

1 DNA isolation protocol effects on nuclear DNA analysis by microarrays, droplet  
2 digital PCR, and whole genome sequencing, and on mitochondrial DNA copy number  
3 estimation

4

5 Elizabeth Nacheva,<sup>1</sup> Katya Mokretar,<sup>1,2</sup> Aynur Soenmez,<sup>2</sup> Alan M Pittman,<sup>3</sup> Colin  
6 Grace,<sup>1</sup> Roberto Valli,<sup>4</sup> Ayesha Ejaz,<sup>2</sup> Selina Vattathil,<sup>5</sup> Emanuela Maserati,<sup>4</sup> Henry  
7 Houlden,<sup>3</sup> Jan-Willem Taanman,<sup>2</sup> Anthony H Schapira,<sup>2</sup> Christos Proukakis.<sup>2</sup> \*

8

- 9 1. Academic Haematology, Royal Free Campus, University College London,  
10 London, UK.
- 11 2. Clinical Neuroscience, Institute of Neurology, University College London,  
12 London, UK.
- 13 3. Molecular Neuroscience, Institute of Neurology, University College London,  
14 London, UK.
- 15 4. Dipartimento di Medicina e Chirurgia, Università dell'Insubria, Varese, Italy.
- 16 5. Department of Epidemiology, The University of Texas MD Anderson Cancer  
17 Center, Houston, Texas, United States of America.

18

19 \* corresponding author

20 e-mail: [c.proukakis@ucl.ac.uk](mailto:c.proukakis@ucl.ac.uk) (CP)

21

## 22 **Abstract**

23 Potential bias introduced during DNA isolation is inadequately explored, although it  
24 could have significant impact on downstream analysis. To investigate this in human  
25 brain, we isolated DNA from cerebellum and frontal cortex using spin columns under  
26 different conditions, and salting-out. We first analysed DNA using array CGH, which  
27 revealed a striking wave pattern suggesting primarily GC-rich cerebellar losses, even  
28 against matched frontal cortex DNA, with a similar pattern on a SNP array. The aCGH  
29 changes varied with the isolation protocol. Droplet digital PCR of two genes also  
30 showed protocol-dependent losses. Whole genome sequencing showed GC-  
31 dependent variation in coverage with spin column isolation from cerebellum. We  
32 also extracted and sequenced DNA from substantia nigra using salting-out and  
33 phenol / chloroform. The mtDNA copy number, assessed by reads mapping to the  
34 mitochondrial genome, was higher in substantia nigra when using phenol /  
35 chloroform. We thus provide evidence for significant method-dependent bias in DNA  
36 isolation from human brain, as reported in rat tissues. This may contribute to array  
37 “waves”, and could affect copy number determination, particularly if mosaicism is  
38 being sought, and sequencing coverage. Variations in isolation protocol may also  
39 affect apparent mtDNA abundance.

40

41

42

43

## 44 **Introduction**

45 Isolation of DNA is possible in several ways, but often little attention is paid to the  
46 protocol, which is not always even reported in detail, with the assumption that the  
47 resulting DNA will be a balanced representation of the original source. Any bias in its  
48 composition could lead to significant downstream effects on copy number  
49 estimation, particularly if mosaicism is being sought, and differential sequencing  
50 coverage. Array-based methods have been used to investigate copy number (CN)  
51 mosaicism although array “waves” are a recognized problem [1–6], and not fully  
52 eliminated bioinformatically [7–10]. Whole genome sequencing (WGS) relative  
53 depth of coverage, now frequently used for CN estimation [11], also varies in a  
54 wave-like pattern [12–14], which is not fully corrected by PCR-free library  
55 construction [15]. Droplet digital PCR (ddPCR) [16] can detect targeted sub-integer  
56 changes expected in mosaicism [17] [18]. Bias in DNA isolation has been reported in  
57 rat tissues, although CNV mosaicism was first considered as an explanation of the  
58 results [12]. To investigate whether DNA isolation bias also occurs in human brain,  
59 we analysed DNA isolated with different protocols (with and without spin columns)  
60 using the above methods. We found a significant effect of the protocol on  
61 downstream results. Care should be given to the selection of DNA isolation method  
62 in all applications, with spin columns requiring particular attention. Furthermore,  
63 mtDNA copy number determination is influenced by the DNA isolation method  
64 chosen [19,20]. We have confirmed this in human substantia nigra, with phenol /  
65 chloroform leading to a higher apparent number. Comparison of mtDNA copy  
66 number would be prone to error unless the exact same conditions were used.  
67

## 68 **Materials and Methods**

### 69 **DNA samples and isolation**

70 Fresh frozen brain material was provided by the Parkinson's UK Tissue bank. Donors  
71 had given informed written consent. Study of brains from the research tissue bank is  
72 approved by the UK National Research Ethics Service (07/MRE09/72). Over the  
73 course of this study, we analysed brain DNA from a total of 11 individuals. This  
74 included six with Parkinson's disease (PD), one with incidental Lewy body disease  
75 (ILBD; PD-like changes found in autopsy in someone who had not been affected by  
76 PD clinically), and four controls. The mean age at death was 79.7 (SD 11.7). Details  
77 are provided in table 1. As not all were used for the same experiments, and some  
78 were used repeatedly, a summary of the isolation method(s) and experiments  
79 performed on each is provided in S1 table.

80

81 **Table 1. Demographic details of individuals whose brains were used.**

Sample ID	gender	age at death	disease duration (years)	Post mortem interval (hours)
PD1	m	63	9	21
PD2	m	69	4	9
PD3	m	73	6	5
PD4	m	68	7	17
PD5	m	78	10	11
PD6	f	83	30	14
ILBD	f	104	-	10

C1	f	78	-	23
C2	m	82	-	48
C3	m	90	-	12
C4	f	89	-	13

82

83

84 DNA isolation protocols used were the following, following manufacturer

85 instructions unless stated.

86 (1) DNeasy® Blood & Tissue spin column (Qiagen), henceforth referred to as SC. We

87 used approximately 25 mg tissue unless otherwise specified. Brain tissue was cut on

88 dry ice, minced and transferred to a 1.5 ml tube. Buffer ATL (180 µl) was added and

89 the samples were homogenized for 1 min with the IKA Eurostar homogenizer. 20 µl

90 of Proteinase K was added to each sample, and digestion was performed at 56 °C, for

91 2 hours, or overnight where stated. When digestions were performed overnight,

92 RNase A (4 µl, 100 mg/ml) was also added the next day.

93 (2) Gentra® Puregene® (Qiagen). This relies on the “salting-out” method, which

94 developed from early work showing that DNA, which carries a negative charge, can

95 be recovered using salt solutions of increasing ionic strength in anion-exchange

96 chromatography [21]. It has been used as a non-toxic alternative to phenol /

97 chloroform. Comparisons with spin columns on bone marrow had shown it to yield

98 more DNA, but any possible biases were not assessed [22]. We used approximately

99 50 mg of brain tissue cut on dry ice, minced and transferred into a 15 ml tube with 3

100 ml Cell Lysis Solution. We performed further steps according to the protocol for 50-

101 100 mg. We included 15 µl Proteinase K overnight incubation at 55°C as

102 recommended for maximal yield, with subsequent treatment with RNase A, the  
103 manufacturer-provided protein precipitation solution, and isopropanol, before 70%  
104 ethanol wash.

