

## Somatic copy number gains of alpha-synuclein (SNCA) in Parkinson's disease and multiple system atrophy brains

Journal:	<i>Brain</i>
Manuscript ID	BRAIN-2017-02040.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Mokretar, Katya; UCL Insitute of Neurology, Department of Clinical Neuroscience; University College London, Academic Department of Haematology</p> <p>Pease, Daniel; UCL institute of neurology, Department of Clinical Neuroscience</p> <p>Taanman, Jan-Willem; UCL institute of neurology, Department of Clinical Neuroscience</p> <p>Soenmez, Aynur; UCL Insitute of Neurology, Department of Clinical Neuroscience</p> <p>Ejaz, Ayesja; UCL institute of neurology, Department of Clinical Neuroscience</p> <p>Lashley, Tammaryn; University College London Institute of Neurology, Queen Square Brain Bank for Neurodegenerative diseases, Department of Molecular Neuroscience</p> <p>Ling, Helen; University College London, Queen Square Brain Bank for Neurodegenerative diseases, Department of Molecular Neuroscience</p> <p>Gentleman, Stephen; Imperial College London, Department of Medicine</p> <p>Houlden, Henry; Institute of Neurology,</p> <p>Holton, Janice; University College London, Queen Square Brain Bank for Neurodegenerative diseases, Department of Molecular Neuroscience</p> <p>Schapira, Anthony; UCL Institute of Neurology, Department of Clinical Neuroscience</p> <p>Nacheva, Elizabeth; University College London, Academic Department of Haematology</p> <p>Proukakis, Christos; UCL Institute of Neurology, Department of Clinical Neuroscience</p>
Subject category:	Genetics
To search keyword list, use whole or part words followed by an *:	<p>Synucleinopathy &lt; NEURODEGENERATION: CELLULAR AND MOLECULAR, Parkinson's disease &lt; MOVEMENT DISORDERS, Parkinson's disease: cellular mechanisms &lt; NEURODEGENERATION: CELLULAR AND MOLECULAR, Genetics: movement disorders &lt; GENETICS, Copy number variation &lt; GENETICS, Genetics: neurodegeneration &lt; GENETICS, alpha-Synuclein &lt; NEURODEGENERATION: CELLULAR AND MOLECULAR</p>

SCHOLARONE™  
Manuscripts

For Peer Review

## **Somatic copy number gains of alpha-synuclein (*SNCA*) in Parkinson's disease and multiple system atrophy brains**

Katya Mokretar <sup>1,2</sup>, Daniel Pease <sup>1</sup>, Jan-Willem Taanman <sup>1</sup>, Aynur Soenmez <sup>1</sup>, Ayesha Ejaz <sup>1</sup>, Tammarn Lashley <sup>3</sup>, Helen Ling <sup>3</sup>, Steve Gentleman <sup>4</sup>, Henry Houlden <sup>5</sup>, Janice L Holton <sup>3</sup>, Anthony H Schapira <sup>1</sup>, Elizabeth Nacheva <sup>2</sup>, Christos Proukakis <sup>1\*</sup>

### **Affiliations**

- 1 Department of Clinical Neuroscience, UCL Institute of Neurology, University College London, London, UK.
- 2 Department of Academic Haematology, University College London.
- 3 Queen Square Brain Bank for Neurodegenerative diseases, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
- 4 Department of Medicine, Imperial College London, London, UK
- 5 Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

\* Corresponding author [c.proukakis@ucl.ac.uk](mailto:c.proukakis@ucl.ac.uk)

Department of Clinical Neuroscience, UCL Institute of Neurology, Royal Free Campus, Rowland Hill Street, University College London, London, NW3 2PF, UK.

Running title: Alpha-synuclein somatic copy number gains

### **Word count**

Abstract 400  
Main text 6,198

## Abstract

The alpha-synuclein protein, encoded by *SNCA*, has a key role in the pathogenesis of Parkinson's disease and other synucleinopathies. Although usually sporadic, Parkinson's disease can result from inherited copy number variants in *SNCA* and other genes. We have hypothesized a role of somatic *SNCA* mutations, leading to mosaicism, in sporadic synucleinopathies. The evidence for mosaicism in healthy and diseased brain is increasing rapidly, with somatic copy number gains of *APP* reported in Alzheimer's brain. Here we demonstrate somatic *SNCA* copy number gains in synucleinopathies (Parkinson's disease and multiple system atrophy), focusing on substantia nigra.

