


RESEARCH

Open Access



Somatic copy number variant load in neurons of healthy controls and Alzheimer's disease patients

Zeliha Gözde Turan^{1*} , Vincent Richter², Jana Bochmann², Poorya Parvizi^{1,3}, Etkä Yapar⁴, Ulas Işıldak¹, Sarah-Kristin Waterholter², Sabrina Leclere-Turbant⁵, Çağdaş Devrim Son¹, Charles Duyckaerts⁵, İdil Yet^{6†}, Thomas Arendt^{2†}, Mehmet Somel^{1†} and Uwe Ueberham^{2*†}

Abstract

The possible role of somatic copy number variations (CNVs) in Alzheimer's disease (AD) aetiology has been controversial. Although cytogenetic studies suggested increased CNV loads in AD brains, a recent single-cell whole-genome sequencing (scWGS) experiment, studying frontal cortex brain samples, found no such evidence. Here we re-addressed this issue using low-coverage scWGS on pyramidal neurons dissected via both laser capture microdissection (LCM) and fluorescence activated cell sorting (FACS) across five brain regions: entorhinal cortex, temporal cortex, hippocampal CA1, hippocampal CA3, and the cerebellum. Among reliably detected somatic CNVs identified in 1301 cells obtained from the brains of 13 AD patients and 7 healthy controls, deletions were more frequent compared to duplications. Interestingly, we observed slightly higher frequencies of CNV events in cells from AD compared to similar numbers of cells from controls (4.1% vs. 1.4%, or 0.9% vs. 0.7%, using different filtering approaches), although the differences were not statistically significant. On the technical aspects, we observed that LCM-isolated cells show higher within-cell read depth variation compared to cells isolated with FACS. To reduce within-cell read depth variation, we proposed a principal component analysis-based denoising approach that significantly improves signal-to-noise ratios. Lastly, we showed that LCM-isolated neurons in AD harbour slightly more read depth variability than neurons of controls, which might be related to the reported hyperploid profiles of some AD-affected neurons.

Keywords: Single-cell whole-genome sequencing, Copy number variation, Alzheimer's disease, Brain, Laser capture microdissection, Fluorescence-activated cell sorting, Denoising

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease of multifactorial aetiology, with numerous genetic and

environmental factors each explaining a small proportion of variance in disease onset and progression [1]. One of the less-studied potential contributors is somatic copy-number variations (CNVs) in neurons, which can include the gain or loss of whole chromosomes (aneuploidy) or of chromosomal segments. It is generally accepted that mature neurons in healthy brains can carry somatic CNVs, but their frequency is uncertain. Early studies estimated aneuploid neuron frequencies between 4 and 40% in neurotypical brains [2–4], while analyses using single-cell whole-genome sequencing (scWGS) estimated aneuploid neuron frequencies at < 1% [5]. Beyond aneuploidy,

[†]İdil Yet, Thomas Arendt, Mehmet Somel and Uwe Ueberham have contributed equally to this work

*Correspondence: turan.zelihagozde@gmail.com; uwe.ueberham@medizin.uni-leipzig.de

¹ Department of Biological Sciences, Middle East Technical University, 06800 Ankara, Turkey

² Paul Flechsig Institute for Brain Research, Leipzig University, Leipzig, Germany

Full list of author information is available at the end of the article



recent scWGS studies also estimated CNV-carrying neurons at around 30% in young adults and 10% in old adults [6].

Over the last two decades, a number of fluorescence in situ hybridization (FISH) and cytogenetic-based studies investigated CNV frequencies in AD and healthy control brains [2, 7–12]. Several of these reported extra copies of chromosomes in the AD brain [7–12]. This, in turn, implies that the chromosomal imbalance might contribute to AD pathogenesis via altered gene expression levels. An example of such imbalance is seen in individuals with Down's syndrome (DS); carrying an extra copy of chromosome 21 appears to facilitate aggregation of amyloid- β (A β) plaques in the brains of DS individuals similar to the AD phenotype [9, 13, 14].

There are various explanations for why post-mitotic neurons in AD brains could carry high frequencies of somatic CNV [15]. According to one view, the high CNV burden in the AD brain originates from neurogenesis in the embryonic period. This excessive somatic mutation may be pathogenic and manifest itself as increased AD risk during ageing [16]. However, Abascal et al. recently showed that somatic mutation (single nucleotide change or indel) accumulation in cells with mitotic capacity and in post-mitotic neurons follow similar trajectories. That is, mutational processes (possibly also including CNVs) appear to occur in a time-dependent manner rather than being division-dependent [17]. Accordingly, CNVs in AD brains may have accumulated during their lifetime. However, this scenario also appears inconsistent with the observation that CNV-bearing neuron frequencies decrease from young to old adulthood [6]. Another view suggests that AD itself might cause dysregulation in neurons, and AD-affected mature neurons might re-enter the cell cycle, resulting in increased CNV load [8, 18], which may then be eliminated at later stages of AD, thus causing neurodegeneration [10].

Over the last decade, advances in next-generation sequencing (NGS) technologies gave fresh impetus to somatic CNV analyses by allowing variants to be determined at the single-cell level [19]. In one such study, van den Bos and colleagues used scWGS to compare the prevalence of aneuploidy in neurons from healthy control and AD patients [5]. Analyzing 1482 neurons from 10 AD patients and 6 control individuals, the authors reported aneuploid prevalence at 0.7% and 0.6% for control and AD neurons, respectively, and concluded that aneuploid cells are not more common in the AD brain.

These findings by van den Bos and colleagues implied that CNVs might have no relationship to AD pathogenesis, in contrast with earlier finds from FISH and cytometry. However, the study by van den Bos and colleagues had a number of limitations. One was that the authors

only estimated aneuploidy (full chromosome gain or loss), while large CNVs, which could also contribute to pathogenesis, remained uncharacterized. Another limitation was that only one brain region was examined, the frontal cortex, while atrophy of the medial temporal lobe and specifically the hippocampus is generally considered to be a strong predictor of AD [20]. The study did not distinguish among neuron types that may carry sensitivity to AD differentially [21]. Thirdly, the study discarded a large fraction of cells (39%) for showing high within-cell variability in genome coverage, although it was unclear to what extent these represented pure technical error versus cells with complex karyotypes. Fourthly, only NeuN positive neurons were included, which substantially restricts the significance of this study due to different reasons: (1) Recently, up to 30% of cortical neurons have been reported being NeuN-negative following diffuse brain injury, which may be related to certain neurons being particularly vulnerable to membrane disruption [22], a process recently associated with AD [23, 24]. (2) Considerable or even complete loss of NeuN immunoreactivity was also reported for neurons affected by ischemic insults (middle cerebral artery occlusion) without significant cell loss [25] or in neurons that just entered the cell death process [26]. Interestingly, these neuronal populations are of special interest because energy and nutritional deficiency and cell loss are essential characteristics of the AD brain [27]. (3) The intensity of NeuN staining is reported to be lower in AD samples [28], and further (4) due to many NeuN negative cortical neurons in FTLD-TDP (frontotemporal lobar degeneration with TDP-43 inclusions) patients, Yousef et al. suggested NeuN staining as an indicator of healthy neurons [29]. However, if NeuN reflects a neuron's health, any selection of NeuN positive cells would lead to a substantial bias for studying any neurodegenerative disease.

These methodological issues could potentially explain the discrepancies between the findings by van den Bos et al. and those based on FISH and cytogenetic studies [7–12]. Notably, a recent technical comparison between FISH and scWGS using mock aneuploid cells reported a tendency of the latter to severely underestimate aneuploidy [30]. It is thus possible that both neurons with CNV and nuclei thereof display altered physicochemical properties. This may result in selection bias against abnormal nuclei with high CNV loads when using the fluorescence activated cell/nuclei sorting (FACS, FANS) isolation method (exerting mechanical stress [31]) and high hydrodynamic pressure [32], applied by van den Bos and colleagues, and artificially inflate euploidy frequencies. Moreover, besides restriction to NeuN positive cells, usage of only intact nuclei could preclude or

bias AD neurons with nuclear envelope stress or rupture [33].

These observations call for additional data and approaches to tackle this issue. Accordingly, here we generated and analyzed scWGS data to establish the frequency of CNVs (both full chromosome aneuploidies and sub-chromosomal CNVs) in five different brain regions that differ in vulnerability to AD [34]. We employed two different single-cell isolation methods, laser capture microdissection (LCM) and FACS, to isolate neuronal nuclei. LCM, despite being technically challenging, has the advantages of allowing for specific neuron types to be chosen, and being neutral towards normal and abnormal nuclei. We further employed a principal component analysis-based denoising approach to eliminate false positive CNV calls that might result from either systematic experimental biases or repetitive regions in the human genome. Finally, we analyzed published datasets to replicate our main results and check the sensitivity and specificity of our bioinformatics pipeline.

Materials and methods

Tissue sources

Frozen postmortem human brain tissues -temporal cortex, hippocampal subfields cornu ammonis (CA) 1, hippocampal subfields cornu ammonis (CA) 3, cerebellum (CB) and entorhinal cortex (EC)- from a total of 13 AD patients and 7 non-demented age-matched controls were obtained from the GIE NeuroCEB Brain Bank (France) (Additional file 1: Table S1-A). AD cases were diagnosed according to the National Institute of Aging and Reagan Institute Criteria [35] and immunohistochemically processed for tau and amyloid pathologies [36, 37]. Control cases were non-demented individuals who died without known neurological disorders. Post-mortem delays and mean ages of control and AD cases were not significantly different. The average age of death was for control cases ($n=7$) 71.057 years (± 5.13 years SEM) and for AD cases ($n=13$) 70.15 years (± 3.63 years SEM) ($p=0.822$). The average post-mortem delays were 31.14 h (± 7.10 h SEM) for control cases and 26.17 h (± 4.08 h SEM) ($p=0.52$). All experiments were conducted at Paul-Flechsig-Institute (Leipzig University, Germany).

Fluorescence-activated cell sorting (FACS)

Neuronal nuclei were extracted following the protocol described in [38]. Briefly, frozen brain samples were thawed in the hypotonic lysis buffer. Neuronal nuclei were stained with propidium iodide and sorted using BD FACSaria II SORP (BD Biosciences). Genomic DNA

was then isolated and amplified as described below (see scWGS library preparation and sequencing).

