortMeRNA (Kopylova et al., 2012), a read-mapping tool (Materials and Methods), we found strong evidence supporting elevated shared content between rat and human reads in the Baek et al. dataset is likely the result of insufficient rRNA depletion in the library preparation prior to sequencing, rather than contamination or misidentification of the cell line (Figure 7B).

Discussion

The HMC3 cell line was created from an *in vitro* microglial sample in 1995 by transfection of the SV40-T antigen in primary human microglial cultures taken from human embryos (Janabi et al., 1995). At the time of its creation, however, the line was never directly compared to mature human microglial cells to confirm its identity. Complicating its provenance further, HMC3 is reported to be a direct derivative of the CHME-5 cell line, which was later found to consist of rat glioma cells in many laboratories' samples (Garcia-Mesa et al., 2017). HMC3 cells were also checked by qualitative immunostaining for the expression of certain markers. Although the HMC3 cells lacked glial fibrillary acidic protein (GFAP) staining and neurofilament staining (used to rule out astrocytic and neuronal contamination), this is not definitive, as other commonly used astrocyte-like lines, such as U87 cells, also lack GFAP expression (Russo et al., 2018). HMC3 also revealed markedly higher basal secretion levels than other microglial cell lines. Additionally, certain chemical transformations done to acquire the immortalized HMC3 cell made it nearly impossible to compare directly with microglial cells (Russo et al., 2018). As HMC3 becomes



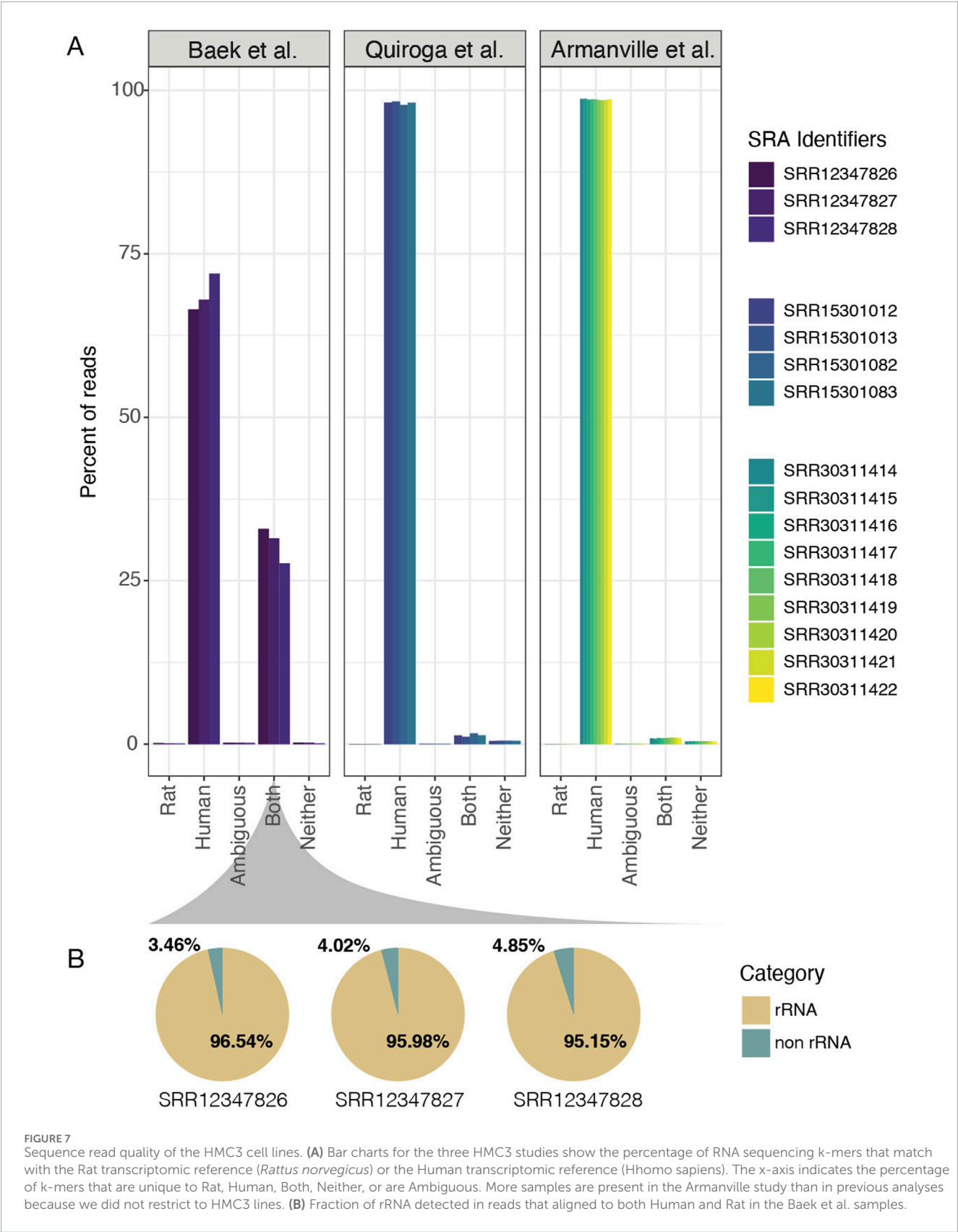
more widely used, its unexpected behavior has prompted researchers to speculate about its true nature as a microglial substitute. This study's goal was to provide an informed statistical opinion on the true lineage of the HMC3 cell line.

