International Journal of
Molecular Sciences

MDPI

Article

Technical Robustness of Transcriptomic Senescence Classification in Human Microglia: An Evaluation of SenFlag and ICE

Burcu Yuksel [1,*] and Levent Aydin [1,2]

[1] Kocaeli Vocational School of Health Services, Kocaeli University, Aydin 41100, Türkiye; levent.aydin@kocaeli.edu.tr
[2] Eureka Patent, Kocaeli Technopark, Kocaeli 41275, Türkiye
* Correspondence: burcu.yuksel@kocaeli.edu.tr

Abstract

The dependence of senescence labels on technical choices becomes apparent when they are used with sparse data from the human brain. We have assessed the SenFlag rules together with five different ICE initialisations in publicly available microglial single-nucleus RNA sequencing datasets. The discovery cohort included 12,220 nuclei obtained from 17 donors; the full ICE perturbation analyses were carried out on a fixed subset of 1823 nuclei. The external cohort consisted of 9371 nuclei from ten donors. All 85 complete-pipeline donor-removal fits were carried out. The median positive-label Jaccard agreement for SenMayo-initialised ICE was 95.3%, as compared with 80.0%, 29.4%, 17.6% and 1.7% for the ATF3, CDKN1A, CDKN2A and MX1 initialisations, respectively. Out of 40 attempts, 39 estimable chains were obtained under six matched molecule-thinning conditions and two MAGIC initialisation controls. The median agreement for SenMayo-initialised ICE was 95.5% at 50% count retention and 0.0% at 25%. The positive labels from the discovery cohort differed when using example constants as opposed to population-relative thresholds (91 versus 12) and decreased following thinning. In the external cohort, three ICE chains completed before depth filtering and one afterwards; no initialisation completed in both conditions. Technical stability, estimability and biological validity are separate concepts. Before biological interpretation, senescence-associated labels must have explicit operational definitions and robustness reporting.

Keywords: cellular senescence; microglia; Alzheimer disease; single nucleus RNA sequencing; ICE; SenFlag; technical robustness

Check for updates

Academic Editor: Maria Anttonietta Panaro

Received: 15 October 2026
Revised: 11 November 2026
Accepted: 13 November 2026
Published: 16 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

To identify senescent cells in human tissues, it is necessary to have an operational definition that links measurable molecular characteristics to the complex cellular state in question. Although consensus recommendations stress the limitations of relying on individual markers and highlight the importance of combining different types of features [1,2], in the case of human brain sequencing data, a transcriptomic label is often taken as proof that a specific nucleus is senescent. The reliability of this interpretation is determined by both the biology involved and by the stability of the computational method.

There are good biological reasons for investigating microglial senescence in Alzheimer's disease. Fancy and his colleagues have shown, using imaging and single-nucleus transcriptomics [3], that human microglia exhibit changes associated with senescence. However,

nuclear RNA profiles only include a portion of the cellular transcriptome. A comparison between human microglial cells and their nuclei found that transcripts related to activation were less easily detected in the nuclear preparations [4]. Although this finding refers to activation and not to senescence, it does call for a careful assessment of the marker coverage when transcriptional features are applied to postmortem brain nuclei.

The biological interpretation is complicated by the fact that microglia in AD can exist in a range of different states (amyloid-beta associated, tau associated) which can have either protective or detrimental effects depending on the context [5]. These disease-associated states in themselves do not equate to senescence. Senescence must be assessed by a convergence of evidence in the cell cycle, damage responses and secretory phenotypes, with the interpretation of the markers being in the context of the cell and the tissue [2]. More generally, p21-high and p16-high cells can have very different distributions, accumulation kinetics and functions [6]. This would therefore support looking at CDKN1A- and CDKN2A-based initialisations separately, whilst acknowledging that they are just the initial sets of cells that we are working with rather than independently validated senescent microglial subtypes.

There are now two different frameworks available for dealing with the classification problem. SenFlag combines transcriptional features using conjunctive rules [7]. ICE incorporates expression imputation with enrichment scoring, along with data-dependent cutoff estimation and marker refinement [8]. Another possible starting gene set is provided by the more extensive senescence-associated panel, SenMayo [9]. The various methods differ in whether they rely on the raw transcripts that have been detected or instead learn a gene set and a cutoff from the population being analysed. As a result, their outputs may react in different ways to the loss of molecules and to changes in the donors contributing to them.

ICE has already been applied to Alzheimer's disease microglia [8], and GSE264648 was generated to investigate brain senescence [3]. Our study evaluates technical label stability and computational estimability across two public human brain cohorts. It does not claim the first discovery of microglial senescence or the first application of ICE to Alzheimer's disease. We look at the effects of removing donors, carrying out a controlled thinning of the observed molecule counts, using random initialisation, and defining the SenFlag thresholds. Throughout the study, we keep the distinction between the reproducibility of the labels and their biological accuracy.

2. Results

2.1. Cohort Reconstruction and Feature Coverage

The discovery cohort contained 12,220 source-annotated microglial nuclei from 17 donors: nine with Alzheimer's disease and eight controls. The nuclei were distributed across 49 donor-region samples. The source matrix had 24,748 gene features. The fixed, diagnosis-independent pilot study consisted of 2431 nuclei, and the part of it that included at least 1000 unique molecular identifiers (UMIs) contained 1823 nuclei taken from all 17 donors and all 49 samples. The threshold was introduced following the observation of low-depth concentration and is regarded as a post hoc sensitivity condition.

The external cohort consisted of 9371 microglial nuclei obtained from ten male donors and 20 samples of the entorhinal cortex or the superior frontal gyrus [10]. The fixed pilot portion contained 1000 nuclei; after applying the same UMI threshold, only 222 nuclei from nine donors and 13 samples were left. The filter kept 13 of the 300 control-labelled pilot nuclei and 209 of the 700 Alzheimer-labelled nuclei. However, these annotations were not regarded as independent confirmation of clinical dementia.