Laser capture microdissection (LCM)

Frozen brain samples at $-80\text{ }^{\circ}\text{C}$ were thawed to $-20\text{ }^{\circ}\text{C}$, sliced using CryoCut Freezing Microtome at $30\text{ }\mu\text{m}$ thickness, and mounted on a membrane slide (Carl Zeiss). After staining with cresyl violet, single cells were cut out and placed into an adhesive cap by PALM Micro-Beam (Carl Zeiss). Neurons of the individual 5603 were collected using both FACS ($n=12$) and LCM ($n=64$).

scWGS library preparation and sequencing

Genomic DNA was amplified using WGA4 (GenomePlex[®] Single Cell Whole Genome Amplification Kit) and then purified using the MinElute PCR Purification Kit (Qiagen). The specific adapters were added to the DNA via Phusion[®] PCR followed by purification with the MinElute PCR Purification Kit (Qiagen). Sample quality was evaluated using agarose gel electrophoresis. Sequencing was performed on the HiSeq2500 platform (Illumina) with paired-end 100 bp (PE100) or 150 bp (PE150) modes.

Read quality control and alignment

The *FastQC* tool (version 0.11.9) was used to check the quality of the raw Illumina reads. The results of *FastQC* were summarized using *MultiQC* (version 1.9) [39]. The mean sequence lengths of the reads (ranging between 101 and 151) were inspected using the output of the *MultiQC* (*general_stats_table*). To avoid biases that would affect the interpretation of the results, all reads were trimmed to a length of 66 (the longest possible length in all reads). Illumina adapter and low-quality bases (the first 35 bp) were removed using *Trimmomatic* [40] with the following parameters: "ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:8:TRUE HEADCROP:35 MINLEN:66 CROP:66". The quality of the trimmed reads was checked again using both *FastQC* and *MultiQC*. Adapter-trimmed paired-end FASTQ files were mapped to the hg19 human reference genome using Burrows-Wheeler Alignment (BWA v.0.7.17) [41] with *aln* and *sampe* options.

Filtering

The output of the BWA aligner in Sequence Alignment/Map (SAM) format was further processed by *SAMtools* v1.10 [42] to obtain high-quality uniquely aligned reads. The applied steps are as follows: (1) keep reads mapped in proper pair and discard reads marked with SAM flag 3852 (using the command "samtools view -f 2 -F 3852 -b file.sam > file.bam"), (2) extract uniquely mapped reads from BAM files ("samtools view -h file.bam | egrep -i

"^@|XT:A:U" | samtools view -Shu - > file.bam2") [43], (3) obtain reads having MAPQ scores 60 ("samtools view -h -q 60 file.bam2 > file.bam3"), (4) sort BAM files ("samtools sort file.bam3 > file.sorted.bam"), (5) filter out PCR duplicates ("samtools rmdup -S file.sorted.bam file_rm.sorted.bam"), (6) index BAM files ("samtools index -b file_rm.sorted.bam"), (7) convert BAM file into BED format using the Bedtools *bamToBed* command (Bedtools v2.27.1) [44].

Coverage

Bedtools v2.27.1 algorithm *genomeCoverageBed* was used to obtain coverage of the bases on each BAM file.

CNV prediction and cell elimination

CNV calling was performed using *Ginkgo* [45]. We had three main reasons for using *Ginkgo* over its most commonly used alternative, *HMMcopy* [46]. First, a recent study [47] performed benchmarking on *Ginkgo* and two other widely used methods *HMMcopy* and *CopyNumber*, and found that *Ginkgo* was the most accurate algorithm for inferring the absolute copy number profiles (although *HMMcopy* was superior in identifying breakpoints and running time). Second, *Ginkgo* provided the advantage of outputting data with normalised coverages per cell, which we could use in our PCA-based denoising method, and further in estimating the genome-wide copy number of each cell, which we used to filter cells for high levels of variability in read depth. Third, our tests on the sensitivity and specificity of *Ginkgo* using trisomy-21 in DS and monosomy-X in males in published data [5, 48] revealed 100% and 94% detection rates across the two published datasets. The command-line version of *Ginkgo* was downloaded from <https://github.com/robertaboutkhalil/ginkgo>. The tool was run under the following settings: (1) variable size of 500 kb bins [43] based on simulations of 76 bp reads aligned with *BWA*, (2) independent segmentation method, (3) *ward* and *euclidean* options for the clustering method and clustering distance metric, respectively. Before the segmentation step, GC correction was performed by *Ginkgo* using the R function *LOWESS* (see [45]). For segmentation, *Ginkgo* uses the CBS algorithm implemented in *DNAcopy* in R [49]. *DNAcopy* runs with the following parameters: $\alpha = 0.0001$, $\text{undo.SD} = 1$, $\text{min.width} = 5$ [50]. We also run *HMMcopy* as described in [43] (using the parameter $e = 0.995$).

The number of reads was divided into the variable size of 500 kb bins that correspond to 5578 genomic windows. Only cells with >50,000 reads were kept in downstream analyses (approximately nine reads per window), resulting in $n = 1337$ cells.

Published datasets

The van den Bos 2016 dataset: Data was downloaded from EBI ArrayExpress with the accession numbers E-MTAB-4184 and E-MTAB-4185 [5]. Only the cells that were reported as having good quality libraries were included in the analysis (AD: 883; control: 586; Down's syndrome: 34). Adapter sequences were trimmed with the following parameters: "ILLUMINACLIP:adapter.fa:2:30:10:8:TRUE MINLEN:51". Single-end reads were aligned to the hg19 human reference genome using *BWA* with *aln* and *samse* options. The remaining steps are the same as those described in sections *Filtering*, except that here we used the SAM flag 3844 (because this dataset was single-end sequenced) and used MAPQ scores 20 (because this dataset did not have enough reads which having the MAPQ 60). Note that due to the missing sample information in the database, the number of cells we analyzed does not match what van den Bos and colleagues reported in their original publication.

The McConnell 2013 dataset: FASTQ files of 110 cells were downloaded from the NCBI SRA database with accession number SRP030642 [48]. Adapter sequence was trimmed with the following parameters: "ILLUMINACLIP:adapter.fa:2:30:10:8:TRUE MINLEN:39". Paired-end reads were aligned to the hg19 human reference genome using *BWA* with *aln* and *sampe* options. The remaining steps are the same as those described in sections *Filtering*.

Statistical modeling of CNV frequencies and index of dispersion (IOD) levels

When modelling CNV frequencies, our null hypothesis was no difference in the frequency of CNVs in the AD brain when compared to healthy controls. The overdispersed and zero-dominated nature of the response variable, i.e. the frequency of CNVs, suggested that the data should be fitted using a zero-inflated negative binomial model. For this reason, we used the "glmmadmb" function (package: glmmADMB) [51] in R 3.6.3 with the following parameters: "zero-inflated=TRUE" and "family=nbinom1". The fixed factors of the model were diagnoses (AD and control), chromosomes (autosomes), sex (male and female), brain regions (temporal cortex, hippocampus CA1, hippocampus CA3, cerebellum, entorhinal cortex), and coverage per cell. The individual effect was added as a random factor. Note that sex could not be used as a fixed factor in the *van den Bos 2016* dataset because cells that remained after filtering only belonged to females. We also compared the difference between AD and control in terms of CNV frequency using *HMMcopy* estimates. The fixed factors of the model were diagnoses

(AD vs. control) and coverage per cell. The individual effect was added as a random factor.

When modelling the index of dispersion (IOD, the ratio between the variance of read coverage and the mean), we used the same approach as above. Levels of the response variable, IOD, was predicted using diagnoses (AD and control), brain regions (temporal cortex, hippocampus CA1, hippocampus CA3, cerebellum, entorhinal cortex) and coverage as explanatory variables using the “glm-madmb” function (package: “glmMADMB”) [51] in R 3.6.3. Individual effects were added as a random factor. The distribution of the IOD was right-skewed and the model was run with the “family = gamma” parameter.

To compare the IOD across different brain regions, we used “lme” function (package: “nlme”) in R 3.6.3 with diagnoses as fixed effects and the individual as a random effect.

Copy number statistics

After reads were mapped into the bins, read counts in each bin were divided by the mean read counts across bins for each cell. This value corresponds to the normalized read counts as calculated by *Ginkgo* (see [45]).

A Z_1 -score for each CNV was calculated using the normalized read counts. It was calculated as the cell mean (mean normalized read counts across autosomes) minus the CNV mean (mean read counts between CNV boundaries) divided by the standard deviation (sd) of CNV:

$$Z_1\text{-score} = (\text{mean}_{\text{cell}} - \text{mean}_{\text{CNV}}) / \text{sd}_{\text{CNV}}$$

The Z_2 -score of each CNV was calculated by calculating the difference between the *Ginkgo*-estimated integer copy number state (1 or 3) and the observed normalized read count, dividing by the standard deviation (sd) of the normalized read counts:

$$Z_2\text{-score} = [a - b] / \text{sd}_{\text{CNV}}$$

$$\text{with } a = \text{ESTIMATED_STATE}_{\text{CNV}} \\ b = \text{mean}(\text{OBSERVED_READCOUNT}_{\text{CNV}}).$$

CNVs with two standard deviations below or above the cell’s mean and CNVs with Z_2 -score smaller than or equal to 0.5 were kept in the analysis. Using these combinations, monosomy X ($\geq 90\%$ of the chromosome’s length) was correctly predicted in 58.1% (217 of 373) of males in the uncorrected data.

Principal component analysis (PCA)

To remove experimental noise from the data, the following steps were applied for every cell: (1) one cell (x) at a time was discarded from the analysis. For the remaining cells, PCA was applied on the normalized read counts using the “prcomp” function with the

parameter “scale = TRUE” in R 3.6.3. (2) n PCs that explained at least 90% of the variance in total were chosen. (3) To remove the effect of the chosen PCs from the focal cell x, a linear regression model with normalized read counts from cell x as a response, and the n PCs as explanatory variables was constructed using the R “lm” function. (4) Residuals from this model were calculated. (5) To prevent errors during a lowest fit of GC content (log transformation of negative residuals produces NaNs), we added the constant 1 to the residuals. If there still remained values ≤ 0 , these were replaced with the smallest positive number for the focal cell x. (6) The resulting value was set as a new value of the focal cell x, and *Ginkgo* was run with the new values.