105 (3) Phenol Chloroform. 450  $\mu$ L STE buffer and 40  $\mu$ L 20% SDS were added to 25 mg  
106 minced brain sample. After 1 hour incubation at 37°C and vortexing, 20 $\mu$ L Proteinase  
107 K were added. The sample was mixed by hand and incubated at 60°C for 4 h. After  
108 vortexing, another 20 $\mu$ L Proteinase K were added, mixed by hand, and incubated  
109 overnight at 37°C with rotation. The next day samples were centrifuged for 30  
110 minutes and supernatant transferred to clean tubes. 400  $\mu$ L phenol was added and  
111 mixed by hand, followed by 10 minutes on ice, and centrifugation for 2 minutes. The  
112 top layer was transferred to a fresh tube. An additional 400  $\mu$ L phenol was added  
113 followed by 5 minutes on ice and centrifugation for 2 minutes. The top layer was  
114 removed again and 400  $\mu$ L of chloroform/isoamyl alcohol (24:1) added and mixed by  
115 hand. After centrifuging for 2 minutes, the top layer was transferred to a fresh tube  
116 and 2 volumes of cold 95% ethanol and inverted. 4% 3M NaAc was added and the  
117 tubes inverted again and placed in -20°C overnight. The next day tubes were  
118 centrifuged for 30 minutes, the supernatant was discarded, and 500  $\mu$ L of 70%EtOH  
119 was added. After a final 2 minute centrifugation, the supernatant was discarded, and  
120 DNA was air dried and resuspended in 50  $\mu$ L TE.

121 We note that there were minor differences in the proteinase K treatment between  
122 Puregene (following manufacturer guidelines) and Phenol Chloroform, with a slightly  
123 higher initial incubation, and addition of more enzyme with rotation at a lower  
124 temperature overnight. We did not use RNase with Phenol Chloroform. Control

125 peripheral blood lymphocyte (PBL) DNA samples were provided by the UCL Institute  
126 of Neurology Neurogenetics department.

### 127 **Microarray work**

128 We designed a custom 8x60k aCGH array using Agilent e-array software, with ~4,400  
129 probes targeting genes relevant to PD, and their surrounding regions (S2 table).

130 Agilent sex-matched human PBL DNA was used as reference unless indicated  
131 otherwise (cat. no: male 5190-4370, female 5190-3797). The recommended 500 ng  
132 DNA was used in all cases, to avoid any possibility of variable waves due to unequal  
133 DNA amount [7], with hybridisation performed according to manufacturer protocol.

134 Analysis was performed using Agilent Genomic Workbench 7.0. Pre-processing  
135 included GC correction (2 kb window size) and diploid peak centralization. The  
136 recommended ADM2 algorithm was used, with threshold 6 unless otherwise stated,  
137 5 consecutive probes and 10 kb size needed for a call, and “fuzzy zero” (FZ) long  
138 range correction on, unless otherwise specified. All data were mapped to hg19.

139 Isochore graphs were produced by Isosegmenter [23].

140 We also used the Infinium® CytoSNP-850k Beadchip (Illumina), which is designed for  
141 enriched coverage of >3,000 dosage-sensitive genes. Hybridisation was performed  
142 according to the manufacturer protocol, using 200 ng DNA. Preliminary analysis was  
143 done using BlueFuse Multi 4.1, CytoChip module (Illumina). B allele frequency was  
144 estimated by HapLOH [24]. Probe IDs, B allele frequencies, Log R ratios, and AB  
145 genotype calls were extracted from BlueFuse output, and AB genotypes were  
146 converted to plus strand alleles using allele and strand designations provided by  
147 Illumina). We phased the samples using SHAPEIT2, with the Thousand Genomes  
148 Project (1KG) haplotypes as a phased reference panel. Specifically, we used the 1KG

149 Phase 1 haplotypes with singleton sites excluded (files downloaded from IMPUTE2  
150 website). Each sample was phased independently using 1KG haplotypes only  
151 (SHAPEIT2 option no-mcmc). We applied the hapLOH profiling hidden Markov model  
152 using the following parameters: number of event states=1, mean event  
153 length=20Mb, event prevalence=0.001, max iterations=100, hapLOH posterior  
154 probability of imbalance threshold= 0.5.

### 155 **Droplet digital PCR**

156 We performed this on the Bio-Rad QX200 system in 20  $\mu$ L reactions using 40 ng DNA,  
157 ddPCR Supermix, and Biorad-designed commercially available primers (*SNCA*-  
158 *dHsaCP1000476*, *EIF2C1*- *dHsaCP1000484*, *TSC2*- *dHsaCP1000061*, *RPP30*-  
159 *dHsaCP1000485*). All were FAM-labelled, except for *RPP30* which was labelled with  
160 HEX and used as reference. Restriction digestion using HaeIII (NEB) was performed in  
161 tandem with the PCR reaction, by including 2u enzyme in a total of 1 $\mu$ l volume made  
162 up with CutSmart buffer. Where specified, DNA was digested in advance (200 ng  
163 with 5u enzyme in 10  $\mu$ l volume), and 1/5 of this was used per ddPCR reaction.  
164 Reactions were performed in duplicate. After droplet generation, PCR was  
165 performed in the Bio-Rad C1000 Touch Thermal Cycler (95°C for 5 mins, 39 cycles of  
166 95°C for 30 seconds and 60°C 1 min, ending with 98°C for 10 mins). CN was then  
167 assessed using the QX200 Droplet Reader and QuantaSoft software (v.1.4.0.99),  
168 combining the two replicates of each reaction. Statistical analysis was performed  
169 using GraphPad Prism v6.0g, GraphPad Software, CA, USA. For comparison of CN of  
170 DNA isolated with different protocols, we first analysed data for normality by the  
171 D'Agostino & Pearson omnibus, but this could not be demonstrated due to the small  
172 sample size; we therefore compared results using non-parametric tests.

## 173 **Whole genome sequencing (WGS)**

174 We prepared dual indexed, paired-end libraries from 2 µg genomic DNA, using  
175 TruSeq DNA PCR Free chemistry (Illumina) according to standard protocols. The  
176 libraries were sequenced 2x101 bases, in one lane of a Rapid Run flowcell on a HiSeq  
177 2500 (cerebellar DNA), and a single lane of a HiSeq 3000 (substantia nigra). fastq files  
178 were trimmed of Illumina adapters and soft clipped to remove low-quality bases  
179 (Q>10). Picard (1.75) tools (FastqToSam) were used to convert the fastq files to  
180 unaligned BAM files. Reads were aligned to hg19 using Novoalign (v3.02.002),  
181 including base score quality recalibration. The generated .bam files were sorted in  
182 co-ordinate order using Picard tools and fed into GATK for local realignment around  
183 indels. Genome coverage metrics were generated by CollectGcBiasMetrics in Picard,  
184 and coverage using CalculateHsMetrics. To calculate chromosome-specific coverage,  
185 the chromosome 18 or 19 sequence was used as bait. To estimate the number of  
186 mtDNA molecules, we repeated the above steps using the revised mitochondrial  
187 genome reference sequence (NC\_012920). We then divided the coverage of mtDNA  
188 by the coverage of the nuclear genome, and further divided by 2 to correct for the  
189 diploid nuclear genome.