The original GEO count matrices were employed to deal with the incomplete representation of features in the externally processed resource. All 9371 source-annotated barcodes were retrieved, and the total count for each nucleus corresponded to the source

nUMI value. When the repeated gene symbols were aggregated, 33,660 unique external symbols were obtained. All 22 required SenFlag genes were found in the recovered matrix; specifically, RRM2 was detected at a count of one in each of two nuclei, even though its previous lack of presence in the processed representation had not been recorded as zero expression. The two feature dictionaries that were analysed had 16,902 symbols in common. The required gene detection frequencies are given in Table S8. Cohort characteristics and analysis denominators are summarised in Table 1.

Table 1. Cohorts and analysis denominators.

Cohort	Donors	Samples	All Microglia	Pilot	UMI Subset
GSE264648	17	49	12,220	2431	1823
GSE147528	10	20	9371	1000	222

2.2. Baseline Labels Depend on Threshold Definition and ICE Initialisation

The fixed constants used in the SenFlag human example found 91 nuclei in the discovery cohort. When a population-relative method was applied using gene-specific 75th-percentile low-expression thresholds, only 12 nuclei were identified. Within the subset of 1823 nuclei examined for sensitivity, the two approaches detected 21 and four nuclei, respectively. The two implementations therefore give rise to different label counts; neither of them has been calibrated against a biological reference in these nuclei.

The additional percentile sensitivity analysis confirmed that the population-relative labels are a function of the chosen cutoff (Table S10). In the full discovery cohort, the 50th, 60th, 70th, 75th, 80th, 90th and 95th percentiles produced 3, 8, 8, 12, 18, 40 and 61 positive nuclei respectively. The 70th and 80th percentile classifications each had 66.7% positive-label Jaccard agreement with the 75th percentile reference. In the 1823 nucleus discovery subset, the same settings produced 2, 2, 2, 4, 5, 10 and 18 positives. All seven settings produced zero positives in each of the external scopes. The external positive-label Jaccard comparisons were undefined (both positive sets were empty) rather than evidence of perfect positive-label agreement.

The ICE initialisations were completed in both the pilot study, which had 2431 nuclei, and its subset consisting of 1823 nuclei. When ICE was initialised using SenMayo, it identified 202 positives before filtering and 42 after refitting on the subset. However, 42 of the original positives came from nuclei that were retained, and 41 of these remained positive after refitting, which yielded a matched positive-label Jaccard index of 95.35%. The respective indices were 52.17% for ATF3, 74.07% for CDKN1A, 29.27% for CDKN2A and 0% for MX1. Hence, the loss of nuclei from a dataset and the reassignment of retained nuclei result in different outcomes.

In the filtered pilot, between-initialisation positive-label Jaccard indices ranged from 0.0% to 83.3% (Table S7). These overlaps describe convergence between methods; neither SenFlag nor any ICE initialisation was treated as a biological reference standard.

2.3. Full-Pipeline Donor Removal Reveals Initialisation-Specific Instability

The entire ICE chain was relearned by removing each donor and all the regions associated with them. All 85 of the initialisation-by-donor fits were carried out. The comparisons made use of only those nuclei that remained after the donors had been removed. SenMayo-initialised ICE achieved a median positive-label Jaccard agreement of 95.3%, with a minimum of 91.7%. At most, two labels changed in any one donor-removal comparison. Agreement was, however, lower for ATF3 and considerably lower for CDKN1A, CDKN2A and MX1 (see Table 2 and Figure 1).

Table 2. The influence of donors throughout the entire pipeline. J stands for Jaccard.

ICE Start	Complete	Median Label J (%)	Minimum Label J (%)	Median Gene J (%)
SenMayo	17/17	95.3	91.7	78.6
ATF3	17/17	80.0	34.8	57.1
CDKN1A	17/17	29.4	0.0	20.3
CDKN2A	17/17	17.6	2.1	16.7
MX1	17/17	1.7	0.0	15.9

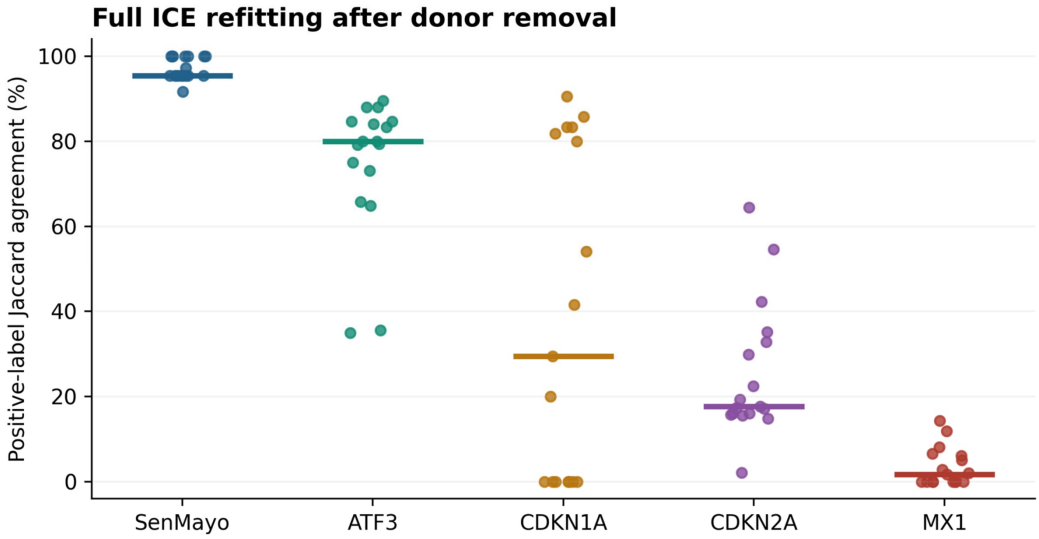


Figure 1. The stability of the positive labels is shown after a complete ICE refitting procedure in which one donor has been removed. Each point corresponds to one of the 17 overlapping donor-removal scenarios for each initialisation. The comparisons involve only the nuclei that were retained. These observations are technical influence analyses, not independent experiments, withheld predictions or estimates of biological accuracy. The horizontal bars indicate the medians.