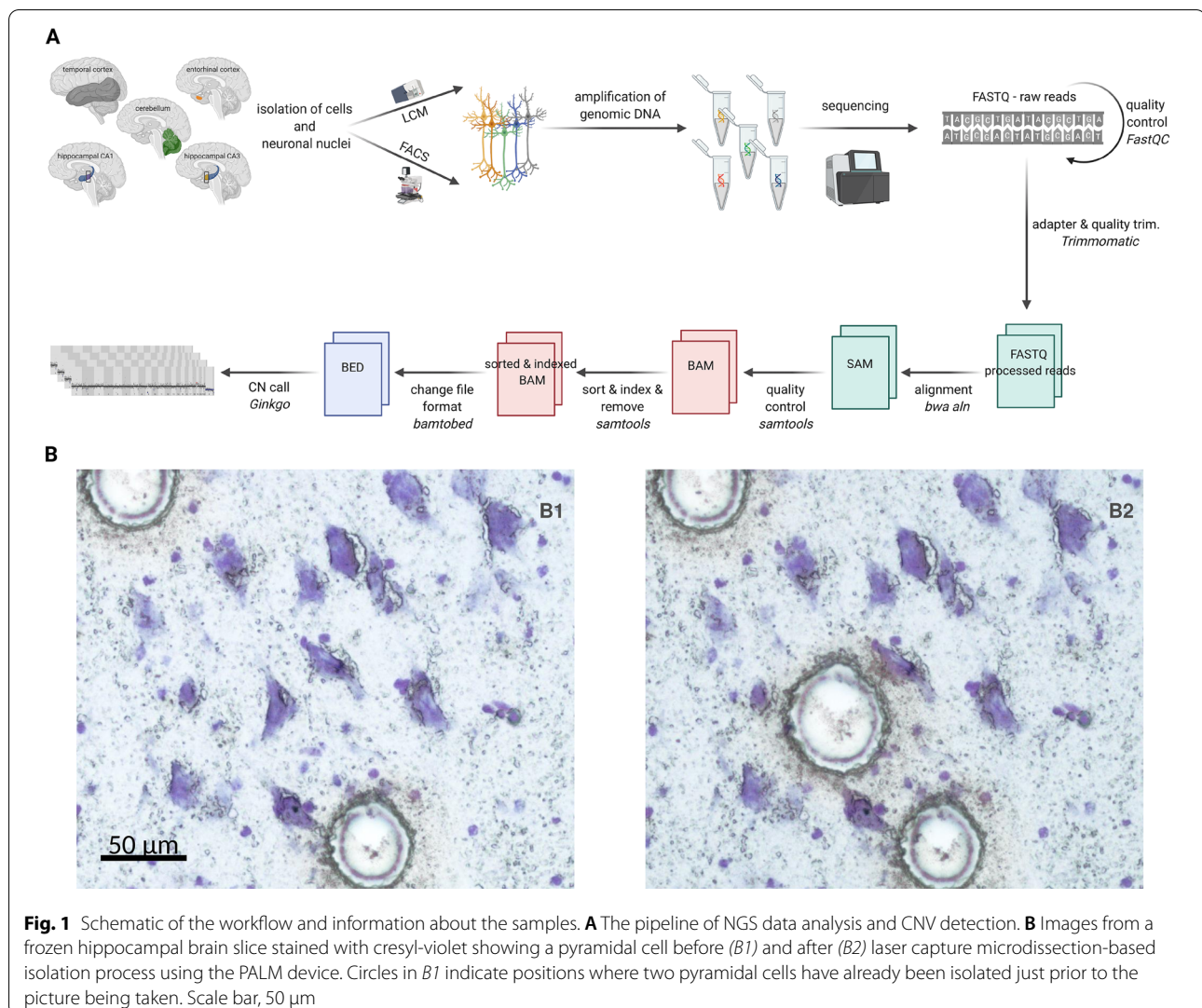
PCA of the normalized read counts across different datasets was performed in R 3.6.3 using the “prcomp” function with the parameter “scale = FALSE”.

Results

Summary of the dataset

We used scWGS to determine the frequency of CNVs in the temporal cortex, hippocampal subfields cornu ammonis (CA) 1, hippocampal subfields cornu ammonis (CA) 3, cerebellum (CB) and entorhinal cortex (EC) of 13 AD patients and 7 age-matched healthy controls (Figs. 1A, 2A,B, Additional file 1: Table S1). The Braak stages of AD patients ranged between III and VI (Fig. 2C). Neuronal nuclei were isolated using either FACS (sorted with propidium iodide, $n = 12$) or LCM (sorted with cresyl violet, $n = 1552$), the latter performed on frozen brain slices (Fig. 1B, see Methods). LCM-isolated non-neuronal “blank” regions were used as negative control ($n = 10$). The LCM method, although more difficult to implement than FACS, was chosen to ensure the selection of nuclei of pyramidal neurons for sequencing, known to be particularly sensitive to AD [21]. For technical comparison, neurons of a single individual were collected both using FACS ($n = 12$) and LCM ($n = 64$) (see Methods). scWGS libraries were prepared using GenomePlex whole-genome amplification and specific adapters were inserted using Phusion® PCR. Paired-end reads were mapped to the human reference genome, followed by stringent filtering to obtain uniquely mapped reads (see Methods). This resulted in a median of 276,446 reads, corresponding to a coverage of 0.006X per LCM-isolated cell (range [133–1,909,016] reads and [0.000003X–0.04X] coverage) (Fig. 3A).

CNVs were predicted using *Ginkgo*, which uses circular binary segmentation (CBS) to estimate deletion or duplication events [45]. Negative controls ($n = 10$) and FACS-isolated neurons ($n = 12$) were analyzed separately and are not included in the main results. *Ginkgo* was run



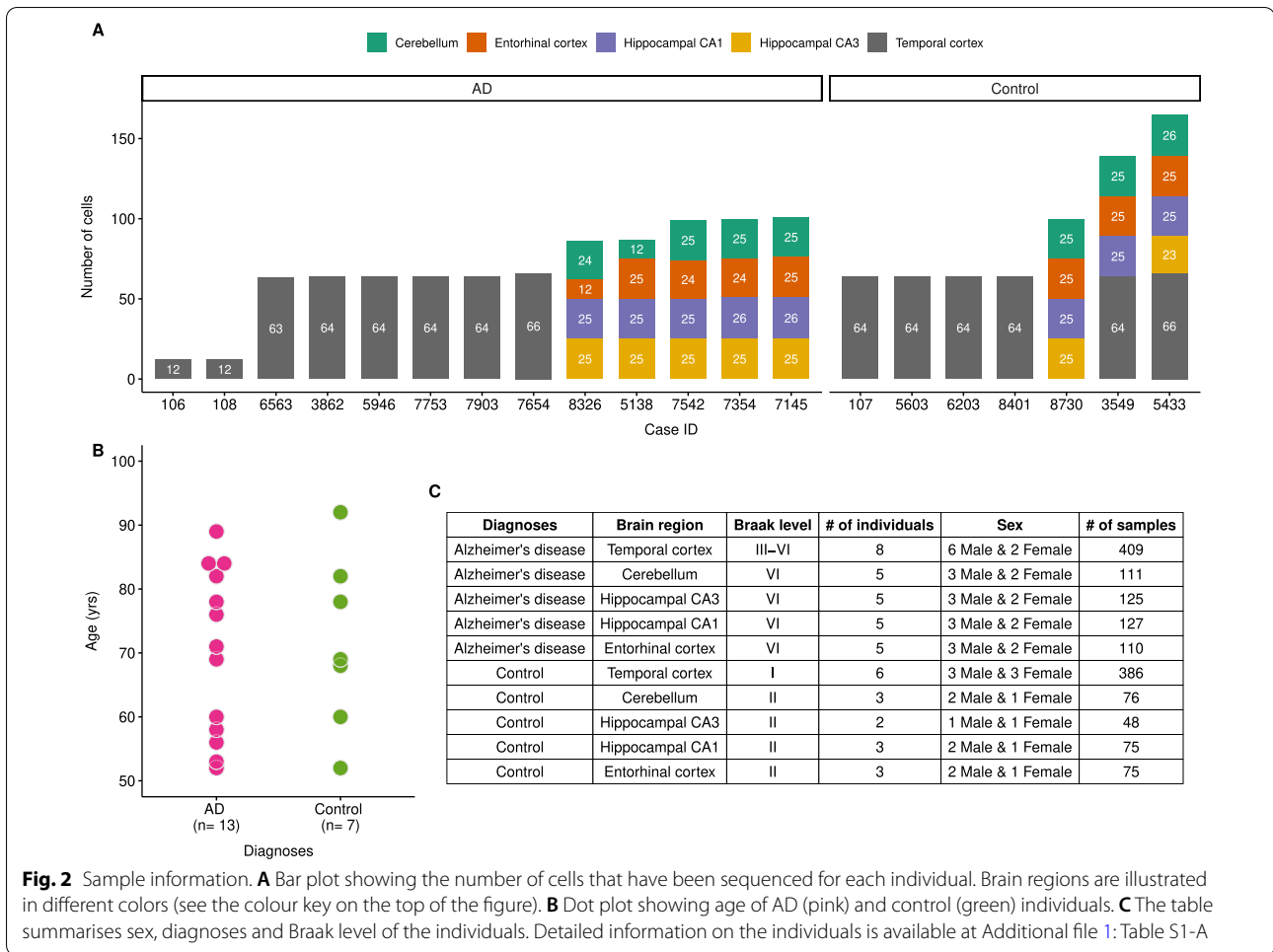
on our dataset with $n=1542$ cells, while in parallel, we also analyzed two published scWGS datasets: one by van den Bos and colleagues (“van den Bos 2016”), comprising $n=1469$ cells from healthy and AD brains (median coverage 0.005X) and another by McConnell and colleagues (“McConnell 2013”), comprising $n=110$ cells from healthy brains (median coverage 0.047X) (Fig. 3D) [5, 48]. Note that the van den Bos 2016 dataset includes only 61% of cells produced in that study, because data from cells filtered for high noise levels were not published and thus could not be included here.