190

## 191 **Results and discussion**

192 We initially analysed DNA isolated from cerebellum and frontal cortex (FC) by spin-  
193 columns (SC) on aCGH. We noted a consistent wave pattern, more prominent in the  
194 cerebellum, even though the cerebellar hybridisations had lower derivative log ratio  
195 spread [dLRs] values (S1 fig), and hybridization of the male to female reference DNA

196 used showed no waves (Fig 1A, using chromosome 1 as an example). Several  
197 aberrations were called in each sample using the standard threshold of 6 (S1 data;  
198 mean 10.7, SD 14.4), of which 1/3 had >10 probes underlying them. Raising the  
199 threshold progressively eliminated these; there were 5.6 at threshold 7 (SD 8.0, data  
200 S2), 2.7 at threshold 8 (SD 2.8; data S3), 1.7 at threshold 9 (SD 1.6, data S4), and 1.06  
201 at threshold 10 (SD 0.9; data S5). From the 17 calls across all samples at this  
202 threshold, 14 were gains at a highly polymorphic 14q32.33 locus. The remaining 3  
203 were a 2 Mb deletion, and two apparent gains, partly overlapping with known CNVs  
204 (fig S2). We did not seek to verify these gains.

205

206 Turning the “fuzzy zero” (FZ) long-range noise correction off, which enhances  
207 mosaicism detection [2], and is recommended for this purpose by Agilent in the  
208 latest Cytogenomics package, led to more extensive calls at threshold 6, following  
209 the “waves”, with apparent losses in GC-rich regions and some gains in GC-poor  
210 regions, many of which persisted even after raising the threshold to 12. These often  
211 followed the genome GC-content isochores [25] (Fig 1B). There was a clear contrast  
212 between chromosome 19, which has the highest gene and CpG island density [26],  
213 and displayed negative waves with prominent losses affecting almost its entire  
214 length, and the similarly-sized chromosome 18, with the lowest gene density and  
215 one of the lowest CpG densities, which showed a mixed picture, with waves in either  
216 direction (S2A fig). Chromosome 19 can be problematic on both aCGH [27] and single  
217 neuron whole genome amplification [28]. A loss of almost the whole chr19 had  
218 indeed been called in one sample by ADM2 with FZ on, but only at threshold 6. To  
219 further investigate the apparent excess of subtle losses in cerebellum, we also

220 hybridised cerebellar DNA with FC of the same brain as reference from 3 PD brains,  
221 including a dye-flip in one. The wave pattern was still generally present, with several  
222 apparent cerebellar relative losses, and reversed by dye flip (fig 1C; S3C –S4 fig).

223

224 **Fig 1. Chromosome 1 in aCGH.**

225 The 10 Mb moving average and the aberration calls by ADM2 (after raising threshold  
226 to 12, with FZ off) are plotted for each sample. Losses are green, gains are red.

227 (A) Brain DNA hybridised against PBL reference DNA. Cerebellar samples are  
228 orange, and FC green. The moving average of a male to female DNA reference  
229 hybridisation is also shown (dark blue).

230 (B) Genome isochores. GC content range for each 100 kb isochore is 30-65%  
231 (blue to orange).

232 (C) Cerebellar DNA hybridised against FC DNA of the same brain for three PD  
233 cases with overnight SC extraction. PD1=purple, PD2=black,  
234 PD4=green. Data for PD2 are derived after combining the dye-flip hybridisation pair.

235 (D) Hybridisations between DNA from the same brain as follows.

236 (1-3) Hybridisations of SC-isolated cerebellar DNA, with Puregene-isolated  
237 DNA from same cerebellum as reference. (1) PD3, 5 mg SC; (2) PD3, 25mg SC; (3)  
238 PD4, 25 mg SC.

239 (4) PD1, Puregene-isolated DNA, cerebellar (test) with FC as reference. Note  
240 the absence of waves and losses. This sample combination, but with spin column  
241 extraction, had led to waves and losses (PD1 in panel C).

242

243

244 To investigate the effect of varying the DNA isolation protocol, we isolated cerebellar  
245 DNA with SC using overnight proteinase K (rather than 2 hours), starting with  
246 approximately 25 or 5 mg tissue in parallel (S3 table), and with the “salting-out”  
247 Puregene kit. We noted that the median DNA yield (ng per mg tissue; S3 table) was  
248 higher with SC when starting with 5 mg (2201) than with 25 mg (544), and even  
249 higher with Puregene (2,784), which was close to the maximum expected (~3,650,  
250 based on 6.6 pg DNA per nucleus, and 85 billion cells in a 154 g cerebellum [29]). We  
251 then performed aCGH of 25 mg overnight SC isolated DNA for two cerebellar  
252 samples, with Puregene-isolated DNA from the same cerebellum as reference; for  
253 one of these, we also hybridised a 5mg SC sample to the Puregene sample (fig 1D; S4  
254 fig). The wave pattern in the 25 mg SC samples (2 and 3 in fig 1D) was similar to the  
255 original hybridization against PBL DNA, although less pronounced, with some  
256 apparent losses called. Waves could therefore be produced even in what was  
257 essentially self-hybridisation, although using only 5mg (sample 1 in fig 1D) minimized  
258 it. Hybridising Puregene-isolated DNA from cerebellum against FC of one brain  
259 (sample 4 in fig 1D and S5 fig) abolished the waves and losses previously seen in the  
260 same pair. Our results suggested a differential bias in cerebellum and FC initially,  
261 with apparent GC-dependent losses, abolished by using a low amount of tissue and  
262 overnight digestion, or Puregene. Using spin columns therefore could lead to  
263 incomplete extraction and introduction of a GC-dependent bias, depending partly on  
264 the tissue amount used. We used overnight proteinase digestion with Puregene,  
265 which should minimize bias, although we cannot exclude the possibility that using a  
266 lower tissue amount, or varying the composition of the solution provided by the  
267 manufacturer, could be of further help.

268

269 To ensure the problem was not limited to our aCGH design, we also analysed freshly  
270 isolated DNA (obtained with the original SC protocol) from four control brains  
271 (cerebellum in all, and FC in three) on a commercially available SNP array. The logR  
272 closely matched the aCGH dLR moving average, with cerebellar losses often called in  
273 similar regions to the aCGH negative waves / possible losses (S6 fig), and losses far  
274 more frequent than gains (115 v 3 on average; S4 table). We next analysed SNP data  
275 using hapLOH [24], which detects regions with significant B-allele frequency (BAF)  
276 deviation, and is valuable in the detection of subtle imbalance expected in  
277 mosaicism [30]. We found no allelic imbalance, suggesting that the apparent losses  
278 affected both chromosomes equally, unlike what one would expect in mosaicism, or  
279 heterozygous CNVs (examples in S7 fig). Based on this, we did not feel that the  
280 CytoSNP losses called were correct, and we only attempted to validate one by PCR  
281 (S7a fig), which was negative (supplementary note), but we cannot exclude the  
282 possibility that some were true.