The results of our study indicate that HMC3 cells are more closely aligned with astrocytic expression profiles than with microglia. Our results demonstrate that the classification method is robust and accurate in our test (Figure 2) and validation sets (Figure 3). Furthermore, we highlight the strength of rules-based features as a viable method to combine gene expression analyses from multiple analyses, i.e., gene-pair qualifiers instead of absolute transcripts per million. While it was advantageous to apply this feature-based tool to multiple datasets with minimum effort in batch correction, the results of this data integration highlight the fact that HMC3 cells classify better with astrocytes than myeloid lineages. One possible explanation is that in the process of creating HMC3, due to the nature of how cells are changed to become immortalized, the cells acquired a stem-like state void of their developmental lineage (Pauklin and Vallier, 2013). However, our transformed cell line-based RF classifier refuted this possibility. Additionally, we also consider rat glioma contamination an unlikely explanation for the HMC3 profile given our read-classification results (Figure 7B). This showed that HMC3 datasets overwhelmingly aligned to the human genome, with any ambiguous reads largely attributable to highly conserved rRNA rather than true rat contamination. Therefore, a

more likely explanation is that the cell culture from which the HMC3 cell line was created was not pure microglia, and that a contaminating astrocyte gave rise to the SV40-transformed clonal cell line that became known as HMC3. Regardless of the true explanation, our results show that using HMC3 cells as a method to model microglial activity for AD research is not justified.

These findings also have implications for AD research. Since HMC3 cells resemble astrocytes more closely than microglia, studies using HMC3 to model microglial biology may inadvertently lead to misinterpretation of microglial roles in AD pathogenesis—particularly in studies investigating immune signaling, phagocytosis of β -amyloid, and neuroinflammation. This misclassification could obscure the true contributions of microglia to AD pathogenesis and yield misleading conclusions about immune responses around β -amyloid plaques or therapeutic targets. In contrast, iPSC-derived microglia, though resource-intensive, more accurately model microglial biology and are preferable for mechanistic and translational AD studies.

Our findings align with recent work by Woolf et al. (2025), who conducted a direct phenotypic and functional comparison of commonly used *in vitro* microglia models, including HMC3 (Woolf et al., 2025). In their study, HMC3 cells failed to express canonical microglia markers (Iba1, CD45, PU.1) and instead stained positive for mural cell markers such as PDGFR β and NG2. Functionally, HMC3 displayed significantly lower phagocytic activity and



secretory responsiveness compared to primary and iPSC-derived microglia—more closely resembling pericytes than cells of myeloid lineage. These findings, derived from an independently sourced and ATCC-validated vial of HMC3 cells, provide an important biological complement to our transcriptomic findings. In addition to validating the classifier on independent primary-cell datasets

(Masuda et al., 2019; Schirmer et al., 2019), we applied the model to independent HMC3 datasets (Armanville et al., 2025; Baek and Yoo, 2021). While these latter datasets are not used to validate cross-cell-type performance, they provide out-of-sample replication of our central finding that HMC3 aligns more closely with astrocytic than microglial signatures. These independent testing sets consistently supported our findings. The convergence of computational and experimental evidence strengthens the robustness and generalizability of our conclusion that HMC3 does not represent a valid microglial model.

It is critical to accurately determine the origin of HMC3 cell lines, as they are widely used in key research on Alzheimer's disease (Russo et al., 2018). Microglial recruitment in the brain and central nervous system plays a vital role in supporting brain health and resilience in the face of neurodegenerative diseases, due to microglia's regulation of several essential immune functions (Miao et al., 2023). This motivated our investigation into the HMC3 line, which we found cannot be accurately classified as microglia. As a result, findings from studies that substitute HMC3 for microglia may be compromised or misleading.