LCM-isolated cells show a high frequency of depth variability

We first evaluated the sensitivity and specificity of our bioinformatics pipeline on scWGS data using trisomy-21

in DS and monosomy-X in males in published data. Analyzing $n=34$ neuronal nuclei from DS individuals [5], trisomy-21 was correctly predicted across all samples without any false positive or false negative calls. In addition, monosomy-X was accurately predicted in 94.2% (338 of 359) of cells from males across the two published datasets [5, 48].

Ginkgo includes an algorithm that uses the distribution of read depth across the genome to infer the average DNA copy number of each cell, which is estimated within a range of 1.5 to 6. It would be expected that the majority of human neurons would carry on average two copies of each autosome, although high frequencies (10–35%) of hyperloid neurons have also been reported, especially in AD brains [10].



Applying *Ginkgo* on the two published datasets, we found that for 99.9% (1577 of 1579) of cells the estimated average copy number lies within [1.9–2]. Using the same algorithm on our dataset, however, only 45% (687 of 1542) of the cells had average copy numbers estimated within the [1.9–2] range; i.e. 55% were non-euploid. Although hyperploid neurons have been described in control brains at ~10% frequency using FISH [10], the observed non-euploidy estimates suggest that our dataset carries particularly high levels of variability in read depth. These differences, in turn, could be related to the LCM protocol used, as the published scWGS experiments had used FACS.

To investigate this possibility, we compared the quality metrics of cells we had collected using FACS or LCM for this study. These metrics were mapping proportion (the number of mapped reads/ the total number of reads), coverage, and index of dispersion (IOD, the ratio between the variance of read coverage and the mean). FACS-isolated cells had higher sequencing coverage and mapping proportions than the LCM-isolated

ones (Wilcoxon two-sided rank-sum test, $p < 0.0001$ and $p < 0.001$ for coverage and mapping proportion, respectively) (Fig. 3A, B). Note that the difference in coverage variability between FACS and LCM has not been reported elsewhere. In addition, FACS-isolated cells had low IOD values, indicating less variation in sequence depth than the rest of the samples (Kruskal–Wallis test, $p = 1.5e-07$) (Fig. 3C). Because our LCM and FACS samples originated from different brain regions with different cell type proportions, we also asked whether such differences could explain the observed LCM vs. FACS differences. To rule out this possibility, we compared the index of dispersion value of the cells that were taken from the temporal cortex of the same individual using FACS ($n = 12$) and using LCM ($n = 64$). We found a significant difference in the direction of higher variability in LCM (Wilcoxon rank-sum test $p < 0.001$), indicating that the observed variability between LCM and FACS can not be simply explained by differences in cell type proportion among brain regions. We note that the higher noise observed in LCM data was not solely due to higher

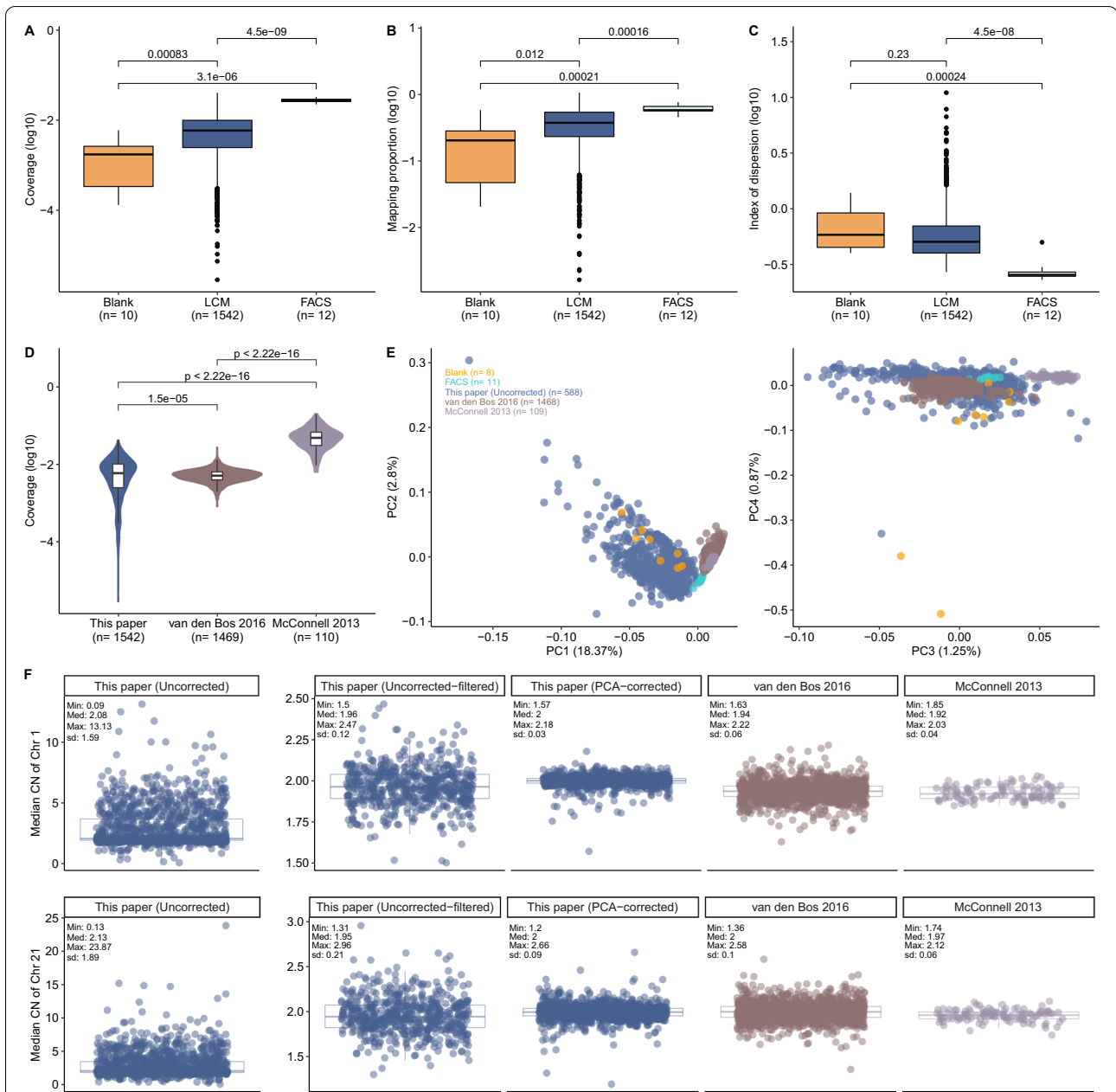
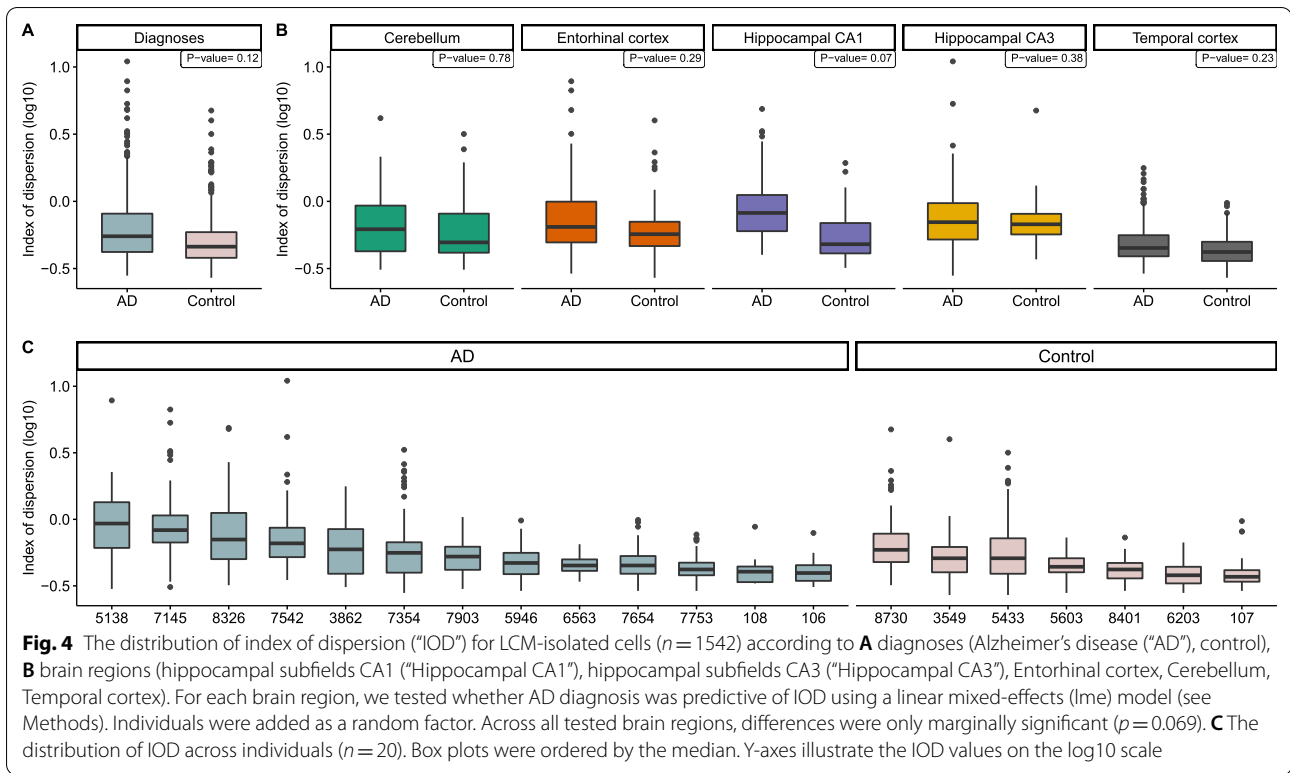