283

284 To determine if the isolation protocol could also affect copy number determination  
285 by ddPCR, we selected two genes where aCGH suggested negative results (S8 fig);  
286 *EIF2C1*, which is also available by the manufacturer as a HEX-labelled reference  
287 assay, and *TSC2*, which is implicated in the neurocutaneous disorder tuberous  
288 sclerosis, and was within losses in 4/110 frontal neurons in a human single neuron  
289 WGS study [31]. The median CN in the original SC samples was less than 2 for both,  
290 and lower in the cerebellum than FC, although normal in PBL samples (S9 fig). We  
291 compared the results of different protocols on cerebellar DNA (fig 2). The overnight

292 25 mg SC isolations had higher median CN for both *EIF2C1* (1.77 v 1.33) and *TSC2*  
293 (1.64 v 1.31), and the 5 mg SC and Puregene isolation values were even closer to 2  
294 (1.85 and 1.89 for *EIF2C1*; 1.86 and 1.92 for *TSC2*, respectively). There was a highly  
295 significant difference in CN between the three conditions tested for all samples  
296 (Friedman test  $p=0.0017$  for *EIF2C1* and  $0.0046$  for *TSC2*), with a significant pairwise  
297 difference between the original and Puregene CN values after Dunn's multiple  
298 comparison correction ( $p=0.0045$  and  $0.0141$  respectively). Modifying the protocol  
299 slightly by using a separate restriction digestion step did not alter ddPCR results (S5  
300 table). To determine if ddPCR results for genes outside the negative "wave" regions  
301 were influenced by isolation method, we also determined CN for *SNCA*, a gene of  
302 major importance in PD, in two cerebellar samples; they were not altered by the  
303 isolation method (S6 table). These data, taken together with array results, indicated  
304 genuine, protocol-dependent, specific losses during DNA isolation, independent of  
305 downstream experiment type.

306

307 **Fig 2. Effect of DNA isolation on copy number determination by ddPCR for**  
308 **cerebellar samples.**

309 (A) *EIF2C1* and (B) *TSC2*. The medians and interquartile ranges from the original  
310 results, and repeats after overnight SC extractions from 25 mg and 5 mg starting  
311 material, and Puregene, are shown.  $n=6$ , except 5 mg SC, where  $n=4$ .

312

313

314 We next compared low coverage WGS of DNA obtained from the cerebellum with  
315 the lowest post-mortem interval (PD3, 5 hours) by a 25 mg SC overnight isolation

316 and by Puregene. We noted a steep decline of coverage with increasing GC content  
317 in the SC sample when using 100 kb bins, while the Puregene showed a decline only  
318 in the highest GC content (Fig 3). The SC sample showed higher coverage of chr19  
319 compared to chr18, while the Puregene sample had no such bias (ratio 1.4 and 0.97  
320 respectively; S7 table). We then isolated and sequenced DNA from substantia nigra  
321 of individuals in parallel using Puregene and the “gold standard” Phenol Chloroform.  
322 The Puregene samples revealed a similar GC bias. One of three brains showed the  
323 same bias with Phenol Chloroform, while one showed none, even at the highest GC  
324 bins (fig S10). The GC “gradient” seen even in the Puregene-isolated samples  
325 suggests either that we have not been able to fully remove bias, as in rat tissues [12],  
326 or a different GC effect related to the sequencing process, although the Illumina  
327 HiSeq provides the most even human genome coverage [15]. The chr18:chr19  
328 coverage ratio did not show major deviation from 1 with either method (S7 table;  
329 Phenol  $1.03 \pm 0.07$ , Puregene  $1.04 \pm 0.02$ ), therefore any long-range GC-effect in  
330 Puregene and phenol / chloroform may be prominent only in very high GC regions.  
331 Phenol / chloroform may have a slight further advantage compared to Puregene, as  
332 evidenced by the lack of a 100-kb scale GC gradient on coverage in some cases,  
333 although the DNA amount did not allow further experimental comparisons. To  
334 determine if WGS GC bias could lead to erroneous copy number calls, even after  
335 appropriate corrections, we analysed all data using QDNAseq [32] in 100 kb bins (S11  
336 fig). There were possible losses, but with minimally negative logR, in the SC  
337 cerebellar sample, which were absent in Puregene. These would probably be  
338 dismissed as noise, although could potentially be misinterpreted as mosaicism.  
339

340 **Fig 3. Whole genome sequencing coverage in relation to GC content.**

341 The mean normalized coverage per 100 kb window of PD3 cerebellum is shown,  
342 after 25 mg overnight SC isolation and Puregene isolation.

343

344

345 We thus demonstrated in human brain that array “waves”, partial losses in ddPCR,  
346 and GC-dependent WGS coverage variation, can be modulated, and almost  
347 abolished, by variation of the DNA isolation protocol. We have compared the effect  
348 of at least two isolation methods on ddPCR for two genes in six cerebellar samples  
349 (and on aCGH results in three of them), and on GC-dependent coverage variation in  
350 WGS for one of these cerebella, and three substantia nigra samples from different  
351 individuals. We therefore believe that we provide strong evidence for uneven GC-  
352 dependent DNA extraction, which was recently noted in rat tissues [12], but never  
353 before investigated in human tissues to our knowledge, although further studies will  
354 help confirm our conclusions. We have not compared PBL DNA isolation, and solid  
355 tissues may be most prone to bias. We have data from a single human frozen  
356 quadriceps muscle biopsy, from which we isolated DNA with the initial spin column  
357 protocol, which we then analysed on the same aCGH design; a similar wave pattern  
358 was seen (figure S12). We note a very recent study using only multiple-ligation  
359 amplification assay (MLPA) for several fixed human tissues and DNA extraction  
360 methods [33]. The number of probes significantly deviating from normality varied  
361 between tissues and methods. Although the methodology used was very different to  
362 ours, and no information on GC-content of targets was provided, a GC-dependent  
363 extraction bias is possible, as acknowledged by the authors.

364

365 We found that using longer proteinase K treatment or less material on spin columns,  
366 or a non-spin column method, reduced GC-dependent bias. In rats, proteinase K  
367 treatment duration had also affected the outcome, but spin columns had not altered  
368 results from blood, although this was not examined in other tissues [12]. Strong  
369 protein binding to GC-rich DNA regions [12] is a likely mechanism that limits their  
370 extraction, particularly if proteinase K digestion is inadequate, or the spin column is  
371 saturated. The cerebellum may be more prone to extraction bias may because it is  
372 packed with small granule cells, and a greater amount of partly protein-bound DNA  
373 in a given tissue mass could result in reduced and more biased overall yield.

374

375 Determination of the number of mtDNA copies is of interest in several fields,  
376 including PD, where lower mtDNA CN was reported in blood and substantia nigra  
377 [34], but with no details on DNA isolation, and cancer, where batch effects were  
378 corrected bioinformatically, but remained unexplained [35]. Although traditionally  
379 done by qPCR, it is now possible to determine the number of mtDNA molecules in a  
380 preparation by the ratio of sequencing reads mapping to the nuclear versus  
381 mitochondrial genome [36-37]). We therefore determined this for each sample, from  
382 the bulk DNA isolation, without seeking to specifically isolate mtDNA. We then  
383 compared the results obtained by different isolation methods (table 2). For the nigra  
384 samples, phenol led to a higher number than Puregene (average increase 2.51-fold,  
385 SD 0.71). This is consistent with a previous report that organic solvent extraction  
386 results in mtDNA enrichment [20]. As we did not use RNase with Phenol, but we did  
387 as per the standard protocol with Puregene, we cannot comment on any possible

388 effect of this, although the potential higher mtDNA recovery when omitting RNase  
389 may only apply to spin columns [20]. The mtDNA number is similar to a human brain  
390 DNA phenol isolation report [19], although much lower than claimed elsewhere [34].