Removing one donor at a time tests whether the classification depends heavily on a particular individual. This approximates a change in the people contributing to a dataset; it does not assess prediction in a new patient.

The median Jaccard indices for the gene sets obtained were 78.6%, 57.1%, 20.3%, 16.7% and 15.9% for SenMayo, ATF3, CDKN1A, CDKN2A and MX1 respectively. Label stability and gene-set stability thus offered complementary information. For instance, when donor A019/12 was removed, there was no positive-label overlap for the CDKN1A and MX1 initialisations, yet both refits produced non-empty positive sets. None of the donors was excluded from the study because of these findings.

For the population-relative SenFlag implementation, gene-specific 75th percentiles were recalculated after each donor was removed. Labels for the retained nuclei were unchanged in all 17 comparisons. This result is based on only four initially positive nuclei. The fixed example rules are deterministic for a retained nucleus and do not provide an independent measure of robustness to donor removal.

2.4. Controlled Molecule Loss and Random Initialisation

Here, molecule thinning randomly removes a proportion of the recorded RNA counts whilst keeping the same nuclei. That approximates lower transcript capture or incomplete molecule detection in a single-nucleus sequencing experiment. It is a controlled technical stress test rather than a model of a change in the underlying biological state of the nuclei.

The eight planned count/initialisation perturbations were all completed. Of the 40 chains in the five ICE starting sets, 39 produced estimable final labels. The CDKN2A chain at 25% retention with thinning seed 202610131 produced a non-finite final cutoff. The non-estimable results are listed in Table S1 and were not assigned zero positives. For the SenMayo-initialised ICE, the matched positive-label Jaccard agreement was 95.5% in each of the three 50% count retention runs and 0.0% in each of the three 25% runs. In each of the three 50% runs, 44 positive nuclei were identified, whilst in each of the three 25% runs, 19–39 were identified. The zero Jaccard agreement at 25% retention simply means that the positive nuclei identified were different from those in the full count reference, not that no positive nuclei were identified. Results for all five starting sets are summarised in Table 3 and visualised in Figure 2. For the two full count MAGIC seed controls, the matched label agreement ranged from 0.0% to 100.0%. These are the controls which assess the effect of initialisation on the results whilst keeping the observed counts constant. The active gene counts and coverage of the original SenMayo set are given for each condition in Table S3.

With the fixed human-example constants, the 21 baseline positives decreased to 1–5 at 50% count retention and to 0–1 at 25%. With the population-relative adaptation, the four baseline positives decreased to 0–1 at either retention level(Figure 3). The paired comparisons include all the nuclei and provide a more direct indication of the effects of the simulated loss of molecules than the exclusion of nuclei with low UMI counts. Yet, binomial thinning is unable to tell the difference between biological absence and molecules that were not detected by the original assay.

Table 3. Agreement based on matched controlled perturbations. The chains which cannot be estimated stay in the denominators of the attempt figures.

ICE Start	Condition	Estimable	Median J (%)	Range J (%)
SenMayo	50%	3/3	95.5	95.5–95.5
ATF3	50%	3/3	66.7	60.5–75.9
CDKN1A	50%	3/3	76.0	1.6–78.3
CDKN2A	50%	3/3	21.5	0.0–26.1
MX1	50%	3/3	0.0	0.0–31.4
SenMayo	25%	3/3	0.0	0.0–0.0
ATF3	25%	3/3	58.1	48.3–70.4
CDKN1A	25%	3/3	1.5	1.4–45.0
CDKN2A	25%	2/3	25.0	13.2–36.8
MX1	25%	3/3	0.0	0.0–0.0
SenMayo	Full + new seed	2/2	97.7	95.3–100.0
ATF3	Full + new seed	2/2	78.7	77.4–80.0
CDKN1A	Full + new seed	2/2	47.6	0.0–95.2
CDKN2A	Full + new seed	2/2	30.7	29.4–31.9
MX1	Full + new seed	2/2	5.5	1.7–9.4

2.5. External Cohort Reveals Limits of Computational Estimability

In the external cohort, both SenFlag implementations yielded no positives amongst all 9371 nuclei, and this was true in both pilot studies. With the example-rule sequence, 12 nuclei passed the combined gates before the final cell-cycle-inhibitor requirement, but none of them also satisfied the CDKN1A-or-CDKN2A detection. It therefore cannot be concluded that senescent microglia are absent.

In the external pilot of 1000 nuclei, three of five ICE chains were complete—one of five in the 222-nucleus sensitivity subset. No starting set produced an estimable chain in both conditions (Table 4). Matched positive-label Jaccard agreement could therefore not be calculated for any of the external starting sets. The non-estimable results were empty

refined gene sets, insufficient initial groups and invalid cutoffs. No alternative thresholds were chosen in order to get a result.