Fig. 3 Comparison between different cell isolation methods and published datasets. Box plots showing the distribution of coverage (A), mapping proportion (B), index of dispersion (C) among FACS-isolated, LCM-isolated and LCM-isolated blank samples. *P*-values were calculated using Kruskal–Wallis test among groups and Wilcoxon rank-sum test between groups. (D) Violin plot showing the distribution of coverage among different datasets. *P*-values were calculated using Wilcoxon rank-sum test between groups. This study, including only LCM: blue; van den Bos 2016: brown; McConnell 2013: purple. (E) A principal components analysis (PCA) was performed using the normalized read counts across autosomal bins ($n = 5243$) in published datasets and this study. Because they dominated the PCs, cells deviating from the [1.9–2] range were not included in the analyses. The number of cells for each dataset are indicated on the plot. X-axes illustrate PC1 and PC3 that explain 18.4% and 1.3% of the total variance, respectively. Y-axes show PC2 and PC4 that explain 2.8% and 0.9% of the total variance, respectively. (F) Boxplots showing the distribution of median CN of chromosome 1 (chr1, upper part of the figure) and chromosome 21 (chr21, lower part of the figure) across bins ($n = 440$ and $n = 68$ for chr1 and chr21, respectively). Each point corresponds to the median CN of each cell. Minimum (“Min”), median (“Med”), maximum (“Max”) and standard deviation (“sd”) of each distribution were shown on the boxplot. Cells that deviated from the [1.9–2] range were excluded from the analyses to be consistent with our filtering criteria (except for the uncorrected datasets). This study [Uncorrected ($n = 1337$), Uncorrected-filtered ($n = 588$), PCA-corrected ($n = 1301$): blue; van den Bos 2016 ($n = 1468$): brown; McConnell 2013 ($n = 109$): purple



genome coverage, as the FACS-based data from the van den Bos 2016 dataset had a median coverage comparable to ours (0.005X vs. 0.006X), but did not show comparable variability as in our LCM data (Fig. 3D). These differences in IOD between LCM and FACS could be potentially explained by the higher sensitivity of the LCM procedure to experimental noise, compared to FACS. Alternatively, they could partly represent abnormal nuclei selected out in FACS but captured by LCM.

We next investigated the possibility that underlying variation may be caused by technical and/or biological factors. For this, we used a generalized linear mixed model (GLMM) to explain IOD (the response variable) per LCM-isolated cell ($n = 1542$) as a function of diagnosis (AD vs. control), genome coverage, and brain region as fixed factors, and individual as a random factor (see Methods). Note that p-values of the pairwise differences between AD and control (Fig. 4A–C) was calculated using a linear mixed-effects model (see Fig. 4 legend). We found that coverage has a significant negative effect on IOD, as may be expected ($z = -21.06$, $p < 0.0001$). Compared to the cerebellum, the region least affected by neurodegenerative diseases [34], we found a significantly high IOD for the entorhinal cortex ($z = 2.61$, $p < 0.05$),

hippocampal CA1 ($z = 3.34$, $p < 0.001$) and hippocampal CA3 ($z = 3.75$, $p < 0.001$), but not for the temporal cortex ($z = -0.28$, $p = 0.78$) (Fig. 4B). Finally, neurons from control individuals have slightly less IOD than AD patients ($z = -1.93$, $p = 0.054$) (Fig. 4A–C). This result might suggest a tendency for neurons of AD patients to carry more variable DNA content and is consistent with cytometry analyses reporting a high occurrence of hyperploid neurons in the AD brain [10]. Although these findings imply a role of biological factors in read count variation within cells, it still remains possible that confounding technical factors influence our data. Given this uncertainty about the source of variability, we continued the analyses by filtering our dataset to remove the most variable cells.

No significant difference in CNV frequency between AD and control in the “uncorrected-filtered” dataset

We then used *Ginkgo* to call CNV events from the “uncorrected-filtered” dataset ($n = 882$ cells from 13 AD patients, and $n = 660$ cells from 7 healthy controls). We found 19,608 events in 882 cells from AD patients (22.2 per cell), and 14,844 events in 660 cells from healthy controls (22.5 per cell). We then tested the observed

frequency difference between AD and control using a GLMM with a negative binomial error distribution (see Methods). The response variable (the frequency of CNVs) was predicted using a combination of fixed factors, including diagnoses, chromosomes, brain regions, sex and coverage (Fig. 6D). The individual effect was added as a random factor. We found no statistically significant difference between AD and control across all tested combinations (GLMM, $p \geq 0.17$; Additional file 3: Table S3).

CNV estimation from low coverage scWGS data is known to be highly sensitive to technical noise, and a large proportion of the called CNV events likely represent false positives. We thus decided to filter both cells and CNV events in our dataset to obtain a more reliable dataset [6, 52, 53]. We started by removing the most highly variable cells among the LCM-isolated ones ($n=1542$) using the following criteria. First, 13% (205 of 1542) of the cells with a low number of reads ($<50,000$) were discarded from the analysis (see Methods). Second, as most cells are expected to be diploid, and also given that the *Ginkgo*-estimated copy number (CN) profiles of 99% of cells in the McConnell 2013 and van den Bos 2016 datasets were observed to lie between [1.9–2], we excluded those cells with CN values beyond this range (54% excluded, 726 of 1337). Third, we filtered out 23 of the remaining 611 cells (4%) that showed extreme CNV intensity, which we defined as three or more chromosomes of a cell carrying predicted CNVs that cover $>70\%$ of their length (Fig. 6A). Information about the remaining cells ($n=588$) is provided in Additional file 2: Table S2 and Additional file 4: Fig. S1.

From these 588 cells, we called 3521 CNVs (~ 5.9 events per cell) in the uncorrected data, which we call the “uncorrected-filtered” dataset. We further applied a number of conservative filtering criteria to remove potential false positives: (1) We only included megabase scale CNVs (≥ 10 Mb), considering that detection of small events with low coverage data will be unreliable. (2) We limited the analyses to 1-somy and 3-somy events, assuming that most somatic CNVs involving chromosomes or chromosome segments would involve loss or duplication of a single copy. (3) We only included CNVs with unique boundaries across all analysed cells, assuming that somatic CNV breakpoint boundaries should be generally randomly distributed across the human genome. (4) We removed CNVs on the proximal portion of the chr19 p-arm, where frequently observed duplications were previously reported as low coverage sequencing artifacts [43]. (5) To ensure the reliability of the CNV signal, we calculated a standard Z -score for each CNV that reflects the deviation in read count distribution in that region compared to the rest of the cell (which we call Z_1 , see Methods), and only accepted CNVs with

absolute values of Z_1 -scores ≥ 2 . (6) We reasoned that read counts in a real CNV should be closely clustered around expected integer values (e.g. 1 or 3). To assess this, we calculated a Z -score for the deviation from the expectation (called Z_2), and only accepted events with absolute values of Z_2 -scores ≤ 0.5 (see Methods, Fig. 6A, Additional file 4: Fig. S3).

After CNV filtering, we found 12 CNV events across 295 cells in 13 AD individuals and 4 CNV events across 293 cells in 7 controls. Among the 295 pyramidal neurons analyzed from the 13 AD patients, we found 10 deletions (3.39% per cell) and 2 duplications (0.68% per cell) (Fig. 6B). These events ranged in size from about 10.14 to 77.01 Mb (median: 19.31 Mb) and were observed in the temporal cortex and the entorhinal cortex. Of the 293 neurons from 7 control brains, 1 deletion (0.34% per cell) and 3 duplications (1.02% per cell) were detected in the temporal cortex with a size range of 10.81 to 54.67 Mb (median: 14.51 Mb) (Fig. 6B). Again testing the CNV frequency differences between AD and control brains using a GLMM, we found no statistically significant effect (GLMM, $p \geq 0.88$) (Additional file 2: Table S2, Additional file 3: Table S3).

We also implemented an alternative algorithm, *HMMcopy* [46], to predict CNVs (see Methods). Overall, 75% (12/16) of the HMMcopy predictions overlapped with the CNV events that we found after filtering the uncorrected Ginkgo predictions. Comparing predicted CNV event frequencies between AD and control we again found no significant difference ($z = -1.34$, $p = 0.18$).

A PCA-based denoising approach minimizes within-cell depth variability

To gain further insight into within-cell variability in our dataset (the uncorrected-filtered version) compared to the two published scWGS datasets, we calculated the median CN of chr1 and chr21 (the largest and smallest chromosomes) across all three. We still found conspicuously higher within-cell variation in our dataset, despite having discarded highly variable cells (Fig. 3F). We then used the autosomal normalized read counts to perform a PCA on the uncorrected-filtered data and published datasets. We also included blank (negative control) samples and FACS-isolated cells to illustrate how reads counts from these two groups relate to others. According to the PCA, LCM-isolated uncorrected-filtered data and blank samples were separated from the published datasets and FACS-isolated cells (Fig. 3E). This result might also highlight distinct profiles of LCM-isolated cells.