Future studies with larger datasets could provide greater statistical power to detect cell-type-specific expression differences, strengthening confidence in our conclusions. Further analysis should identify the genes that most effectively distinguish astrocytes and microglia from other cell types, and evaluate whether HMC3 has any valid, limited applications. Additionally, to the power of data integration between bulk-RNA and single-cell RNA experiments, we believe this strategy for data integration may help boost power as new tools seek to use AI to identify novel drug targets (Zhao et al., 2022). This will be increasingly important as higher-resolution data emerges across the spectrum of human diseases and the cells that drive their expression.

Our computational findings strongly indicate that HMC3 is not a true human microglial model of myeloid origin and call into question its use in prior research, urging caution in continuing to use HMC3 as a proxy for microglia.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used. Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used. When human data were used for this study, specifically results from scRNA, snRNA, or sorted bulk sequencing, only combined

expression counts were analyzed from previous publications; thus, no identifiable data were collected or used for the results of this study.

Author contributions

KR: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing. BK: Conceptualization, Data curation, Formal Analysis, Software, Writing – original draft. MM: Formal Analysis, Writing – original draft. BS: Formal analysis, Writing – original draft. KS: Data curation, Investigation, Writing – original draft. BK: Data curation, Writing – review and editing. YC: Formal analysis, Validation, Visualization, Writing – review and editing. SS: Formal analysis, Validation, Visualization, Writing – review and editing. AN: Formal analysis, Methodology, Writing – review and editing, Investigation. DH: Conceptualization, Data curation, Methodology, Supervision, Writing – original draft, Writing – review and editing. MB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

The authors would like to thank the countless and often nameless individuals who build, maintain, and provide their sequencing data to the research community. We would also like to thank Samuel Payne, Ph.D., for the tremendous time and effort he spends working with our undergraduate authors and other bioinformatics majors on their capstone projects at Brigham Young University. Brigham Young University's Office of Research Computing, which provided supercomputing resources for our analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Clarify writing, summarize the scope statement, and provide preliminary drafts of the abstract and discussions. It also helped generate the title of our manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to

ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be

evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2025.1681811/full#supplementary-material>