391

392 **Table 2. Effect of DNA isolation on mtDNA copy number estimated by sequencing.**

Source	Isolation method	mtDNA copy number	Ratio
PD5 SN	Ph:Chl	1380	1.72
	Puregene	802	
PD6 SN	Ph:Chl	2398	2.73
	Puregene	877	
ILBD SN	Ph:Chl	1225	3.08
	Puregene	397	
PD3 CER	SC	508	0.98
	Puregene	518	

393 The ratio of the number estimated for each sample with different isolation methods  
394 is shown. Ph:Chl = Phenol / chloroform.

395

396

397 Our results highlight the often overlooked effects of DNA isolation on copy number  
398 determination, sequencing coverage variation, and mtDNA copy estimation. Array  
399 and sequencing “waves” may be largely due to isolation-induced relative losses.

400 Raising the ADM2 threshold, and keeping the “fuzzy-zero” correction, reduces false  
401 positive calls, although may not eliminate them unless high values are used at the

402 expense of sensitivity. Further studies will be helpful for further validation, and  
403 detailed assessment in other tissues, but we believe that studies should carefully  
404 select and fully report the DNA isolation protocol. For spin columns, the amount of  
405 tissue loaded, and the proteinase digestion duration, might require optimisation,  
406 and avoiding spin columns may sometimes be preferable. Comparing WGS coverage  
407 of chromosomes with different GC content, or performing selective ddPCR, as we  
408 have done, can help exclude major GC bias. When comparing different samples, the  
409 same protocol should be followed. Suspected CN mosaicism should be confirmed by  
410 allelic imbalance, direct visualization by FISH, or breakpoint demonstration. mtDNA  
411 number comparisons should be treated with caution unless the exact same  
412 conditions were used.

413

414

415

#### 416 **Acknowledgements**

417 Tissue samples and associated anonymized data were supplied by the Parkinson's UK  
418 Tissue Bank, funded by Parkinson's UK, a charity registered in England and Wales  
419 (258197) and in Scotland (SC037554). We are grateful to Dr Udo Koehler of MGZ  
420 Medical Genetics Centre for performing the Beadchip hybridization, the UCL  
421 Institute of Neurology sequencing facility, and to all patients and controls who  
422 donated their brains to research.

423

424 **References**

- 425 1. O'Huallachain M, Karczewski KJ, Weissman SM, Urban AE, Snyder MP.  
426 Extensive genetic variation in somatic human tissues. *Proc Natl Acad Sci U*  
427 *S A.* 2012;109: 18018–23. doi:10.1073/pnas.1213736109
- 428 2. Valli R, Marletta C, Pressato B, Montalbano G, Lo Curto F, Pasquali F, et al.  
429 Comparative genomic hybridization on microarray (a-CGH) in  
430 constitutional and acquired mosaicism may detect as low as 8% abnormal  
431 cells. *Mol Cytogenet.* 2011;4: 13. doi:10.1186/1755-8166-4-13
- 432 3. Aghili L, Foo J, DeGregori J, De S. Patterns of somatically acquired  
433 amplifications and deletions in apparently normal tissues of ovarian  
434 cancer patients. *Cell Rep.* 2014;7: 1310–9.  
435 doi:10.1016/j.celrep.2014.03.071
- 436 4. Kasak L, Rull K, Vaas P, Teesalu P, Laan M. Extensive load of somatic CNVs  
437 in the human placenta. *Sci Rep.* 2015;5: 8342. doi:10.1038/srep08342
- 438 5. Lindgren D, Höglund M, Vallon-Christersson J. Genotyping techniques to  
439 address diversity in tumors. *Adv Cancer Res.* 2011;112: 151–82.  
440 doi:10.1016/B978-0-12-387688-1.00006-5
- 441 6. Sakai M, Watanabe Y, Someya T, Araki K, Shibuya M, Niizato K, et al.  
442 Assessment of copy number variations in the brain genome of  
443 schizophrenia patients. *Mol Cytogenet.* 2015;8: 46. doi:10.1186/s13039-  
444 015-0144-5
- 445 7. Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment  
446 of genomic waves in signal intensities from whole-genome SNP genotyping  
447 platforms. *Nucleic Acids Res.* 2008;36: e126. doi:10.1093/nar/gkn556

- 448 8. van de Wiel MA, Brosens R, Eilers PHC, Kumps C, Meijer GA, Menten B, et  
449 al. Smoothing waves in array CGH tumor profiles. *Bioinformatics*. 2009;25:  
450 1099–104. doi:10.1093/bioinformatics/btp132
- 451 9. Leo A, Walker AM, Lebo MS, Hendrickson B, Scholl T, Akmaev VR. A GC-  
452 wave correction algorithm that improves the analytical performance of  
453 aCGH. *J Mol Diagn*. 2012;14: 550–9. doi:10.1016/j.jmoldx.2012.06.002
- 454 10. Marioni JC, Thorne NP, Valsesia A, Fitzgerald T, Redon R, Fiegler H, et al.  
455 Breaking the waves: improved detection of copy number variation from  
456 microarray-based comparative genomic hybridization. *Genome Biol*.  
457 2007;8: R228. doi:10.1186/gb-2007-8-10-r228
- 458 11. Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and  
459 coverage: key considerations in genomic analyses. *Nat Rev Genet*. Nature  
460 Publishing Group, a division of Macmillan Publishers Limited. All Rights  
461 Reserved.; 2014;15: 121–32. doi:10.1038/nrg3642
- 462 12. van Heesch S, Mokry M, Boskova V, Junker W, Mehon R, Toonen P, et al.  
463 Systematic biases in DNA copy number originate from isolation  
464 procedures. *Genome Biol*. 2013;14: R33. doi:10.1186/gb-2013-14-4-r33
- 465 13. Koren A, Handsaker RE, Kamitaki N, Karlić R, Ghosh S, Polak P, et al.  
466 Genetic Variation in Human DNA Replication Timing. *Cell*. Elsevier;  
467 2014;159: 1015–26. doi:10.1016/j.cell.2014.10.025
- 468 14. Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, et al. Cell  
469 Lineage Analysis in Human Brain Using Endogenous Retroelements.  
470 *Neuron*. 2015;85: 49–59. doi:10.1016/j.neuron.2014.12.028
- 471 15. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al.  
472 Characterizing and measuring bias in sequence data. *Genome Biol*.