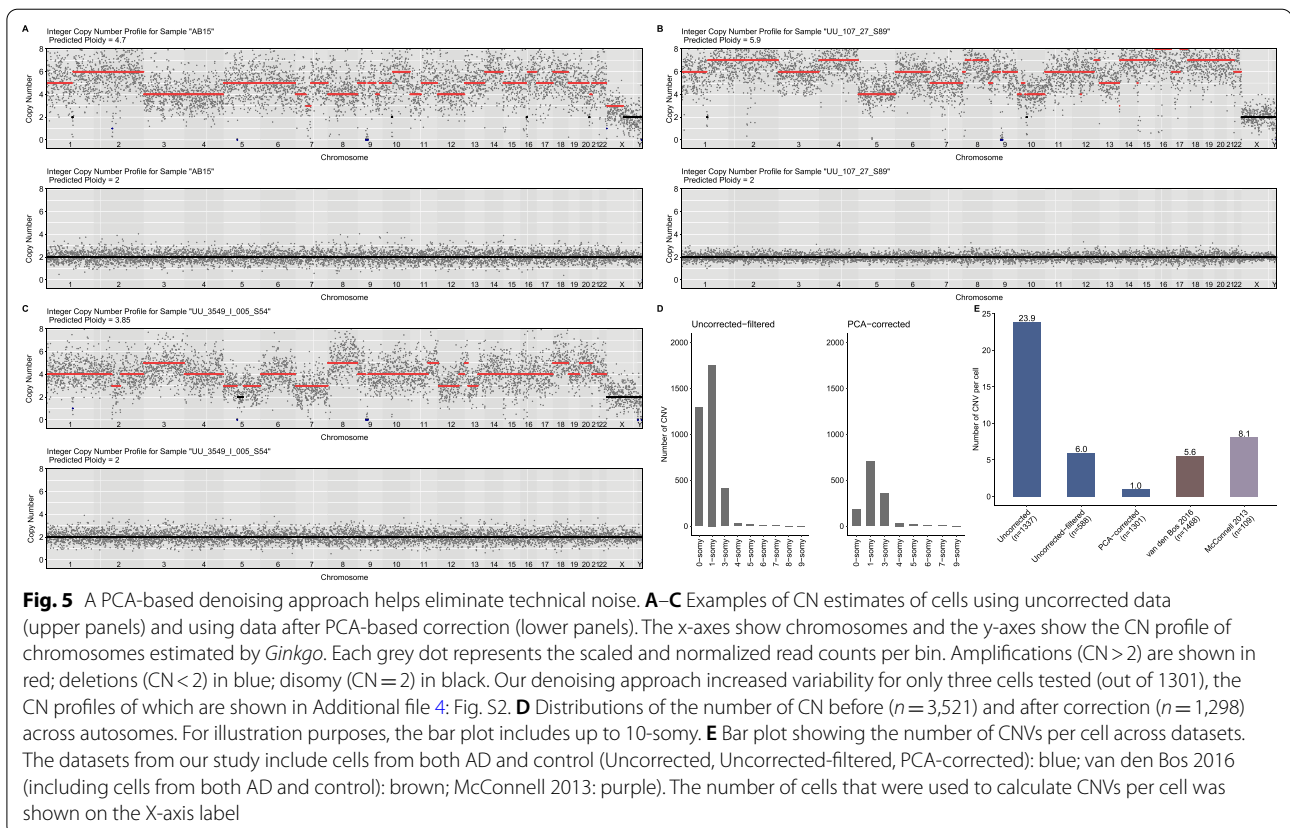
We then sought an approach that could reduce this elevated within-cell variability in read depth, assuming it is of technical origin and possibly related to the LCM procedure. Experimental biases could involve

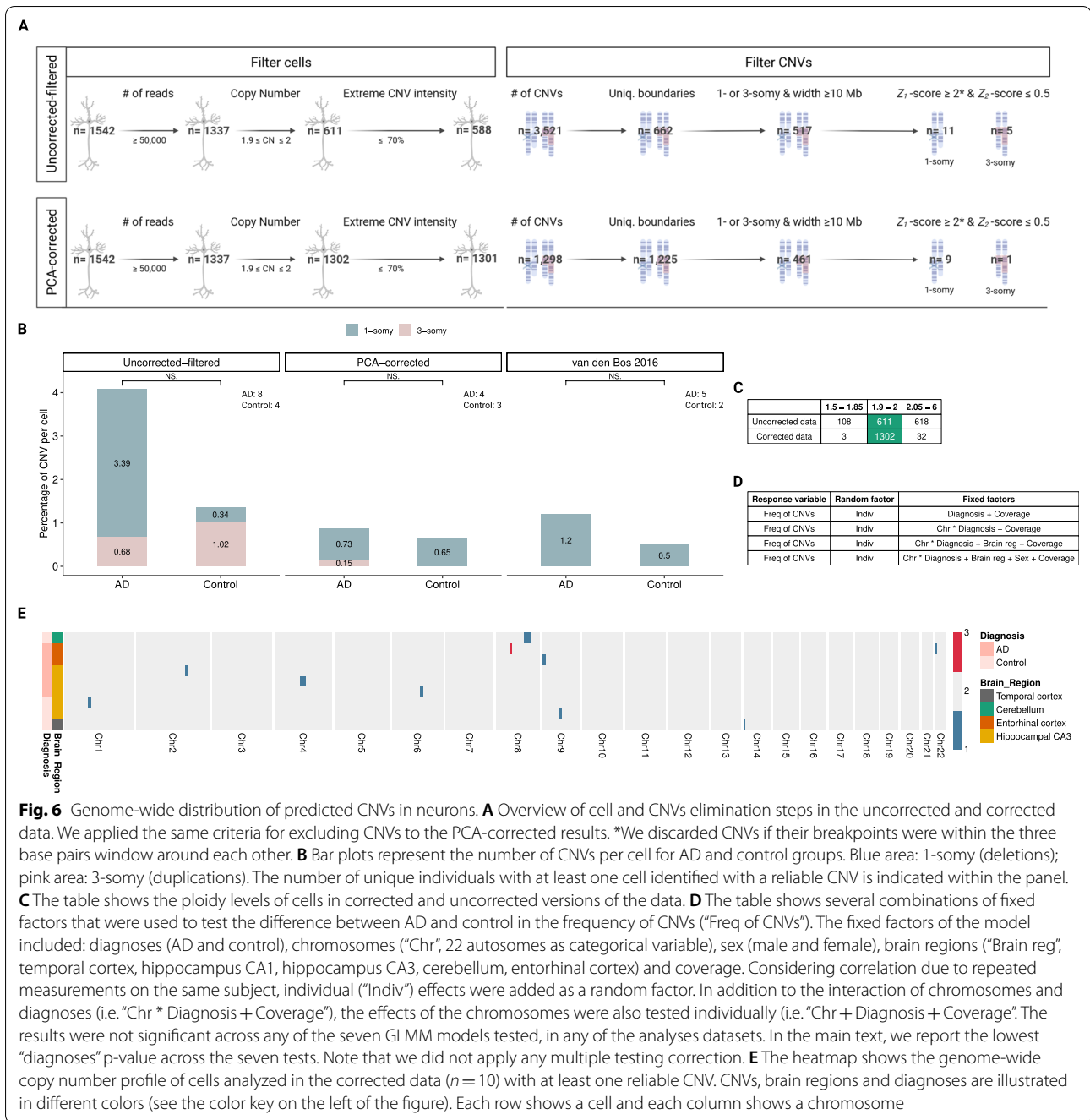
cross-contamination across cells during isolation, or biases that arise during DNA amplification. Although the former should be mainly random, the latter may follow systematic patterns, such as some chromosome segments being more or less prone to be amplified.

We thus devised a procedure for removing putative patterns of systematic read depth variation across cells (see Methods). The algorithm starts by choosing a focal cell x in the dataset, and calculating principal components (PCs) from the normalized read counts per autosome across the rest of the cells (except cell x). It then collects all PCs explaining $\geq 90\%$ of the variance. Treating these as representatives of systematic variation, it removes their values from the normalized read counts of cell x using multiple regression analysis. These steps are performed on all cells individually, creating a “denoised” dataset. The final dataset contains residuals from the multiple regressions instead of the normalized read counts. Notably, this procedure should remove experimentally-induced variation in read depth shared among cells, and also any recurrently occurring somatic CNVs. Rare somatic CNVs, instead, would be mostly unique to each cell and randomly distributed in the genome, and thus would not be affected.

After filtering cells with a low number of reads ($n=205$) and denoising our dataset with this approach,

CN and CNV prediction were performed using *Ginkgo*. We further compared the results between the PCA-corrected and “uncorrected-filtered” datasets. Examples of cells having “noisy” profiles before and after correction are shown in Fig. 5A–C, which suggests a dramatic reduction in within-cell variability. Beyond visual inspection, we also analyzed three statistics. First, we studied the CN profile of cells after PCA correction. We found 97% (1302 of 1337) now lie between 1.9 and 2 (Fig. 6C). This result is comparable to the two published datasets described above and much higher than uncorrected data (45%). Second, we calculated the number of CNV events per cell (sum of the number of CNV/ number of cells) across datasets. In the van den Bos 2016 and McConnell 2013 datasets, we estimated 5.6 and 8.1 CNVs per cell, respectively (Fig. 5E). In our dataset, in the uncorrected version, we found 23.9 CNVs per cell, in the “uncorrected-filtered” data 6.0 CNVs per cell, and in the PCA-corrected data, we estimated on average 1.0 CNV event per cell. The denoising leads to lower CNV estimates in our data, which is more conservative and possibly more realistic than the higher estimates without correction. Third, we estimated the standard deviation in CN among cells for chr1 and chr21. For chr1 and chr21, the standard deviations in the PCA-based data were 4 and 2.3 times lower than in the “uncorrected-filtered” data,





respectively, and comparable to CN standard deviations in the two published datasets.

Subchromosomal CNVs are enriched in deletions in the PCA-corrected data

Based on these three statistics, we decided to study this PCA-corrected version of our dataset. For downstream analysis, we further eliminated cells that deviated from the ploidy range of [1.9–2] (2.6%, 35 of 1337) or showed

extreme CNV intensity (0.08%, 1 of 1302) (Fig. 6A). We thus created a denoised dataset of 1301 pyramidal neurons from 20 individuals.

Estimating CNVs in this dataset using *Ginkgo*, we found 1298 CNVs in total (~1 event per cell). To remove false positives, we also performed the same CNV prediction and downstream analyses on our PCA-corrected data (Additional file 4: Fig. S3). After these steps, we found a total of 9 deletion events (0.7% per cell) and 1 duplication

event (0.08% per cell) across 1301 cells in 20 individuals among all tested brain regions (except for the hippocampal CA1 where no CNV event was found). This excess of deletions is unexpected under the null hypothesis of equal expectation of duplication and deletions (two-sided binomial test $p=0.021$), but consistent with previous observations of more deletions than duplications among somatic mutations [6, 48, 50, 53].

No significant difference between AD and control after PCA-correction or in the van den Bos dataset

Studying CNV frequencies with respect to diagnosis, we found 6 CNV events across 688 cells in 13 AD individuals and 4 CNV events across 613 cells in 7 controls (Fig. 6E). Performing the formal test for the hypothesis of AD versus control differences with this data, we again found no significant difference between the groups (GLMM, $p \geq 0.80$; Additional file 3: Table S3). Information about the CNVs and cells can be found in Additional file 2: Table S2.