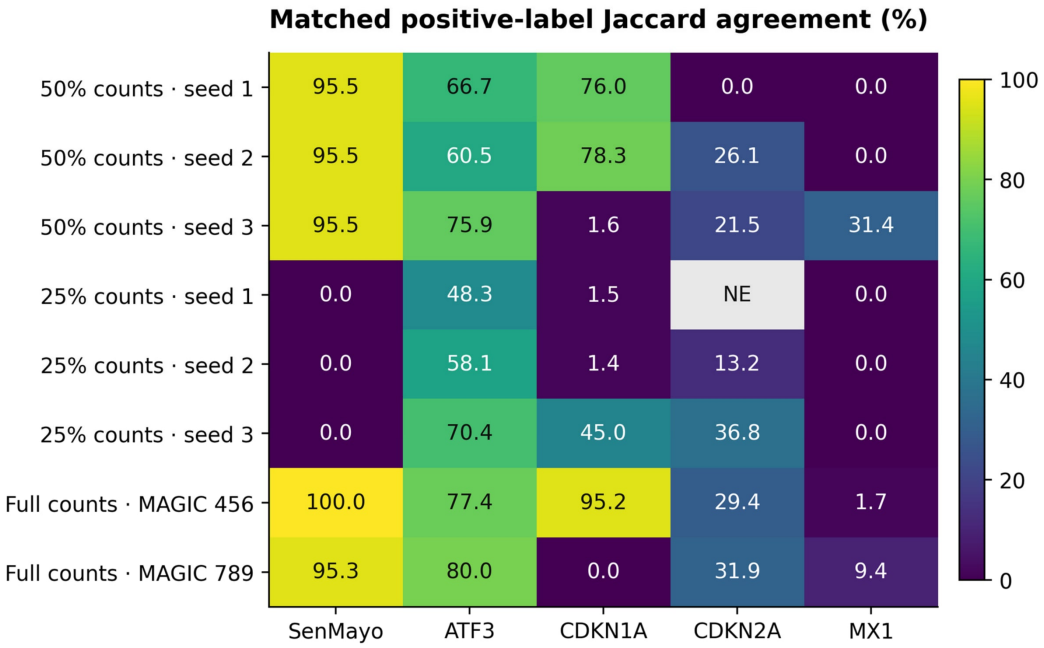


Figure 2. The Jaccard agreement for the positively labelled cases after controlled thinning and changes to the MAGIC initialization is shown. In each case, the same 1823 nuclei are kept. Thinning of the counts was carried out at retention probabilities of 0.50 and 0.25 using three random seeds; in two further fits, all the counts were retained and only the MAGIC seed was changed. All the ICE steps were relearned. NE means that the chain did not result in an estimable final cutoff, not that the agreement is zero. The control runs consist of technical repetitions and not of independent biological replicates.

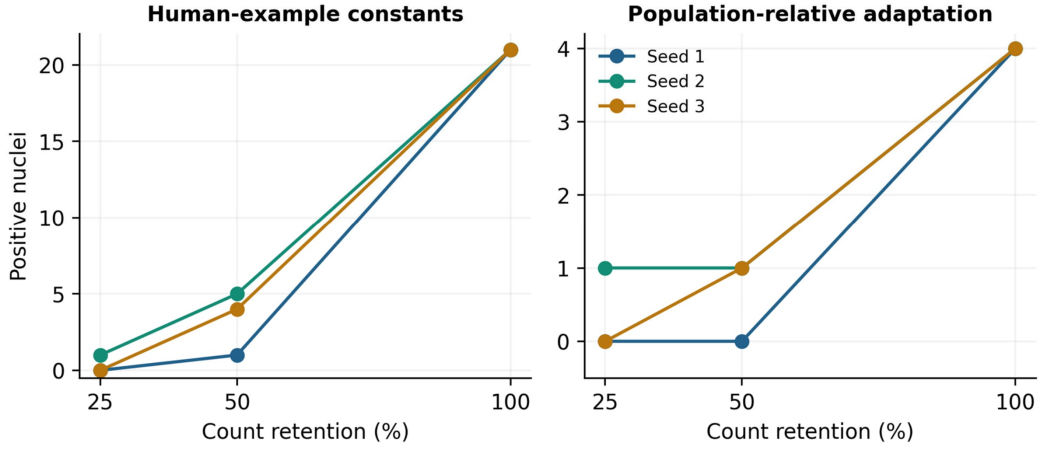


Figure 3. The number of positive SenFlag counts is shown for the molecule-thinning process in the fixed discovery subset. The lines indicate the paired thinning-seed labels rather than the biological trajectories. For the reference case, there are 21 positives when the fixed human-example constants are used and four when population-relative adaptation is applied. At each reduced depth, the same nuclei are kept and the population-relative thresholds are reapplied. The positive counts do not represent the actual number of senescent cells.

Table 4. External ICE outcomes. NE denotes non-estimable, rather than zero positive nuclei.

ICE Start	Positive in 1000	Positive in 222	Matched Jaccard
SenMayo	NE	NE	Not estimable
ATF3	42	NE	Not estimable
CDKN1A	90	NE	Not estimable
CDKN2A	NE	3	Not estimable
MX1	95	NE	Not estimable

The additional subsampling audit comprised four 222-nucleus conditions with 20 fits per condition (Table S11). The original composition and filtered composition subsets gave 53/100 and 55/100 estimable initialisation chains, respectively. The UMI-bin matched counterparts gave 47/100 and 54/100. There was a difference in completion rates depending on the starting set. Of the 20 UMI matched pairs for each of the five starting sets, 49/100 of the paired initialisation comparisons had a discordant completion status (21 completed only under the original composition and 28 only under the filtered composition).

Of the 191 non-estimable subsampling chains, 94 returned invalid/empty refined gene sets. A total of 90 had an initial score positive group of just two nuclei (below the minimum required for marker refinement). Six had an absent or constant starting gene and one had a non-finite final cutoff. The small groups were defined by ICE scores, not by control or Alzheimer annotation.