We selected sporadic cases with relatively young onset or short disease duration, and first excluded high level copy number variant mosaicism by DNA analysis using digital PCR for *SNCA*, and / or customized array comparative genomic hybridisation. To detect low level *SNCA* copy number variant mosaicism, we used fluorescent in-situ hybridisation with oligonucleotide custom-designed probes for *SNCA*, validated on brain and fibroblasts with known copy number variants. We determined *SNCA* copy number in nigral dopaminergic neurons and other cells in frozen nigra sections from 40 Parkinson's disease and 5 multiple system atrophy cases, and 25 controls, in a blinded fashion.

Parkinson's disease cases were significantly more likely than controls to have any *SNCA* gains in dopaminergic neurons ( $p=0.0036$ ), and overall ( $p=0.0052$ ). The average proportion of dopaminergic neurons with gains in each nigra was significantly higher in Parkinson's disease than controls ( $0.78\% \text{ v } 0.45\%$ ;  $p=0.017$ ). There was a negative correlation between the proportion of dopaminergic neurons with gains and onset age in Parkinson's disease ( $p=0.013$ ), but not with disease duration, or age of death in cases or controls. Cases with tremor at onset were less likely to have gains ( $p=0.035$ ). All multiple system atrophy cases had gains, and the highest levels in dopaminergic neurons were in two of these (2.76%, 2.48%). We performed selective validation with different probes after dye-swapping. All three control probes used showed minimal or no gains ( $\leq 0.1\%$  in dopaminergic neurons). We also found occasional *SNCA* gains in frontal neurons of Parkinson's disease cases, and the putamen of one multiple system atrophy case.

We present evidence of somatic *SNCA* gains in brain, commoner in nigral dopaminergic neurons of Parkinson's disease than controls, negatively correlated with onset age, and possibly commonest in some multiple system atrophy cases. Somatic *SNCA* gains may be a risk factor for sporadic synucleinopathies, or a result of the disease process.

**Keywords**

Alpha-synuclein

Parkinson's disease

Multiple system atrophy

Somatic mutation

Mosaicism

**Abbreviations**

Array-CGH array comparative genomic hybridisation

FISH fluorescecent in situ hybridisation

MSA multiple system atrophy

PBS phosphate-buffered saline

## Introduction

Parkinson's disease is usually sporadic, although it can result from inherited mutations, including copy number variants, in *SNCA* and other genes (Mullin and Schapira, 2015; Petrucci *et al.*, 2016). The  $\alpha$ -synuclein protein appears crucial to the pathogenesis of sporadic Parkinson's disease, yet *SNCA* mutations are very rare in DNA from peripheral blood leucocytes. Mosaicism, the presence in an organism of cells with genetic differences due to post-zygotic somatic mutations, is increasingly recognized (Lupski, 2013; Macosko and McCarroll, 2012). Somatic mutations may affect the risk of sporadic diseases (Forsberg *et al.*, 2013), and neurodegeneration (Leijja-Salazar *et al.*, 2018). Inherited (constitutional) copy number variants contribute to variation between individuals (Zarrei *et al.*, 2015), while somatic gains or losses lead to copy number mosaicism. This appears widespread (Abyzov *et al.*, 2012; O'Huallachain *et al.*, 2012), established early in development (Mkrtchyan *et al.*, 2010), and has been repeatedly shown in single frontal neurons from healthy individuals (Cai *et al.*, 2014; Gole *et al.*, 2013; McConnell *et al.*, 2013). Aberrant neuronal DNA synthesis may also arise as a result of neurodegeneration (Frade *et al.*, 2015). Brain mosaicism also exists for point mutations (Hazen *et al.*, 2016; Lodato *et al.*, 2015), aneuploidy (Arendt *et al.*, 2010; Knouse *et al.*, 2014; Rehen *et al.*, 2005), and retrotransposon insertions (Baillie *et al.*, 2011; Evrony *et al.*, 2015; Thomas *et al.*, 2012). In Alzheimer disease somatic *APP* gains occur preferentially in frontal neurons (Bushman *et al.*, 2015), and somatic point mutations have been reported (Beck *et al.*, 2004; Frigerio *et al.*, 2015).