We also repeated the same analysis on the van den Bos 2016 dataset, from which originally only aneuploidy was reported. Here we identified 11 CNV events across 883 cells in 10 AD individuals and 3 CNV events across 585 cells in 6 controls. The difference was in the same direction as in our dataset, but again not significant (GLMM, $p \geq 0.79$) (Fig. 6B, Additional file 2: Table S2, Additional file 3: Table S3).

Discussion

Here we discuss technical aspects and the biological outcomes of our work.

The sources of variability among LCM-isolated cells

To the best of our knowledge, this is the first study to use LCM to collect neuronal nuclei for scWGS. Our results showed that LCM-isolated cells showed significantly higher within-cell read depth variation compared to FACS-isolated ones. One random source of high variation could be cross-contamination of LCM-isolated cells during the isolation [54], which in turn might be reflected in the downstream analysis as duplications. In line with this possibility, we found that the number of duplications (≥ 3 -somy) is higher than the number of deletions (0- and 1-somy) in the uncorrected data (deletion to duplication ratio: 0.39). We applied several elimination steps to remove “noisy” cells and to filter nominally false positive CNVs. After these elimination steps, the deletion to duplication ratio increased to 6.46 in the uncorrected-filtered data.

In addition to filtering the uncorrected data, we devised a PCA-based denoising approach to remove systematic

variation across the genome, which could be experimentally-induced, but could also reflect convergent somatic CNVs shared among different individuals. Segments systematically deviating from the genome average have also been described in other neuronal scWGS datasets [6]. Our results showed that PCA-based denoising can strongly reduce within-cell variance in CN among cells. If the noise that was removed is experimentally-induced, then our result means that this noise was partly shared among cells and not entirely random. One source of systematic bias might be genome-wide variation in the propensity to DNA degradation and/or DNA amplification, perhaps due to GC content, chromatin structure or nuclear location of chromosomal segments [50]. Such biases would be shared among cells and effectively removed by PCA.

Beyond technical biases, biological factors could also explain the higher read-depth variability in LCM-isolated than FACS-isolated neurons. Chronister and colleagues recently reported that CNV frequencies in neurons (4%–23.1%) are higher than non-neuronal cells (4.7%–8.7%) [6]. Moreover, cytological studies suggested that AD brains harbour hyperploid neurons more frequently than healthy controls [10]. Consistent with the latter report, we found that neurons from AD patients tend to have higher IOD than control individuals. Also, the cerebellum, which is relatively spared from AD, had lower IOD than the entorhinal cortex and hippocampal areas (but not the temporal cortex). This might be interpreted as a reflection of biological factors on the read-depth variation which is captured efficiently in LCM data. Indeed, if the FACS procedure eliminated cells having abnormal karyotypes, this would result in a cell population with artificially uniform and “clean” ploidy levels. In conclusion, we predict that although random factors (e.g. contamination) and systematic biases most likely contribute to relatively high variation in LCM-collected scWGS data, biological variation may also be a contributor.

PCA-based denoising: advantages and caveats

scWGS is a promising method for predicting CNVs per cell using shallow sequencing. However, as in our study, within-cell variation that may represent false-positive CNVs hinder analyses in low coverage data. Our PCA-based denoising method can be used as a practical solution for in silico cleaning of such data. The approach is based on the idea that somatic CNVs are randomly distributed in the human genome and are particular to each cell. One possible drawback of this approach is that if some neurons from the same individual share the same CNV due to shared developmental ancestry, our method will eliminate such real signals. A more subtle approach could take into account possible clonal relatedness

among cells [55]. Another drawback could arise if certain genomic regions are predisposed to undergo copy number changes; in that case, our method may cause overcorrection. In our dataset, observing an unexpectedly high frequency of CNVs (23.9 events) per cell in the uncorrected version, we chose to remove $\geq 90\%$ of the common variance. After applying PCA-based correction, the CNV rate per cell decreased by 96% (Fig. 5D, E). This, in turn, resulted in a lower number of CNVs per cell in the corrected data, even compared to published datasets. This difference might be attributable to the overcorrection of normalized read counts.

We note that our PCA-based approach could also be used to detect recurrent breakpoints in single-cell cancer genomics. Because clonal cancer cells would also inherit the same CNVs, shared CNV breakpoints identified in PCAs can be used to study clonal evolution.

Limitations

Our study has a number of limitations. First, we only focused on relatively large (≥ 10 Mb) CNVs for sake of sensitivity. However, smaller CNVs may still be much more common and could have contributions to neurodegenerative disease. Future studies on somatic genomic variation in AD might therefore focus on a smaller scale (< 10 Mb) CNVs, for which improvement of experimental protocols and/or the use of higher coverage data appears to be needed [47]. Second, our PCA-based denoising is expected to have removed any CNVs and aneuploidies that are shared among neurons (instead of being cell-specific), due to common origin in the same individual or due to recurrent mutations. Therefore our results only pertain to single cell-specific CNVs. Third, our analysis of published data from van den Bos et al. (2016) could not include a large fraction of cells that they had discarded for showing high depth variability.

Finally, recent work has suggested that CNV-bearing neurons may be eliminated through lifetime in neurotypical individuals [6], and work on hyperploid neurons has also suggested selection against hyperploidy during AD progression [10]. This raises the possibility that dynamic elimination may have obscured a possible signal of AD-control difference in neuronal CNV loads, because our sample size did not allow studying disease stage as a separate factor.

Conclusion

Our main motivation in this study was to describe the relative prevalence of CNVs in the AD brain, where the evidence has been equivocal. Contrary to earlier cytogenetic work [7–12], an scWGS study had reported no difference in neuronal aneuploidy levels in the frontal cortex of AD

patients versus controls [5]. However, the CNV load in different brain regions and relative frequency to the healthy age-matched controls had remained unclear. For example, the entorhinal cortex and hippocampal CA1 have roles in memory formation and learning and are the earliest and most heavily affected regions in AD [56]. On the other hand, hippocampal CA3 is less affected, and neurons in the cerebellum are thought to be relatively spared from neurodegenerative disease [34]. Here we tackled the same question by comparing AD patients and controls using LCM-isolated cells across five different brain regions, either using the raw data ($n = 588$ cells after filtering) or using a denoising approach ($n = 1301$ cells after filtering). To our knowledge, this is the first dataset that includes scWGS data from pyramidal neurons isolated from AD and control brains in multiple brain regions. Although our AD sample contained slightly higher CNV frequencies than the control sample, none of the comparisons was statistically significant. Our analysis of the van den Bos 2016 dataset yielded a qualitatively similar result, also consistent with the original observation of no significant difference in aneuploidy levels in this dataset [5]. Overall, our results call for further research into the possible role of CNVs in AD pathogenesis.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40478-022-01452-2>.

Additional file 1: Table S1. Sample information. A: Demographic and pathology information on each donor.

Additional file 2: Table S2. Information about the cells and significant CNVs. A-B: Uncorrected data. C-D corrected data. E-F: van den Bos

Additional file 3: Table S3. Tables show the GLMM results across different datasets and models. Note that we did not apply any multiple testing correction. A-G: Uncorrected filtered data. H-N: PCA-corrected data. O-Q: van den Bos. R-X: Uncorrected data.

Additional file 4: Fig. S1. Information about the samples of uncorrected and PCA-corrected data. **Figure S2.** CN profiles of three cells deviating from the range of 1.9 and 2 after correction. **Figure S3.** Examples of CNVs that failed to pass and passed the filtering criteria in the uncorrected-filtered data and PCA-corrected data.

Acknowledgements

We thank all colleagues at the METU CompEvo group, in particular, Sinan Can Aan and an anonymous reviewer for helpful suggestions, İsmail Sađlam for discussion, and Kivılcım Bařak Vural for their support and help. Figures 1A and 6A were created with BioRender.com.

Author contributions

MS, İY, UU, TA, CD conceived and supervised the study. SLT and CD prepared brain tissue and performed neuropathological diagnosis. VR, JB and SKW carried out the LCM and library preparation. ZGT performed analyses with support from PP, EY, and UI. ZGT and MS wrote the manuscript. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The work was supported by an ERA.NetRus grant "significans" (coordinator: Thomas Arendt

[FKZ:DLR/01DJ16018], partners: Ivan Iourov, Mehmet Somel, Charles Duyckaerts). The project was funded by a grant from Alzheimer Forschung Initiative (AFI; #17021) to U.U. and TÜBITAK Grant 215Z495. ZGT was supported by a YÖK PhD fellowship. Publication was funded by the Open Access Publishing Fund of Leipzig University, which is supported by the German Research Foundation within the program Open Access Publication Funding. We acknowledge support from Alzheimer Forschung Initiative e.V. for Open Access Publishing.

Availability of data and materials

All data from this study have been submitted to the European Nucleotide Archive (ENA) repository under accession number PRJEB51941. The codes and additional information can be found in the Github repository (https://github.com/zgturan/brain_CN).

Declarations

Ethics approval and consent to participate

The samples were obtained from brains collected in a Brain Donation Program of the Brain Bank NeuroCEB run by a consortium of Patients Associations: ARSEP (association for research on multiple sclerosis), CSC (cerebellar ataxias), LECMA (European league against Alzheimer disease) and France Parkinson. The consents, that have been validated by the Ethical Committee Ile de France 6, were signed by the patients themselves or their next of kin in their name, in accordance with the French Bioethical Laws. The Brain Bank NeuroCEB has been declared at the Ministry of Higher Education and Research and has received approval to distribute samples (agreement AC-2013-1887). The autopsy protocol has been approved by the Biomedicine Agency as requested by the French Law.

Consent for publication

All authors have approved the manuscript and agree with its submission.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biological Sciences, Middle East Technical University, 06800 Ankara, Turkey. ²Paul Flechsig Institute for Brain Research, Leipzig University, Leipzig, Germany. ³Usher Institute, The University of Edinburgh, Edinburgh, UK. ⁴Department of Biology, Lund University, Lund, Sweden. ⁵Laboratoire de Neuropathologie Escourrolle, Hôpital de La Salpêtrière, Paris, France. ⁶Department of Bioinformatics, Hacettepe University, 06100 Ankara, Turkey.