References

- Abud, E. M., Ramirez, R. N., Martinez, E. S., Healy, L. M., Nguyen, C. H. H., Newman, S. A., et al. (2017). iPSC-derived human microglia-like cells to study neurological diseases. *Neuron* 94 (2), 278–293.e9. doi:10.1016/j.neuron.2017.03.042
- Armanville, S., Tocco, C., Haj Mohamad, Z., Clarke, D., Robitaille, R., and Drouin-Ouellet, J. (2025). Chemically induced senescence prompts functional changes in human microglia-like cells. *J. Immunol. Res.* 2025 (1), 3214633. doi:10.1155/jimr/3214633
- Author anonymous (2024). 2024 alzheimer's disease facts and figures. *Alzheimer's & Dementia J. Alzheimer's Assoc.* 20 (5), 3708–3821. doi:10.1002/alz.13809
- Baek, M., Yoo, E., Choi, H. I., An, G. Y., Chai, J. C., Lee, Y. S., et al. (2021). The BET inhibitor attenuates the inflammatory response and cell migration in human microglial HMC3 cell line. *Sci. Rep.* 11 (1), 8828. doi:10.1038/s41598-021-87828-1
- Baek, M., Chai, J. C., In Choi, H., Yoo, E., Binas, B., Lee, Y. S., et al. (2022). Analysis of differentially expressed long non-coding RNAs in LPS-induced human HMC3 microglial cells. *BMC Genomics* 23 (1), 853. doi:10.1186/s12864-022-09083-6
- Bargou, R. C., Mapara, M. Y., Zugck, C., Daniel, P. T., Pawlita, M., Döhner, H., et al. (1993). Characterization of a novel hodgkin cell line, HD-MyZ, with myelomonocytic features mimicking hodgkin's disease in severe combined immunodeficient mice. *J. Exp. Med.* 177 (5), 1257–1268. doi:10.1084/jem.177.5.1257
- Bhattacharjee, A., Rodrigues, E., Jung, J., Luzentales-Simpson, M., Enterina, J. R., Galleguillos, D., et al. (2019). Repression of phagocytosis by human CD33 is not conserved with mouse CD33. *Commun. Biol.* 2 (1), 450. doi:10.1038/s42003-019-0698-6
- Blasi, E., Barluzzi, R., Bocchini, V., Mazzolla, R., and Bistoni, F. (1990). Immortalization of murine microglial cells by a v-Raf/v-Myc carrying retrovirus. *J. Neuroimmunol.* 27 (2–3), 229–237. doi:10.1016/0165-5728(90)90073-v
- Bohlen, C. J., Friedman, B. A., Dejanovic, B., and Morgan, S. (2019). Microglia in brain development, homeostasis, and neurodegeneration. *Annu. Rev. Genet.* 53 (1), 263–288. doi:10.1146/annurev-genet-112618-043515
- Cancer Cell Line Encyclopedia Consortium, and Genomics of Drug Sensitivity in Cancer Consortium, (2015). Pharmacogenomic agreement between two cancer cell line data sets. *Nature* 528 (7580), 84–87. doi:10.1038/nature15736
- Carette, J. E., Raaben, M., Wong, A. C., Herbert, A. S., Obernosterer, G., Mulherkar, N., et al. (2011). Ebola virus entry requires the cholesterol transporter niemann-Pick C1. *Nature* 477 (7364), 340–343. doi:10.1038/nature10348
- Cedres, N., Ekman, U., Poulakis, K., Shams, S., Cavallin, L., Muehlboeck, S., et al. (2020). Brain atrophy subtypes and the ATN classification scheme in alzheimer's disease. *Neuro-Degenerative Dis.* 20 (4), 153–164. doi:10.1159/000515322
- Condello, C., Peng, Y., Schain, A., and Grutzendler, J. (2015). Microglia constitute a barrier that prevents neurotoxic protofibrillar Aβ42 hotspots around plaques. *Nat. Commun.* 6 (1), 6176. doi:10.1038/ncomms7176
- Costa-Verdera, H., Meneghini, V., Fitzpatrick, Z., Abou Alezz, M., Fabyanic, E., Huang, X., et al. (2025). AAV vectors trigger DNA damage response-dependent pro-inflammatory signalling in human iPSC-Derived CNS models and mouse brain. *Nat. Commun.* 16 (1), 3694. doi:10.1038/s41467-025-58778-3
- Drexler, H. G., Pommerenke, C., Eberth, S., and Nagel, S. (2018). Hodgkin lymphoma cell lines: to separate the wheat from the chaff. *Biol. Chem.* 399 (6), 511–523. doi:10.1515/hsz-2017-0321
- Elliott, K., Wong, C. K., Yau, C., Castro, M. A. A., Lee, J. A., Karlberg, B. J., et al. (2025). Classification of Non-TCGA cancer samples to TCGA molecular subtypes using compact feature sets. *Cancer Cell* 43 (2), 195–212.e11. doi:10.1016/j.ccell.2024.12.002