Heritability for Parkinson's disease is less than 35% (Keller *et al.*, 2012; Wirdefeldt *et al.*, 2011), and for MSA under 10% (Federoff *et al.*, 2016). There is therefore a clear need to search for additional aetiological factors. We have hypothesized a role of somatic *SNCA* mutations in sporadic Parkinson's disease (Proukakis *et al.*, 2013). Somatic *SNCA* gains in the neuroectodermal lineage could lead to disease, if over a threshold in a particular region or cell type, or increase the risk synergistically with other factors (inherited risk alleles, epigenetic changes, environmental influences), and could act as initiators or facilitators of pathology spread. Broad spatial dispersal of early developmental mutations could explain Parkinson's disease widespread pathology (Proukakis *et al.*, 2013). Multiple system atrophy (MSA), a predominantly glial synucleinopathy, has significant neuronal pathology (Cykowski *et al.*, 2015), and certain *SNCA* inherited mutations also lead to MSA phenotypes (Fujishiro *et al.*, 2013; Kiely *et al.*, 2013). We found no coding *SNCA* point mutations in

>500 Parkinson's disease / Dementia with Lewy bodies brains in mostly cerebellar DNA, with a 5-10% mutation level detection threshold (Proukakis *et al.*, 2014). Inherited *SNCA* copy number gains (duplications, triplications), while rare, are more frequent than inherited point mutations, with severity and onset age related to gene dosage (Kasten and Klein, 2013). *SNCA* is within a chromosomal fragile site (Rozier *et al.*, 2004), where breaks may arise from DNA replication stress (Arlt *et al.*, 2009). DNA rearrangements have long been suspected in neurogenesis (Chun and Schatz, 1999), and DNA breaks recently demonstrated (Schwer *et al.*, 2016; Wei *et al.*, 2016).

In this work, we present evidence of somatic *SNCA* gains in nigral dopaminergic neurons from synucleinopathies, where they are commoner than in controls, and negatively correlated with the age of onset in Parkinson's disease. We detect some evidence of gains in other cell types and brain regions, and validate our results with different probes. Somatic *SNCA* gains may be a risk factor for sporadic synucleinopathies, or a result of the disease process.

## Methods and materials

### Patients

Material from fresh frozen brain of patients with Parkinson's disease and MSA, and controls, was provided by the Parkinson's UK Tissue Bank and the Queen Square Brain Bank. The latter also provided limited fixed sections. Fixed sections of a brain carrying a heterozygous *SNCA* duplication were provided by the South West Dementia Brain Bank. Donors had given informed written consent. Study of brains from the research tissue banks is approved by the UK National Research Ethics Service (07/MRE09/72), and we had additional approval by the National Research Ethics Service London – Hampstead (10/H0729/21). In total, 41 Parkinson's and 5 MSA patients, and 30 controls, were studied to some extent, but due to limited material available, not all experiments were performed on each. Demographic data, and details of experiments and key results, are in Supplementary Table 1.

Patients were selected for either (1) relatively short disease duration ( $\leq 10$  years), where vulnerable neurons with somatic mutations may not have all died, or (2) relatively young onset ( $\leq 60$  years), as these may have higher mosaicism level / mutation load (Frank, 2010; Proukakis *et al.*, 2013); only eight fulfilled both these criteria. All controls had had alpha-synuclein immunohistochemistry except for one (C13). They had an older age at death, as aged individuals with negative alpha-synuclein immunohistochemistry were considered more suitable, since younger “controls” could have developed pathology later in life. We excluded cases and controls with a first degree relative with Parkinson's disease. All cases had been screened for the *LRRK2* G2019S mutation. One young onset case (PD3; onset 44), who underwent genetic testing in life, had a very rare heterozygous *PARK2* coding variant of uncertain significance (c.1195G>A; p.Glu399Lys; frequency 0.0015% in Europeans in Exome Aggregation Consortium <http://exac.broadinstitute.org>, accessed 6/10/2017). All brain bank reports were available, and review of pathology sections where required was performed by a senior neuropathologist (JLH for Queen Square, and SG for Parkinson's UK). Staging was performed according to the McKeith (McKeith *et al.*, 2005) and Braak (Braak *et al.*, 2003) criteria.