When the same original 222 filtered nuclei were kept and their counts thinned, each fit gave rise to two or three estimable starting sets as opposed to just one in the full count case. This was true at both 50% and 25% count retention. Only two of the matched label comparisons were estimable in both the full count reference and the thinned fit (both for CDKN2A init). Their positive-label Jaccard agreements were 16.7% at 50% retention and 18.2% at 25% retention. All the other comparisons were non-estimable rather than assigned zero agreement (Table S11).

The audit, which involved an external examination of the feature-universe, recomputed the Bonferroni-adjusted p-values for markers using all 33,660 symbols. The order of the gene sets that had been learned by the four completed chains did not change. It thus follows that the difference in feature representation could not account for the completion pattern that was reported. Full details can be found in the Supplementary Outputs.

3. Discussion

This secondary analysis reveals several distinct aspects of technical robustness concerning transcriptomic senescence labelling. The same nuclei could end up with very different labels when the donor composition was altered, when the number of molecules was decreased, or when the threshold rules were applied in a different way. Moreover, the successful completion of a chain, the stability of its positive labels and the stability of its learned genes were all separate outcomes.

The difference between the fixed example and the population-relative classifications was not just apparent at the 75th percentile. The percentile sweep also had an effect on the positive counts and the label membership within the population-relative approach. A lower percentile will result in a more restrictive low-expression gate. Sparse integer counts also lead to tied thresholds and plateaus: at the 75th percentile, 12 of the 14 low-expression cutoffs were zero in the full discovery cohort and all 14 were zero in the full external cohort (Table S10). An unchanging label across some percentiles can therefore be due to the identical numerical cutoffs rather than independent evidence of biological robustness. These results do support the reporting of the percentile, the resulting gene-specific cutoffs

and the positive-label overlap. They do not establish the 75th percentile or any of the other percentiles tested here as the optimal biological threshold.

The ICE method initialised with SenMayo demonstrated considerably better label stability than the four single-gene initialisations when the donors in the discovery cohort were removed. However, this finding does not prove that it has greater sensitivity or specificity. It is possible for a method to consistently label the incorrect state, and different starting genes can correspond to separate stress or inflammatory programmes. Additionally, the marker refinement process replaces the initial set, so SenMayo-initialised ICE should not be referred to as an unmodified SenMayo score. The donor-removal results pertain to the learned pipeline and not to the original panel's inherent validity.

Controlled thinning separates the effect of retaining fewer RNA counts from the effect of removing nuclei. The same nuclei remained in every condition, yet their labels changed. With respect to ICE initialised with SenMayo, the median Jaccard index at 25% count retention was 0.0%, as opposed to 95.5% at 50%. SenFlag's requirement for co-detection of several transcripts also produced fewer positives as counts were reduced. These effects demonstrate the sensitivity to the available measurements; they do not prove that the labels which were lost were false positives or that the labels which were retained were true cases of senescence. The numerical-seed controls also show that count depth and computational initialisation should be considered separately.

The external analysis gives us information about computational estimability. It is not an external validation of a locked predictor as ICE was relearned in each condition. The UMI filter also altered the external sample composition significantly. The differences between conditions cannot therefore be all down to sequencing depth. We do not treat external non-estimability as zero agreement and we do not use the external dataset to validate the discovery ranking.

The external subsampling controls show that the original filtering contrast cannot be ascribed to control depletion alone. Non-estimability also arose in subgroups that preserved the original pilot composition. Reproducing the filtered donor-region composition did not consistently lead to a reduction in completion (in fact, it did so in both directions). The differences between the schemes were scheme-specific and in both directions. These are the descriptive sensitivity results from overlapping resamples taken from a fixed cohort and do not constitute an estimate of an AD effect.

The same nucleus thinning controls show that computational estimability is not a monotonic function of RNA capture depth. More chains were completed after thinning, yet the two estimable matched label comparisons showed very poor agreement. UMI-bin matching controls the overall library size distribution (not the gene expression profiles, diagnosis specific depth or any of the other assay quality properties) and only applies to library size distributions which are shared by the sampling schemes. Only 13 of the control-labelled nuclei in the original pilot met the 1000 UMI threshold so the original control proportion cannot be recovered at the fully filtered depth without replacement. The new analyses therefore account for a number of technical effects, whilst leaving their individual contribution to the original filtering contrast unaccounted for.

Before you apply ICE to a new dataset, you should check the feature ids and the raw count representation, check that the starting genes are present and detected, and check the UMI distributions and the number of nuclei per donor and region. A small pilot under prespecified conditions can be run to check that there is a finite initial cutoff, that there are sufficiently populated initial groups to enable marker selection and that the refined gene set is non-empty. The final scores and final cutoff should be checked for finite values. These checks may identify the cause of non-estimability prior to a larger analysis but cannot guarantee that the analysis will complete or that the results will be biologically valid. The

reasons for failure should be recorded. Thresholds should not be changed simply to get some positive assignments upon inspection of the disease labels.

These results support the idea of separating transcriptional evidence from a cellular diagnosis. The initial study that made the discovery included associations between imaging and pathology [3]; however, those measurements were not connected to the individual nucleus labels in our case. Standards for detecting senescence call for evidence that goes beyond a single computational assignment [1]. The present analyses do not contradict the biological observations of the source study and do not show that any of the initialisation measures in question detect senescence more accurately.

Experimental approaches can complement transcriptomic classifications by measuring cell-cycle arrest, senescence-associated β-galactosidase activity and secretory phenotypes, as outlined for cultured mammalian cells by Kang et al. [11]. These assays require validation for the tissue and cell type being studied, consistent with the in vivo senescence guidelines of Ogrodnik et al. [2]. We did not perform these assays; the present results therefore concern computational robustness and cannot establish a senescent microglial phenotype.