Received: 4 July 2022 Accepted: 27 September 2022

Published online: 30 November 2022

References

- Bishop NA, Lu T, Yankner BA (2010) Neural mechanisms of ageing and cognitive decline. *Nature* 464:529–535
- Pack SD, Weil RJ, Vortmeyer AO, Zeng W, Li J, Okamoto H et al (2005) Individual adult human neurons display aneuploidy: detection by fluorescence in situ hybridization and single neuron PCR. *Cell Cycle United States* 4:1758–1760
- Rehen SK, Yung YC, McCreight MP, Kaushal D, Yang AH, Almeida BSV et al (2005) Constitutional aneuploidy in the normal human brain. *J Neurosci* 25:2176–2180
- Yurov YB, Iourov IY, Monakhov VV, Soloviev IV, Vostrikov VM, Vorsanova SG (2005) The variation of aneuploidy frequency in the developing and adult human brain revealed by an interphase FISH study. *J Histochem Cytochem* 53:385–390
- van den Bos H, Spierings DCJ, Taudt A, Bakker B, Porubský D, Falconer E et al (2016) Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. *Genome Biol* 17:116
- Chronister WD, Burbulis IE, Wierman MB, Wolpert MJ, Haakenson MF, Smith ACB et al (2019) Neurons with complex karyotypes are rare in aged human neocortex. *Cell Rep* 26:825–835.e7
- Yang Y, Geldmacher DS, Herrup K (2001) DNA replication precedes neuronal cell death in Alzheimer's disease. *J Neurosci* 21:2661–2668
- Mosch B, Morawski M, Mittag A, Lenz D, Tarnok A, Arendt T (2007) Aneuploidy and DNA replication in the normal human brain and Alzheimer's disease. *J Neurosci* 27:6859–6867
- Iourov IY, Vorsanova SG, Liehr T, Yurov YB (2009) Aneuploidy in the normal, Alzheimer's disease and ataxia-telangiectasia brain: differential expression and pathological meaning. *Neurobiol Dis* 34:212–220
- Arendt T, Brückner MK, Mosch B, Lösche A (2010) Selective cell death of hyperloid neurons in Alzheimer's disease. *Am J Pathol* 177:15–20
- Yurov YB, Vorsanova SG, Liehr T, Kolotii AD, Iourov IY (2014) X chromosome aneuploidy in the Alzheimer's disease brain. *Mol Cytogenet* 7:20
- Thomas P, Fenech M (2007) Chromosome 17 and 21 aneuploidy in buccal cells is increased with ageing and in Alzheimer's disease. *Mutagenesis England* 23:57–65
- Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L et al (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nat Engl* 349:704–706
- Potter H, Granic A, Caneu J (2016) Role of trisomy 21 mosaicism in sporadic and familial Alzheimer's disease. *Curr Alzheimer Res* 13:7–17
- Arendt T, Stieler J, Ueberham U (2017) Is sporadic Alzheimer's disease a developmental disorder? *J Neurochem Engl* 143:396–408
- Potter H (1991) Review and hypothesis: Alzheimer disease and Down syndrome—chromosome 21 nondisjunction may underlie both disorders. *Am J Hum Genet* 48:1192–1200
- Abascal F, Harvey LMR, Mitchell E, Lawson ARJ, Lensing SV, Ellis P et al (2021) Somatic mutation landscapes at single-molecule resolution. *Nature* 593:405–410
- Herrup K, Arendt T (2002) Re-expression of cell cycle proteins induces neuronal cell death during Alzheimer's disease. *J Alzheimers Dis Netherlands* 4:243–247
- Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A et al (2012) Genome-wide copy number analysis of single cells. *Nat Protoc* 7:1024–1041
- Clerx L, van Rossum IA, Burns L, Knol DL, Scheltens P, Verhey F et al (2013) Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. *Neurobiol Aging* 34:2003–2013
- Fu H, Hardy J, Duff KE (2018) Selective vulnerability in neurodegenerative diseases. *Nat Neurosci* 21:1350–1358
- Hernandez ML, Chatlos T, Gorse KM, Lafrenaye AD (2019) Neuronal membrane disruption occurs late following diffuse brain trauma in rats and involves a subpopulation of NeuN negative cortical neurons. *Front Neurol* 10:1238–1238
- Mrdenovic D, Pieta IS, Nowakowski R, Kutner W, Lipkowski J, Pieta P (2022) Amyloid β interaction with model cell membranes—what are the toxicity-defining properties of amyloid β ? *Int J Biol Macromol* 200:520–531
- Sallaberry CA, Voss BJ, Majewski J, Biernat J, Mandelkow E, Chi EY et al (2021) Tau and membranes: interactions that promote folding and condensation. *Front Cell Dev Biol*. <https://doi.org/10.3389/fcell.2021.725241>
- Unal-Cevik I, Kiliç M, Gürsoy-Ozdemir Y, Gurer G, Dalkara T (2004) Loss of NeuN immunoreactivity after cerebral ischemia does not indicate neuronal cell loss: a cautionary note. *Brain Res Netherlands* 1015:169–174
- Davoli MA, Fourtounis J, Tam J, Xanthoudakis S, Nicholson D, Robertson GS et al (2002) Immunohistochemical and biochemical assessment of caspase-3 activation and DNA fragmentation following transient focal ischemia in the rat. *Neurosci United States* 115:125–136
- Ardanaz CG, Ramírez MJ, Solas M (2022) Brain metabolic alterations in Alzheimer's disease. *Int J Mol Sci*. <https://doi.org/10.3390/ijms23073785>
- Gonzalez-Rodriguez M, Villar-Conde S, Astillero-Lopez V, Villanueva-Anguita P, Ubeda-Banon I, Flores-Cuadrado A et al (2021) Neurodegeneration and astrogliosis in the human CA1 hippocampal subfield are related to hsp90ab1 and bag3 in Alzheimer's disease. *Int J Mol Sci* 23:165
- Yousef A, Robinson JL, Irwin DJ, Byrne MD, Kwong LK, Lee EB et al (2017) Neuron loss and degeneration in the progression of TDP-43 in frontotemporal lobar degeneration. *Acta Neuropathol Commun* 5:68
- Andriani GA, Maggi E, Piqué D, Zimmerman SE, Lee M, Quispe-Tintaya W et al (2019) A direct comparison of interphase FISH versus low-coverage single cell sequencing to detect aneuploidy reveals respective strengths and weaknesses. *Sci Rep* 9:10508

31. Chongtham MC, Todorov H, Wettschereck JE, Gerber S, Winter J. Isolation of nuclei and downstream processing of cell-type-specific nuclei from micro-dissected mouse brain regions—techniques and caveats. *bioRxiv*. 2020;2020.11.18.374223.
32. Chongtham MC, Butto T, Mungikar K, Gerber S, Winter J (2021) INTACT vs. FANS for cell-type-specific nuclei sorting: a comprehensive qualitative and quantitative comparison. *Int J Mol Sci*. <https://doi.org/10.3390/ijms21105335>
33. Kamikawa Y, Saito A, Imaizumi K (2022) Impact of nuclear envelope stress on physiological and pathological processes in central nervous system. *Neurochem Res*. <https://doi.org/10.1007/s11064-022-03608-x>
34. Xu J, Patassini S, Rustogi N, Riba-Garcia I, Hale BD, Phillips AM et al (2019) Regional protein expression in human Alzheimer's brain correlates with disease severity. *Commun Biol* 2:43
35. Consensus Recommendations for the Postmortem Diagnosis of Alzheimer's Disease. *Neurobiol Aging*. 1997;18:S1–2.
36. Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 82:239–259
37. Thal DR, Rüb U, Orantes M, Braak H (2002) Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology United States* 58:1791–1800
38. Matevosian A, Akbarian S. Neuronal nuclei isolation from human post-mortem brain tissue. *J Vis Exp*. 2008. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3233860/>
39. Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 06(32):3047–3048
40. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
41. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
43. Knouse KA, Wu J, Hendricks A (2017) Detection of copy number alterations using single cell sequencing. *J Vis Exp*. <https://doi.org/10.3797/55143>
44. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842
45. Garvin T, Aboukhalil R, Kendall J, Baslan T, Atwal GS, Hicks J et al (2015) Interactive analysis and assessment of single-cell copy-number variations. *Nat Methods* 12:1058–1060
46. Lai D, Ha G. HMMcopy: A package for bias-free copy number estimation and robust CNA detection in tumour samples from WGS HTS data.
47. Mallory XF, Edrisi M, Navin N, Nakhleh L (2020) Assessing the performance of methods for copy number aberration detection from single-cell DNA sequencing data. *PLoS Comput Biol* 16:e1008012
48. McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C et al (2013) Mosaic copy number variation in human neurons. *Science* 342:632–637
49. Seshan V, Olshen A. DNACopy: DNA copy number data analysis. R package version 1.60.0. 2019.
50. Knouse KA, Wu J, Amon A (2016) Assessment of megabase-scale somatic copy number variation using single-cell sequencing. *Genome Res* 26:376–384
51. Fournier DA, Skaug HJ, Ancheta J, Ianelli J, Magnusson A, Maunder M et al (2011) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim Methods Softw* 27:233–249
52. Knouse KA, Wu J, Whittaker CA, Amon A (2014) Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proc Natl Acad Sci USA* 111:13409–13414
53. Cai X, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, Poduri A et al (2014) Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep* 8:1280–1289
54. Kummari E, Guo-Ross SX, Eells JB. Laser capture microdissection—a demonstration of the isolation of individual dopamine neurons and the entire ventral tegmental area. *J Vis Exp*. 2015:e52336.
55. Wang R, Lin D-Y, Jiang Y (2020) SCOPE: a normalization and copy-number estimation method for single-cell DNA sequencing. *Cell Syst* 10:445–52. e6
56. Wang X, Michaelis ML, Michaelis EK (2010) Functional genomics of brain aging and Alzheimer's disease: focus on selective neuronal vulnerability. *Curr Genomics* 11:618–633

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