### **DNA analysis by array comparative genomic hybridisation (array-CGH) and droplet digital PCR**

DNA was extracted using Blood and Tissue DNeasy kit (Qiagen), or phenol-chloroform, as described (Nacheva *et al.*, 2017). Array-CGH was performed using our custom 8x60k array (Agilent), densely tiled for relevant genes (*SNCA*, *PARK2*, *GBA*, *PARK7*, *PINK1*, *LRRK2*, *ATP13A2*, *VPS35*, *NURR1*, *POLG*, *PANK2*, *PLA2G6*, *EIF4G1*, *MAPT*) (Nacheva *et al.*, 2017). Analysis was performed with Agilent Genomic Workbench 7.0 using the ADM2 algorithm (threshold 6, 5 consecutive probes and 10 kilobase size needed for a call). The “fuzzy zero” long range correction was used, because removing it may help detect mosaicism (Valli *et al.*, 2011), but also allows artifacts due to GC-related DNA isolation bias to be called (Nacheva *et al.*, 2017). Data were mapped to hg19. We only considered samples with adequate quality hybridisation (derivative log ratio spread <0.3 as recommended by Agilent). Droplet digital PCR was performed on the Bio-Rad QX200 as described (Nacheva *et al.*, 2017). The *SNCA* probe was dHsaCP1000476 (FAM-labelled), and HEX-labelled *RPP30* (dHsaCP1000485) was used as reference. Reactions were performed in  $\geq 2$  replicates. In the case of samples extracted with spin columns, which may be particularly prone to isolation bias (Nacheva *et al.*, 2017), any samples with copy number >2.2 or <1.8 were re-analysed after phenol-chloroform extraction; the latter result is provided. “Artificial mosaics” for *SNCA* copy number were prepared by mixing DNA from triplication fibroblasts, or blood from a duplication patient (Kara *et al.*, 2014), with control DNA.

### **Fluorescent in situ hybridisation (FISH)**

We used labelled oligonucleotide “SureFISH” probes (Agilent Technologies) (Fig. 1A). Probes for *SNCA* were custom-designed, aiming to keep them as small as possible to minimize the chance of adjacent signals merging, which may lead to false negatives (Bushman *et al.*, 2015). Commercially available control probes were chosen, initially on the same chromosome, but far from the fragile site around *SNCA*. A ~50 kb red probe was used initially, and a ~100kb green probe for further validation. Dual-colour (red / green) FISH was performed, with nuclei stained blue using DAPI. Whole Chromosome Paint for chromosome 4 (Kreatech) was used in the same protocol, by replacing the reference probe with whole chromosome paint.

After initial optimisation (Supplementary Note 1), we used 10  $\mu\text{m}$  sections from snap-frozen brain tissue unless specified. Slides were left at room temperature for 30 minutes and then incubated for 30 minutes at 37°C in pepsin (0.05% in 0.01% HCl). To inactivate pepsin,

slides were transferred to 0.01M MgCl<sub>2</sub> in PBS for 5 minutes, and then to a MgCl<sub>2</sub> solution with 3% formaldehyde for 10 minutes. They were then washed twice for 5 minutes in PBS and dehydrated in an ethanol series. After air-drying, they were denatured in 70% formamide solution at 73°C for exactly 2 minutes, and then dehydrated again in an ethanol series. The probes were mixed (3 µl hybridisation buffer, 1 µl sterile water, 1 µl of each of the *SNCA* and control probes) and denatured at 73°C for 5 minutes. 5 µl of the denatured mix was applied to the denatured and air-dried tissue on the slide, covered with a coverslip and incubated for 72 hours at 37°C. After hybridisation, slides were washed in Agilent SureFISH buffers 1 (2 minutes, 73°C) and 2 (1 minute, room temperature), incubated for 20 minutes in 0.1% Sudan Black B (in 70% ethanol) to reduce autofluorescence (Oliveira *et al.*, 2010), washed 3 times for 5 minutes in PBS, and DAPI was applied. Fixed slides first underwent deparaffinisation. They were incubated overnight at 65°C, and then immersed in xylene three times, for 10 minutes, at room temperature. They were then dehydrated twice, for 5 minutes, in absolute ethanol, treated with 0.1M citric acid solution (pH 6) for antigen retrieval using microwave heating for 20 minutes, and then washed twice, for 5 minutes, with dH<sub>2</sub>O.