Many factors prevent the results from being generalised. Instead of carrying out a full ICE refitting on all 12,220 discovery nuclei, it was done on a fixed pilot set. The UMI threshold and population-relative SenFlag adaptation were post hoc methodological choices. The external cohort was small, consisted only of men and was highly sparse, having an age and pathology distribution which is not adequate to allow any broad conclusions about the disease effect. The donor-removal fits include most of the same training nuclei and therefore cannot be regarded as 17 independent validations. The molecule-thinning repetitions together with two further numerical seeds provide only limited perturbation checks and not a complete parameter search.

Furthermore, the way in which ICE selects its internal markers involves making nucleus-level comparisons between groups that have been defined by algorithms. These p-values form part of the learning rule and are not evidence at the donor level of differential expression associated with Alzheimer's disease. To assess the procedure's behaviour, we kept the original method and did not treat the selection p-values as biological inferences. It is necessary to make this distinction since failing to take biological replication into account can lead to an overestimation of the differential-expression evidence at the single-cell level [12]. We give a description of the agreement without providing disease-effect p-values or confidence intervals that are derived from the overlapping technical replicates.

This study cannot determine whether Alzheimer's disease increases microglial senescence. No donor-level disease-effect model was fitted, and the nuclei were not linked to an independent biological reference for senescence. Positive-label counts and technical agreement therefore cannot establish either an increase or an absence of disease-associated senescence. Answering that question would require an appropriately controlled analysis with donors as the biological replicates and independent validation of the senescence phenotype.

The practical consequence is that the full operational definition together with its failure conditions must be documented before senescence labels are used in a biological association analysis. At the very least, studies should give information on required-gene coverage, changes to the denominator, the overlap between matched labels, the overlap between the learned gene sets and all other non-estimable conditions. When assessing external transfer, missing features should not be regarded as unexpressed genes, or the thresholds should not be altered after looking at the disease labels. With these reporting options, it is still possible to evaluate robustness even if no nucleus-level biological truth is available.

4. Materials and Methods

4.1. Study Design and Public Data

A computational method evaluation was carried out using processed public human brain data from GSE264648 [3] and GSE147528 [10]. The original microglial annotations were kept, and no other cell types were relabelled. The discovery dataset includes the entorhinal cortex, the middle temporal gyrus and the somatosensory cortex; the external dataset includes the entorhinal cortex and the superior frontal gyrus. The source count totals, the barcode matching, the feature dictionaries and the source-file hashes were all checked. The public, de-identified metadata were used; no additional participants or experimental specimens were recruited.

In the fixed discovery pilot, up to 50 nuclei were sampled per donor-region combination without taking the diagnosis into account. The external pilot also sampled 50 per combination, using seed 20261012. The existing lists of pilot cells were kept when carrying out later comparisons. A raw UMI total of at least 1000 was used to define a sensitivity subset, not as a universally validated quality threshold. Nuclei that were removed were not put back in. The source annotations and the original quality control were retained; however, additional verification of donor overlap at the genotype level was not available. Although the cohort provenance was different, it was not claimed that there had been a complete exclusion of donor overlap.

4.2. Feature Reconstruction and Normalisation

The discovery matrix had 24,748 features; the original GEO 10× count files used for external data were matched to the microglial barcodes provided by the author. Symbols that occurred more than once were added together, with the Ensemble-to-symbol provenance kept. There was no automatic conversion to the current symbols. A lack of entry in the feature dictionary was treated differently from a feature that was present but had zero observed counts. When the original and the processed external matrices were compared, the 31,988 Ensemble identifiers that were common to both were considered; nine of the current identifiers per region were outside the scope of this comparison.

For ICE, each nucleus was normalised by taking the logarithm of one plus ten thousand times the number of its gene counts divided by its total UMI count. All the original feature rows were included in both the raw normalisation and the marker-selection objects. When carrying out MAGIC and enrichment scoring, only those genes which showed expression in at least one of the analysed nuclei were used. Neither cohort integration nor diagnosis-specific batch correction was performed. Separate records were kept for gene detection frequencies and active gene-set coverage.

4.3. Operational SenFlag Implementations

We evaluated the fixed raw-count constants in the authors' human example and a population-relative adaptation [7]. The 75th percentile was used because the original SenFlag methods specify this percentile for HMG genes and apply the same strategy to NFIA [7]. We used it as a population-relative reference, not a validated optimum for human microglia. Extending the same percentile to all ten proliferation genes was our operational choice, not a uniquely specified SenFlag rule. This setting was not fitted to disease labels. Low HMG, proliferation, and NFIA expression were combined with detection of at least one V-ATPase subunit, one cathepsin, CCND1, and either CDKN1A or CDKN2A. The exact gene lists and fixed example constants are provided in the archived code. Population thresholds were recalculated within each analysis scope, donor-removal subset, and thinned count matrix. No imputation was used for SenFlag.

In a post hoc sensitivity analysis added during peer review, we varied the common percentile for the 14 low-expression genes across the 50th, 60th, 70th, 75th, 80th, 90th, and 95th percentiles. This grid was fixed before generating the new sweep results. The raw counts, gene lists, positive-detection gates, and original nucleus selections were held constant. Percentiles included zero counts and used linear interpolation equivalent to R quantile type 7. Counts at or below the cutoff passed the low-expression gate. Cutoffs were recomputed separately within the full cohort, balanced pilot, and UMI-filtered pilot for each dataset. We recorded positive counts, gene-specific cutoffs, and positive-label Jaccard overlap with the 75th-percentile reference and fixed example rules (Table S10). Empty positive unions were undefined. No disease comparison or selection of an optimal threshold was performed.