- 473 2013;14: R51. doi:10.1186/gb-2013-14-5-r51
- 474 16. Huggett JF, Cowen S, Foy CA. Considerations for Digital PCR as an Accurate  
475 Molecular Diagnostic Tool. *Clin Chem*. 2014;  
476 doi:10.1373/clinchem.2014.221366
- 477 17. Kluwe L. Digital PCR for discriminating mosaic deletions and for  
478 determining proportion of tumor cells in specimen. *Eur J Hum Genet*.  
479 Nature Publishing Group; 2016;24: 1644–1648. doi:10.1038/ejhg.2016.56
- 480 18. Miotke L, Lau BT, Rumma RT, Ji HP. High sensitivity detection and  
481 quantitation of DNA copy number and single nucleotide variants with  
482 single color droplet digital PCR. *Anal Chem*. 2014;86: 2618–24.  
483 doi:10.1021/ac403843j
- 484 19. Devall M. A comparison of mitochondrial DNA isolation methods in frozen  
485 post-mortem human brain tissue—applications for studies of  
486 mitochondrial genetics in brain disorders [Internet]. [cited 15 Feb 2016].  
487 Available:  
488 [http://www.biotechniques.com/BiotechniquesJournal/2015/October/A-](http://www.biotechniques.com/BiotechniquesJournal/2015/October/A-comparison-of-mitochondrial-DNA-isolation-methods-in-frozen-post-mortem-human-brain-tissueapplications-for-studies-of-mitochondrial-genetics-in-brain-disorders/biotechniques-360963.html)  
489 [comparison-of-mitochondrial-DNA-isolation-methods-in-frozen-post-](http://www.biotechniques.com/BiotechniquesJournal/2015/October/A-comparison-of-mitochondrial-DNA-isolation-methods-in-frozen-post-mortem-human-brain-tissueapplications-for-studies-of-mitochondrial-genetics-in-brain-disorders/biotechniques-360963.html)  
490 [mortem-human-brain-tissueapplications-for-studies-of-mitochondrial-](http://www.biotechniques.com/BiotechniquesJournal/2015/October/A-comparison-of-mitochondrial-DNA-isolation-methods-in-frozen-post-mortem-human-brain-tissueapplications-for-studies-of-mitochondrial-genetics-in-brain-disorders/biotechniques-360963.html)  
491 [genetics-in-brain-disorders/biotechniques-360963.html](http://www.biotechniques.com/BiotechniquesJournal/2015/October/A-comparison-of-mitochondrial-DNA-isolation-methods-in-frozen-post-mortem-human-brain-tissueapplications-for-studies-of-mitochondrial-genetics-in-brain-disorders/biotechniques-360963.html)
- 492 20. Guo W, Jiang L, Bhasin S, Khan SM, Swerdlow RH. DNA extraction  
493 procedures meaningfully influence qPCR-based mtDNA copy number  
494 determination. *Mitochondrion*. 2009;9: 261–5.  
495 doi:10.1016/j.mito.2009.03.003
- 496 21. Bendich A, Pahl HB, Korngold GC, Rosenkranz HS, Fresco JR. Fractionation  
497 of Deoxyribonucleic Acids on Columns of Anion Exchangers; Methodology

- 498           1. J Am Chem Soc. American Chemical Society; 1958;80: 3949–3956.  
499           doi:10.1021/ja01548a038
- 500   22.   Aplenc R, Orudjev E, Swoyer J, Manke B, Rebbeck T. Differential bone  
501           marrow aspirate DNA yields from commercial extraction kits. *Leukemia*.  
502           2002;16: 1865–6. doi:10.1038/sj.leu.2402681
- 503   23.   Cozzi P, Milanesi L, Bernardi G. Segmenting the Human Genome into  
504           Isochores. *Evol Bioinform Online*. 2015;11: 253–61.  
505           doi:10.4137/EBO.S27693
- 506   24.   Vattathil S, Scheet P. Haplotype-based profiling of subtle allelic imbalance  
507           with SNP arrays. *Genome Res*. 2013;23: 152–8.  
508           doi:10.1101/gr.141374.112
- 509   25.   Costantini M, Clay O, Auletta F, Bernardi G. An isochore map of human  
510           chromosomes. *Genome Res*. 2006;16: 536–41. doi:10.1101/gr.4910606
- 511   26.   Antonarakis SE. Vogel and Motulsky’s Human Genetics: Problems and  
512           Approaches. M.R. Speicher et al, editor. Springer-Verlag Berlin Heidelberg;  
513           2010.
- 514   27.   Jacobs K, Mertzaniidou A, Geens M, Thi Nguyen H, Staessen C, Spits C. Low-  
515           grade chromosomal mosaicism in human somatic and embryonic stem cell  
516           populations. *Nat Commun*. 2014;5: 4227. doi:10.1038/ncomms5227
- 517   28.   Cai X, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, Poduri A, et al.  
518           Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-  
519           Number Variation in the Human Brain. *Cell Rep*. 2014;8: 1280–9.  
520           doi:10.1016/j.celrep.2014.07.043
- 521   29.   Azevedo FA, Carvalho LR, Grinberg LT, Farfel JM, Ferretti RE, Leite RE, et  
522           al. Equal numbers of neuronal and nonneuronal cells make the human

- 523 brain an isometrically scaled-up primate brain. *J Comp Neurol.*  
524 2009/02/20. 2009;513: 532–541. doi:10.1002/cne.21974
- 525 30. Vattathil S, Scheet P. Extensive Hidden Genomic Mosaicism Revealed in  
526 Normal Tissue. *Am J Hum Genet. Elsevier; 2016;98: 571–578.*  
527 doi:10.1016/j.ajhg.2016.02.003
- 528 31. McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron  
529 C, et al. Mosaic copy number variation in human neurons. *Science.*  
530 2013;342: 632–7. doi:10.1126/science.1243472
- 531 32. Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, et  
532 al. DNA copy number analysis of fresh and formalin-fixed specimens by  
533 shallow whole-genome sequencing with identification and exclusion of  
534 problematic regions in the genome assembly. *Genome Res.* 2014;24:  
535 2022–32. doi:10.1101/gr.175141.114
- 536 33. Atanesyan L, Steenkamer MJ, Horstman A, Moelans CB, Schouten JP, Savola  
537 SP. Optimal Fixation Conditions and DNA Extraction Methods for MLPA  
538 Analysis on FFPE Tissue-Derived DNA. *Am J Clin Pathol. Oxford University*  
539 *Press; 2017;14: aqw205. doi:10.1093/ajcp/aqw205*
- 540 34. Pyle A, Anugrha H, Kurzawa-Akanbi M, Yarnall A, Burn D, Hudson G.  
541 Reduced mitochondrial DNA copy-number is a biomarker of Parkinson’s  
542 disease. *Neurobiol Aging. Elsevier; 2015;*  
543 doi:10.1016/j.neurobiolaging.2015.10.033
- 544 35. Reznik E, Miller ML, Şenbabaoglu Y, Riaz N, Sarungbam J, Tickoo SK, et al.  
545 Mitochondrial DNA copy number variation across human cancers. *Elife.*  
546 *eLife Sciences Publications Limited; 2016;5: e10769.*  
547 doi:10.7554/eLife.10769

- 548 36. Ye F, Samuels DC, Clark T, Guo Y. High-throughput sequencing in  
549 mitochondrial DNA research. *Mitochondrion*. 2014;17: 157–63.  
550 doi:10.1016/j.mito.2014.05.004
- 551 37. D’Erchia AM, Atlante A, Gadaleta G, Pavesi G, Chiara M, De Virgilio C, et al.  
552 Tissue-specific mtDNA abundance from exome data and its correlation  
553 with mitochondrial transcription, mass and respiratory activity.  
554 *Mitochondrion*. 2015;20: 13–21. doi:10.1016/j.mito.2014.10.005  
555  
556  
557  
558  
559  
560

## 561 **Supporting information**

562 **S1 fig. Derivative Log Ratio spread (dLRs) values of aCGH brain to PBL reference**

563 **DNA hybridisations.** CER= cerebellum, FC = frontal cortex. The median and

564 interquartile ranges are shown.

565 **S2 fig. CNVs called in only one sample with ADM2 threshold 10, FZ on.** See also

566 data S5. The chromosomal location, individual probe DLR with region of call

567 highlighted, and CNV list in this region are shown for each.