- Galatro, T. F., Holtman, I. R., Lerario, A. M., Vainchtein, I. D., Brouwer, N., Sola, P. R., et al. (2017). Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat. Neurosci.* 20 (8), 1162–1171. doi:10.1038/nn.4597
- Garcia-Mesa, Y., Jay, T. R., Ann Checkley, M., Luttge, B., Dobrowolski, C., Valadkhan, S., et al. (2017). Immortalization of primary microglia: a new platform to study HIV regulation in the central nervous system. *J. Neurovirology* 23 (1), 47–66. doi:10.1007/s13365-016-0499-3
- Gerrits, E., Yang, H., Boddeke, E. W. G., and Eggen, B. J. L. (2020). Transcriptional profiling of microglia; current state of the art and future perspectives. *Glia* 68 (4), 740–755. doi:10.1002/glia.23767
- Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., et al. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.* 261, 169–176. doi:10.1016/j.jbiotec.2017.06.1198
- Gosselin, D., Skola, D., Coufal, N. G., Holtman, I. R., Schlachetzki, J. C. M., Sajti, E., et al. (2017). An environment-dependent transcriptional network specifies human microglia identity. *Sci. (New York, N.Y.)* 356 (6344), eaal3222. doi:10.1126/science.aal3222
- Gupta, R., Dittmeier, M., Wohlleben, G., Vera, N., Bischler, T., Luzak, V., et al. (2024). Atypical cellular responses mediated by intracellular constitutive active TrkB (NTRK2) kinase domains and a solely intracellular NTRK2-Fusion oncogene. *Cancer Gene Ther.* 31 (9), 1357–1379. doi:10.1038/s41417-024-00809-0
- Hansen, D. V., Hanson, J. E., and Sheng, M. (2018). Microglia in alzheimer's disease. *J. Cell Biol.* 217 (2), 459–472. doi:10.1083/jcb.201709069
- Haque, A., Engel, J., Teichmann, S. A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 9 (1), 75–12. doi:10.1186/s13073-017-0467-4
- James, B. D., Leurgans, S. E., Hebert, L. E., Scherr, P. A., Yaffe, K., and Bennett, D. A. (2014). Contribution of alzheimer disease to mortality in the United States. *Neurology* 82 (12), 1045–1050. doi:10.1212/wnl.0000000000000240
- Janabi, N., Peudener, S., Héron, B., Ng, K. H., and Tardieu, M. (1995). Establishment of human microglial cell lines after transfection of primary cultures of embryonic microglial cells with the SV40 large T antigen. *Neurosci. Lett.* 195 (2), 105–108. doi:10.1016/0304-3940(94)11792-h
- Kaur, G., and Dufour, J. M. (2012). Cell lines: valuable tools or useless artifacts. *Spermatogenesis* 2 (1), 1–5. doi:10.4161/spmg.19885
- Keil, J. M., Qalieh, A., and Kwan, K. Y. (2018). Brain transcriptome databases: a user's guide. *J. Neurosci. Official J. Soc. Neurosci.* 38 (10), 2399–2412. doi:10.1523/jneurosci.1930-17.2018
- Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28 (24), 3211–3217. doi:10.1093/bioinformatics/bts611
- Kotecki, M., Reddy, P. S., and Cochran, B. H. (1999). Isolation and characterization of a near-haploid human cell line. *Exp. Cell Res.* 252 (2), 273–280. doi:10.1006/excr.1999.4656
- Kursa, M. B., Jankowski, A., and Rudnicki, W. R. (2010). Boruta – a system for feature selection. *Fundam. Inf.* 101 (4), 271–285. doi:10.3233/fi-2010-288
- Love, M. I., Huber, W., and Simon, A. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Marzouka, N.-A.-D., and Eriksson, P. (2021). multiclassPairs: an R package to train multiclass pair-based classifier. *Bioinformatics* 37 (18), 3043–3044. doi:10.1093/bioinformatics/btab088
- Masuda, T., Sankowski, R., Staszewski, O., Böttcher, C., Amann, L., Sagar, C. S., et al. (2019). Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* 566 (7744), 388–392. doi:10.1038/s41586-019-0924-x
- McGeer, P. L., Kawamata, T., Walker, D. G., Akiyama, H., Tooyama, I., and McGeer, E. G. (1993). Microglia in degenerative neurological disease. *Glia* 7 (1), 84–92. doi:10.1002/glia.440070114