4.4. ICE Learning and Computational Failure Rules

The ICE functions run_gsea, run cutoff and gs_generate were kept in the commit f200599156d3834573eb3cf40a0fc00d2edd0652 [8]. The chain that was evaluated consisted of MAGIC, the initial enrichment and cutoff, one pass of source marker refinement and finally enrichment and cutoff. Five starting sets were specified: SenMayo and four single-gene initialisations based on ATF3, CDKN1A, CDKN2A or MX1. Each single-gene initialisation was expanded to 20 genes using the highest finite Pearson correlations, including the target gene. This refers to the implementation that was evaluated, not a claim that all possible ICE settings have been benchmarked.

For MAGIC 3.0.0, a diffusion time of 3, 100 principal components and a random state value of 123 were used, except for the initialization controls [13]. The ssGSEA analysis used GSVA 1.50.0 with normalise = FALSE [14]. For each perturbed matrix, the initial correlation expansion, the active gene membership, the marker refinement and both cutoffs were recalculated. The source cutoff settings were start = 0.01, end = 0.6, peak threshold 0.5 and a minimum peak distance of 2% of the analysed nuclei. In Seurat 5.0.1, the marker selection kept genes that had an average log2 fold change greater than 0.25, a positive-group detection above 0.5 and a Bonferroni-adjusted p-value below 0.05, with a maximum of 50 source-ordered genes. These thresholds were not adjusted in relation to diagnosis.

Invalid or empty refined sets, initial groups with too few nuclei, and non-finite cutoffs were recorded as non-estimable. No fallback percentile or rescued gene set was used. The valid final scores were verified by referring to their stored cutoffs to reproduce the labels precisely. Before carrying out the donor-removal analysis, the reference scores, the ordered learned genes and the labels were reproduced. Together with the software session information, the seeds and all the condition-level outcomes are included in the code archive.

4.5. Perturbation Designs

In the case of donor influence, all the regions associated with each of the 17 discovery donors were eliminated. The entire chain was then relearned using the remaining nuclei without carrying out any replacement sampling. No predictions were made for the donor that had been omitted. The comparison reference was the part of the 17-donor fit that had been retained. A separate fixed-score control was used to recalculate only the thresholds; it is given as supplementary diagnostic evidence and not regarded as a full-pipeline test.

To achieve a controlled depth, each of the observed integer gene counts was independently thinned according to retention probabilities of 0.50 and 0.25, using the seeds 202610131, 202610132 and 202610133. The same 1823 nuclei were retained, and the UMI filter was not reapplied. Shared seeds across depths produced paired random-number streams, not nested molecule samples. Two additional analyses employed all the original counts together with MAGIC seeds 456 and 789. Any nucleus with a zero total UMI count

would have been treated as a technical failure rather than excluded or assigned invented counts. No such nucleus occurred. Although the plan and the rules regarding failure were established before these new outputs were generated, the audit had been prompted by earlier results and was not preregistered.

We performed a post hoc subsampling audit of the external cohort. Subsets of 20,222 nuclei were generated, which approximately preserved the original pilot's donor-region proportions (67 control-labelled, 155 Alzheimer's labelled nuclei). A total of 20 further subsets exactly replicated the filtered donor-region counts (13 control-labelled, 209 Alzheimer's labelled nuclei). Both conditions were samples without replacement from the original unfiltered 1000-nucleus pilot. A second set of 20 paired comparisons retained the composition targets and matched the overall UMI distributions in the 100 UMI bins (Table S11). Joint integer allocation was used to select the feasible donor region by UMI-bin quotas which were closest to the original pilot histogram scaled to 222 nuclei. The selection rules and random seeds were all fixed prior to generating the new ICE outcomes.

For a separate depth control, we kept the same original 222 filtered nuclei. We binomially thinned all the observed gene counts at 50% and 25% retention with three fixed seeds per level. The UMI filter was not reapplied. Each subset and each thinning condition relearned the full ICE chain under the original settings. Computational estimability and failure stages were recorded separately for each starting set. The two original external fits were re-run as reference checks. The completion frequencies are just descriptive summaries of the overlapping technical resamples; they are not disease-effect estimates and are not independent biological replicates.

4.6. Agreement and Scope of Statistical Inference

Positive-label agreement was measured by Jaccard overlap using matched nuclei. Jaccard overlap is the number of nuclei positive in both assignments divided by the number positive in either assignment. Comparisons with an empty positive union were undefined. The number of gained and lost labels, the absolute number of positives, and the learned gene-set overlap were also recorded. We did not choose the overall negative-dominated agreement as the main measure of stability. The comparisons between methods were merely descriptive and did not designate any one method as the ground truth.

The structure at the donor level was kept unchanged. In the case of summarised rates, the nuclei were first grouped according to donor-region samples, and the different regions were given equal weight within each donor; pooled counts were clearly stated as pooled. There was no Alzheimer's disease-control, pathology or senescence-accuracy confirmation analysis carried out. The internal marker-selection statistics were used solely for the purpose of computational refinement. The technical-run ranges and the donor-removal medians are not confidence intervals. The tables at the condition level include outcomes that cannot be estimated rather than removing them from the denominator of the completion rate.

5. Conclusions

Transcriptomic senescence labels in human brain microglia are dependent on the operational rules, the learning initialisation and the data that you analyse. SenMayo-initialised ICE was comparatively stable through the process of discovery cohort donor removal, whereas the other initialisations were much more variable. Loss of controlled molecules altered the labels but the nuclei remained the same. The external cohort showed non-estimable conditions. This does support the need for explicit reporting of technical robustness prior to any downstream biological interpretation. It does not establish that this is the best senescence diagnostic method, nor does it establish that there is an increase in senescent microglia in Alzheimer's disease.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms272210273/s1.