568 **S3 fig. aCGH results for brain DNA of chromosomes 19 (above), and 18 (below).**

569 Analysis by ADM2 (FZ off). The ADM2 threshold is 12 for chr.19, and 6 for chr.18, as

570 most changes were not visible at higher thresholds. 5 Mb moving averages are

571 shown

572 (a) Cerebellum and FC DNA with PBL DNA as reference. Arrows show gains in low

573 GC regions.

574 (b) The human GC content isochore plot (orange=high, blue=low; range 30-65%).

575 (c) PD cerebellar DNA with FC from same brain as reference for PD1, 2, and 4.

576 For PD2, analysis of the combined dye-flip pair is shown. Note that for chr18,

577 in the arrowed low GC region, even the two samples where gains were not

578 called have a slightly positive moving average.

579 **S4 fig. Dye-flipped hybridisations of PD2 cerebellum and FC DNA.**

580 (1) Cerebellum (test) v FC (reference), red. (2) FC (test) v cerebellum (reference), dye

581 flip specified during data import, blue. (3) Male to female reference PBL DNA

582 hybridisation, brown, for comparison. Moving averages are shown over 10 Mb for  
583 chr.1, and 5 Mb for chr.18 and 19.

584 **S5 fig. Chromosome 19 (above) and 18 (below) in aCGH analysis of additional DNA**

585 **extractions with different protocols.** Analysis by ADM2 (FZ off, threshold 12 for  
586 chr.19, 6 for chr.18), with 5 Mb moving averages, and GC isochores; range 30-65%).

587 (1-3): Hybridisations of spin column-extracted cerebellar DNA, with Puregene  
588 extracted DNA from same cerebellum as reference.

589 (1) PD3, 5 mg spin column extraction; (2) PD3, 25mg spin column extraction; (3) PD4,  
590 25 mg spin column extraction.

591 (4) PD1, Puregene DNA, cerebellar, with FC as reference. Note the absence of waves  
592 and losses, unlike the same combination but after SC isolations, shown in fig S3C,  
593 sample 1).

594

595 **S6 fig. Detailed comparison of genome-wide calls by aCGH and SNP array in**

596 **samples C1 (A) and C2 (B).** These samples had the highest number of losses on SNP  
597 array. The five columns for each chromosome are as follows:

598 (1) Deletion calls by SNP array (red dots / bars)

599 (2) Common deletion calls. Yellow lines / bars represent areas called as losses by  
600 SNP array and aCGH (using ADM2, threshold 8, FZ off)

601 (3) aCGH (ADM2, threshold 8, FZ off). Losses are green, gains are red.

602 (4) aCGH (ADM2, threshold 12, FZ off).

603 (5) aCGH (ADM2, threshold 12, FZ off). An additional filter was used to filter calls  
604 that have a level of <15% gain or loss. Note that almost all the losses at these  
605 settings are also called by the SNP array.

606 The rare gains on SNP array are shown as green dots between columns 2 and 3,  
607 and highlighted with a blue arrow.

608 **S7 fig. Examples of chromosome 1 SNP array losses.** The left hand column shows  
609 the relevant data including **with B allele frequency comparison to aCGH from C1**  
610 **cerebellum**. The right hand column shows the SNP logR over this region in selected  
611 other three samples where it was also somewhat negative, although losses were not  
612 always called.

613 A. Loss around *FBXO42* (chr1:16,619,350-16,773,880; 154.5 kb). This was  
614 examined further by PCR, and not confirmed (see supplementary note).

615 B. Loss in 1q22 region (chr1: 155,540,660-155,819,657; 279 kb).

616

617 **S8 fig. aCGH results around ddPCR target genes in all hybridisations using original**  
618 **brain SC isolations.**

619 Probe dLRs in each hybridisation are shown, grouped by type, with ddPCR target  
620 location indicated by a blue line. For PD2 Cerebellum to FC, the combined the dye-  
621 flip data were used.

622 (A) EIF2C1, 325 kb shown (chr1:36199314-3652540)

623 (B) TSC2, 144 kb shown (chr16:2027153-2172009)

624 **S9 fig. Copy number of EIF2C1 and TSC2 determined by ddPCR in cerebellum (CER),**  
625 **frontal cortex (FC), and peripheral blood leucocytes (PBL).** The median and  
626 interquartile ranges are shown in all cases. (a) EIF2C1. CER and FC from four brains,  
627 CER only from another three, and three control PBL DNA samples. Kruskal-Wallis  $p <$   
628 0.001 (b) TSC2. CER and FC from three brains, and cerebellum only from another  
629 four, and four control PBL DNA samples. Kruskal-Wallis  $p=0.0001$

630 **S10 fig. WGS coverage in relation to GC content for the three SN samples after**  
631 **Phenol / Chloroform and Puregene isolation.**  
632 The mean normalized coverage per 100 kb window is shown (y-axis) and the %  
633 content of each window (x-axis). The base quality for each GC content is also shown.  
634

635 **S11 fig. QDNAseq analysis of Cerebellar DNA WGS data with different isolation**  
636 **methods in 100 kb bins. (A) Phenol / Chloroform. (B) Puregene.**  
637 The total PE read number was 97,978,308 for SC, and 62,120,676 for Puregene. The  
638 right-hand figure in A shows the results after downsampling to 62,115,269 reads,  
639 done with Picard DownsampleSam (strategy=high accuracy). The estimated  
640 minimum standard deviation due purely to read counting ( $E\sigma$ ) and the observed  
641 standard deviation ( $\sigma_{\Delta}$ ) are shown. The y-axes show the log ratio (left) and  
642 probability assigned to the aberration called (right). The observed losses in A had a  
643 minimally negative log ratio, and are indicated by arrows for clarity. The same  
644 analysis of the SN WGS data revealed no aberrations (data not shown).

645

646 S12 fig. aCGH analysis of a muscle sample isolated by spin column. The moving  
647 averages are shown for chr1 (over 10 Mb), and 18 and 19 (5 Mb). Aberrations called  
648 with FZ off are highlighted (threshold 12 for chr1 and 19, 6 for chr18).

649  
650 **S1 table. Summary of all samples and experiments. CER= cerebellum. ON=**  
651 **overnight. S/C= spin column. PG= puregene. Ref= used as reference DNA in aCGH.**  
652 **Individual experiments explained in text.**

653

654 **S2 table. aCGH custom design targets and probes.** The fragile site within which  
655 *SNCA* is located and its flanking regions were included (chr4:87-97 Mb), as  
656 constitutional CNVs may involve most of this.

657

658 **S3 table. Effect of DNA extraction protocol on yield and purity of cerebellar DNA.**  
659 For each sample, the exact tissue mass used in spin column extractions (in mg), the  
660 DNA yield (in ng DNA per mg tissue), and the 260/280 ratio are shown. Yield mean  
661 and SD are shown at bottom. ANOVA for yield in four samples where all protocols  
662 were used:  $p=0.039$ .

663

664 **S4 table. Summary of CytoSNP calls in each sample.**

665

666 **S5 table. *TSC2* CN in ddPCR performed with and without restriction digestion as a**  
667 **separate step.**

668 Two control cerebellar samples analysed, with the standard protocol, and with  
669 restriction digestion as a separate step (“pre-digested”), in DNA extracted with spin

670 columns overnight, or Puregene.