- Miao, J., Ma, H., Yang, Y., Liao, Y., Lin, C., Zheng, J., et al. (2023). Microglia in alzheimer's disease: pathogenesis, mechanisms, and therapeutic potentials. *Front. Aging Neurosci.* 15 (June), 1201982. doi:10.3389/fnagi.2023.1201982
- Mukamel, E. A., and Ngai, J. (2019). Perspectives on defining cell types in the brain. *Curr. Opin. Neurobiol.* 56 (June), 61–68. doi:10.1016/j.conb.2018.11.007
- O'Neil, D., Glowatz, H., and Schlumpberger, M. (2013). Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr. Protoc. Mol. Biol.* 103 (1), 4–19. doi:10.1002/0471142727.mb0419s103
- Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C., and Yuan, J. (2017). Single-cell RNA sequencing: unraveling the brain one cell at a time. *Trends Mol. Med.* 23 (6), 563–576. doi:10.1016/j.molmed.2017.04.006
- Pauklin, S., and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. *Cell* 155 (1), 135–147. doi:10.1016/j.cell.2013.08.031
- Quiroga, I. Y., Cruikshank, A. E., Bond, M. L., Reed, K. S. M., Evangelista, B. A., Tseng, J. H., et al. (2022). Synthetic amyloid beta does not induce a robust transcriptional response in innate immune cell culture systems. *J. Neuroinflammation* 19 (1), 99. doi:10.1186/s12974-022-02459-1
- Rai, M. A., Hammonds, J., Pujato, M., Mayhew, C., Krishna, R., and Paul, S. (2020). Comparative analysis of human microglial models for studies of HIV replication and pathogenesis. *Retrovirology* 17 (1), 35. doi:10.1186/s12977-020-00544-y
- Rawat, P., and Spector, S. A. (2017). Development and characterization of a human microglia cell model of HIV-1 infection. *J. Neurovirology* 23 (1), 33–46. doi:10.1007/s13365-016-0472-1
- Righi, M., Mori, L., De Libero, G., Sironi, M., Biondi, A., Mantovani, A., et al. (1989). Monokine production by microglial cell clones. *Eur. J. Immunol.* 19 (8), 1443–1448. doi:10.1002/eji.1830190815
- Russo, D., Cappoli, N., Coletta, I., Mezzogori, D., Paciello, F., Pozzoli, G., et al. (2018). The human microglial HMC3 cell line: where do we stand? A systematic literature review. *J. Neuroinflammation* 15 (1), 259. doi:10.1186/s12974-018-1288-0
- Schirmer, L., Velmeshev, D., Holmqvist, S., Kaufmann, M., Werneburg, S., Jung, D., et al. (2019). Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature* 573 (7772), 75–82. doi:10.1038/s41586-019-1404-z
- Spurgat, M. S., and Tang, S. J. (2022). Single-cell RNA-Sequencing: Astrocyte and microglial heterogeneity in health and disease. *Cells* 11 (13), 2021. doi:10.3390/cells11132021
- Srinivasan, K., Friedman, B. A., Etxeberria, A., Huntley, M. A., van der Brug, M. P., Foreman, O., et al. (2020). Alzheimer's patient microglia exhibit enhanced aging and unique transcriptional activation. *Cell Rep.* 31 (13), 107843. doi:10.1016/j.celrep.2020.107843
- Timmerman, R., Burm, S. M., and Bajramovic, J. J. (2018). An overview of *in vitro* methods to study microglia. *Front. Cell. Neurosci.* 12 (August), 242. doi:10.3389/fncel.2018.00242
- Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., et al. (2017). Defining a cancer dependency map. *Cell* 170 (3), 564–576.e16. doi:10.1016/j.cell.2017.06.010
- Vainchtein, I. D., and Molofsky, A. V. (2020). Astrocytes and microglia: in sickness and in health. *Trends Neurosci.* 43 (3), 144–154. doi:10.1016/j.tins.2020.01.003
- Voloshin, N., Tyurin-Kuzmin, P., Karagayur, M., Akopyan, Z., and Kulebyakin, K. (2023). Practical use of immortalized cells in medicine: current advances and future perspectives. *Int. J. Mol. Sci.* 24 (16), 12716. doi:10.3390/ijms241612716
- Wang, S., Li, H., Song, M., Tao, Z., Wu, T., He, Z., et al. (2021). Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. *PLoS Genet.* 17 (5), e1009557. doi:10.1371/journal.pgen.1009557
- Weiner, H. L. (2025). Immune mechanisms and shared immune targets in neurodegenerative diseases. *Nat. Rev. Neurol.* 21 (2), 67–85. doi:10.1038/s41582-024-01046-7
- Weuve, J., Hebert, L. E., Scherr, P. A., and Evans, D. A. (2014). Deaths in the United States among persons with Alzheimer's disease (2010–2050). *Alzheimer's & Dementia J. Alzheimer's Assoc.* 10 (2), e40–e46. doi:10.1016/j.jalz.2014.01.004
- Wickstead, E. S. (2023). Using stems to bear fruit: deciphering the role of Alzheimer's disease risk loci in human-induced pluripotent stem cell-derived microglia. *Biomedicine* 11 (8), 2240. doi:10.3390/biomedicine11082240
- Woolf, Z., Stevenson, T. J., Lee, K., Blake, H., Foliaki, J. M., Ratiu, R., et al. (2025). *In vitro* models of microglia: a comparative study. *Sci. Rep.* 15 (1), 15621. doi:10.1038/s41598-025-99867-z
- Zentgraf, J., and Rahmann, S. (2021). Fast lightweight accurate xenograft sorting. *Algorithms Mol. Biol. AMB* 16 (1), 2. doi:10.1186/s13015-021-00181-w
- Zhang, Ye, Sloan, S. A., Clarke, L. E., Caneda, C., Plaza, C. A., Blumenthal, P. D., et al. (2016). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* 89 (1), 37–53. doi:10.1016/j.neuron.2015.11.013
- Zhao, B.-W., Su, X. R., Hu, P. W., Ma, Y. P., Zhou, X., and Hu, L. (2022). A geometric deep learning framework for drug repositioning over heterogeneous information networks. *Briefings Bioinforma.* 23 (6), bbac384. doi:10.1093/bib/bbac384