Author Contributions: Conceptualisation, B.Y. and L.A.; methodology, B.Y. and L.A.; software, B.Y. and L.A.; validation, B.Y. and L.A.; formal analysis, B.Y. investigation, B.Y.; resources, B.Y. and L.A. data curation, B.Y. and L.A.; writing—original draft preparation, B.Y. and L.A.; writing—review and editing, B.Y. and L.A.; visualisation, B.Y. and L.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study involved only the secondary analysis of publicly available, de-identified human brain single-nucleus RNA sequencing datasets (GEO accessions GSE264648 and GSE147528). No new human participants or animals were recruited or studied, no new biological specimens were collected, and no identifiable personal information was accessed.

Informed Consent Statement: Not applicable. Informed consent was not required for this study because it involved only the secondary analysis of publicly available, de-identified datasets (GEO accession numbers GSE264648 and GSE147528). No new participants were recruited, and no identifiable personal information was accessed.

Data Availability Statement: The source data can be obtained via the GEO accessions GSE264648 and GSE147528. The processed derived labels, the fixed cell selections, the analysis plans, the source identifiers and the scripts are included in the reproducibility archive. Large regenerable MAGIC intermediates and the installed software packages are not included; the settings used to generate them, and the session information are provided. The reproducibility archive listed above can be provided as Supplementary Material in the journal; if separate deposition is carried out, a public repository identifier can be added.

Acknowledgments: During the preparation of this work, the authors used ChatGPT (GPT-6) to support language refinement, clarity, and organisation of the text. All scientific content, analyses, interpretations, and conclusions were conceived, verified, and finalised by the authors. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Conflicts of Interest: Author Levent Aydin had secondary employment with Eureka Patent Ltd. Co., Türkiye. The remaining 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. Also, we did not receive any funding for this study.

References

1. Gorgoulis, V.; Adams, P.D.; Alimonti, A.; Bennett, D.C.; Bischof, O.; Bishop, C.; Campisi, J.; Collado, M.; Evangelou, K.; Ferbeyre, G.; et al. Cellular Senescence: Defining a Path Forward. *Cell* **2019**, *179*, 813–827. [CrossRef]
2. Ogrodnik, M.; Carlos Acosta, J.; Adams, P.D.; d'Adda di Fagagna, F.; Baker, D.J.; Bishop, C.L.; Chandra, T.; Collado, M.; Gil, J.; Gorgoulis, V.; et al. Guidelines for minimal information on cellular senescence experimentation in vivo. *Cell* **2024**, *187*, 4150–4175. [CrossRef]
3. Fancy, N.N.; Smith, A.M.; Caramello, A.; Tsartsalis, S.; Davey, K.; Muirhead, R.C.J.; McGarry, A.; Jenkyns, M.H.; Schneegans, E.; Chau, V.; et al. Characterisation of premature cell senescence in Alzheimer's disease using single nuclear transcriptomics. *Acta Neuropathol.* **2024**, *147*, 78. [CrossRef]
4. Thrupp, N.; Sala Frigerio, C.; Wolfs, L.; Skene, N.G.; Fattorelli, N.; Poovathingal, S.; Fourne, Y.; Matthews, P.M.; Theys, T.; Mancuso, R.; et al. Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep.* **2020**, *32*, 108189. [CrossRef]
5. Bang, J.Y.; Yoo, Y. Rationale and emerging evidence for microglial replacement in Alzheimer's disease. *Mol. Cells* **2025**, *48*, 100265. [CrossRef]
6. Yan, J.; Chen, S.; Yi, Z.; Zhao, R.; Zhu, J.; Ding, S.; Wu, J. The role of p21 in cellular senescence and aging-related diseases. *Mol. Cells* **2024**, *47*, 100113. [CrossRef]

7. Altulea, A.; Mackedenski, S.; Nehme, J.; Borghesan, M.; Bird, T.G.; Demaria, M. SenFlag gene signature identifies senescent cells in mouse and human tissues through a conserved core transcriptional program. *EMBO J.* **2026**, *45*, 5580–5606. [CrossRef]
8. Xu, P.; Zhang, H.; Zhu, S.; Kong, Y. ICE: Robust detection of cellular senescence from weak single-cell signatures using imputation-based marker refinement. *Genome Biol.* **2026**, *27*, 89. [CrossRef]
9. Saul, D.; Kosinsky, R.L.; Atkinson, E.J.; Doolittle, M.L.; Zhang, X.; LeBrasseur, N.K.; Pignolo, R.J.; Robbins, P.D.; Niedernhofer, L.J.; Ikeno, Y.; et al. A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nat. Commun.* **2022**, *13*, 4827. [CrossRef]
10. Leng, K.; Li, E.; Eser, R.; Piergies, A.; Sit, R.; Tan, M.; Neff, N.; Li, S.H.; Diehl Rodriguez, R.; Suemoto, C.K.; et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* **2021**, *24*, 276–287. [CrossRef]
11. Kang, E.; Kang, C.; Lee, Y.-S.; Lee, S.-J.V. Brief guide to senescence assays using cultured mammalian cells. *Mol. Cells* **2024**, *47*, 100102. [CrossRef]
12. Squair, J.W.; Gautier, M.; Kathe, C.; Anderson, M.A.; James, N.D.; Hutson, T.H.; Hudelle, R.; Qaiser, T.; Matson, K.J.E.; Barraud, Q.; et al. Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **2021**, *12*, 5692. [CrossRef]
13. van Dijk, D.; Sharma, R.; Nainys, J.; Yim, K.; Kathail, P.; Carr, A.J.; Burdziak, C.; Moon, K.R.; Chaffer, C.L.; Pattabiraman, D.; et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **2018**, *174*, 716–729.e27. [CrossRef]
14. Hänzelmann, S.; Castelo, R.; Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **2013**, *14*, 7. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.