671

672 **S6 table. SNCA CN by ddPCR in two control cerebellar samples, using DNA**

673 **extracted with different protocols.**

674

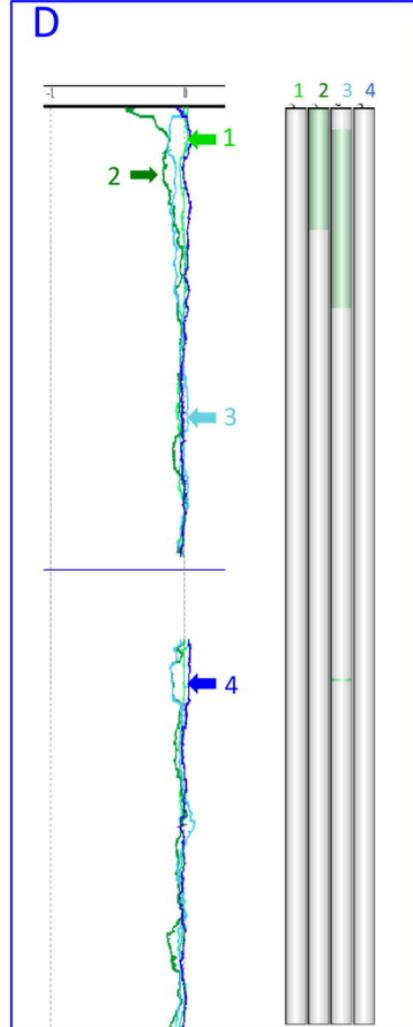
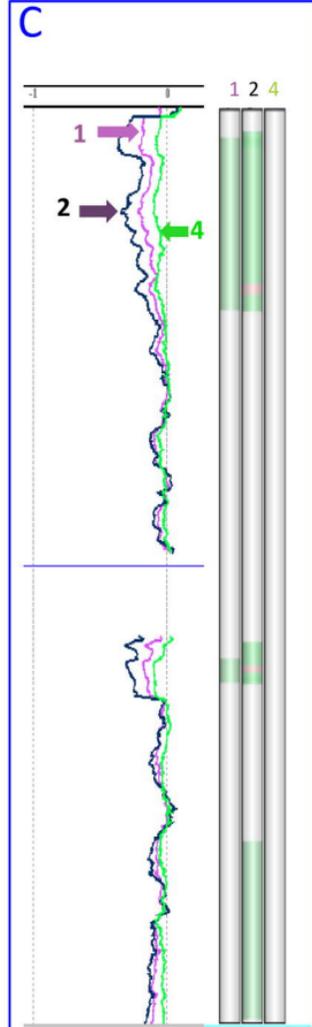
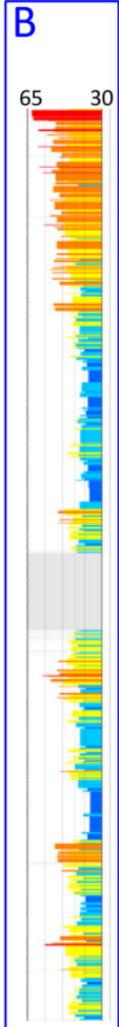
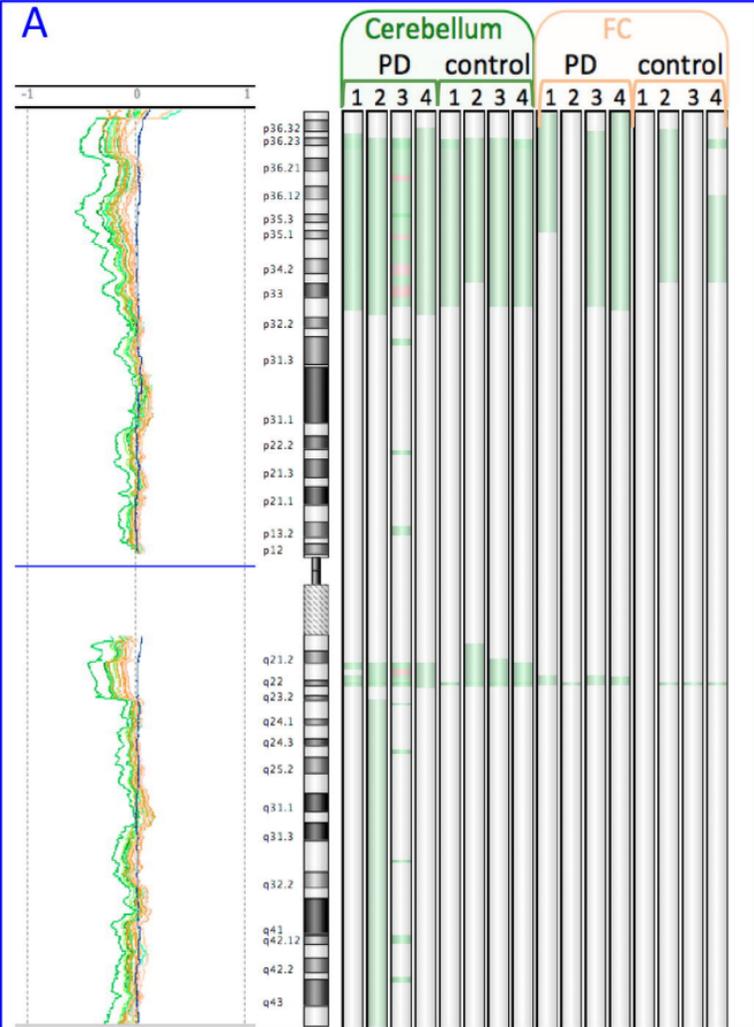
675 **S7 table. WGS summary results of samples isolated with different methods.**

676 **S1-S5 data.** All the brain DNA aberrations reported against PBL reference DNA with

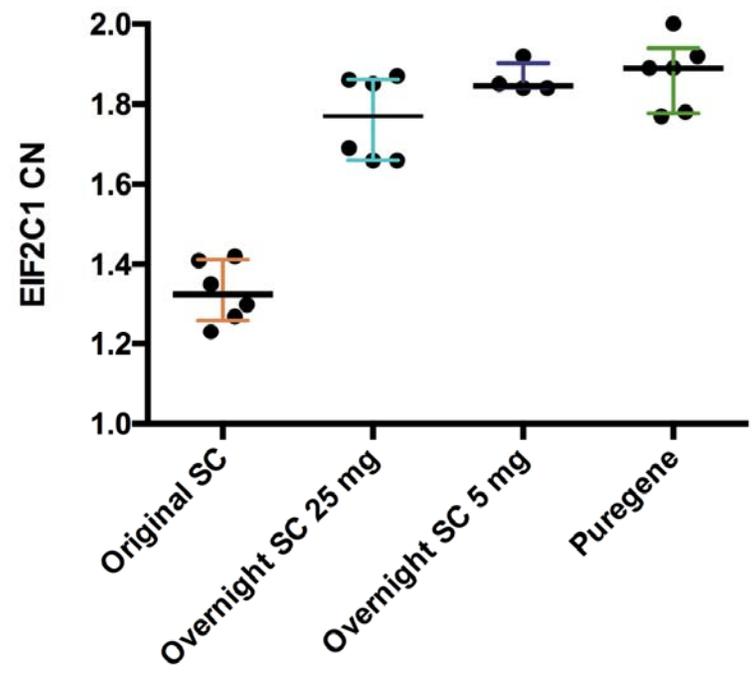
677 ADM2 at varying thresholds, FZ on. Thresholds are as follows: S1- 6, S2- 7, S3- 8, S4-

678 9, S5 -10.

679 **Supplementary note. Attempted PCR validation of a CytoSNP-called deletion.**



A



B