To prepare fibroblasts for FISH, cells were incubated overnight with colcemid (100 µL per 5 ml of cells, 1,000,000 cells per mL). The next day cells were harvested in 0.075M warm KCl for 15 minutes at room temperature and then fixed in Carnoy's fixative (3:1 methanol: acetic acid). Fresh slides were made and FISH performed using the protocol for frozen slides.

Slides were visualised on a BX61 automated epifluorescence microscope (Olympus, Southend-on-Sea, UK) equipped with an ORCAII-ER CCD camera (Hamamatsu Photonics UK Ltd., Welwyn Garden City, UK) and images captured using SmartCapture 3 software (DSUK Limited, Cambridge, UK). To ensure no bias, all copy numbers were determined on the microscope, by the same experienced observer (KM), blinded to disease status, and demographics (except for certain later experiments which were not blinded, as specified in the results). In substantia nigra, we detected dopaminergic neurons by their neuromelanin content, which is a reliable marker, probably more robust than tyrosine hydroxylase which can be lost in Parkinson's disease (Kordower *et al.*, 2013). It was also used to identify dopaminergic neurons in a recent RNA sequencing study (Nichterwitz *et al.*, 2016). Cells were therefore classified as neuromelanin-positive, corresponding to dopaminergic neurons, and neuromelanin-negative, which likely were mostly glial, although we cannot exclude other neuronal cell types. Cells of each type were counted in a systematic unbiased way (Supplementary fig. 1), aiming for 150 neuromelanin-positive and 300 neuromelanin-negative. Substantia nigra samples were scored as positive or negative for each cell type, if

they had any cells with gains. Additionally, for each section, the copy number of *SNCA* and the reference was individually recorded for each nucleus. The key result was the proportion of nuclei of each type with a gain of *SNCA* and normal reference (and vice versa). We also calculated the numbers of cells with gains of both probes, or losses of any, the mean copy number of *SNCA*, and the total number of signals for each probe. These are discussed in the results section. Similar metrics were recorded for other brain regions analysed, but without neuromelanin scoring.

### **Immunohistochemistry**

To detect *SNCA* copy numbers in frontal cortex, FISH using *SNCA* probe was combined with immunohistochemistry for NeuN, widely used for cortical neuronal identification (Westra *et al.*, 2010). Immunohistochemistry was performed after FISH hybridisation and washing, but before Sudan Black addition. Slides were blocked with goat serum, and stained with the primary antibody (mouse monoclonal anti-NeuN B377; Sigma; 1:100) and secondary antibody (goat anti-mouse AlexaFluor 568, 1:200) for 60 minutes each.

### **Statistics**

Statistical analysis was focused on Parkinson's disease against control, due to the small number of MSA cases, and performed using GraphPad Prism v6, GraphPad Software, CA, USA. We analysed data for normality by the D'Agostino & Pearson omnibus; where this could not be demonstrated, non-parametric statistical tests were used. All p values are 2-sided where applicable. *SNCA* digital PCR results between Parkinson's and controls were compared by unpaired t-test. The relationship between the observed and expected copy number in artificial mosaics was determined by linear regression. Fisher's exact and Chi-square with Yates correction was used for 2x2 tables. Non-parametric ANOVA (Kruskal-Wallis) was used to compare proportion of cells with *SNCA* gains, mean *SNCA* copy number, and mean ratio of red *SNCA* to green reference FISH signals. Pairwise comparisons of Parkinson's and control for each cell type were performed with Dunn's correction for multiple comparisons. Spearman correlation analysis was used to investigate a relationship between the proportion of cells with gains and age of onset, disease duration, and age of death. Mann-Whitney test was used to compare percentage gains and pathology.

## Results

### **DNA analysis by array-CGH and droplet digital PCR excludes high level copy number variant mosaicism of *SNCA* and other relevant genes**

To detect a high level of copy number variant mosaicism, we first analysed brain DNA by array-CGH customized for Parkinson's genes including *SNCA* and *PARK2* (Nacheva *et al.*, 2017), and droplet digital PCR for *SNCA*, which is more suitable than qPCR for copy number determination, including mosaicism (Kluwe, 2016). To estimate the sensitivity of these assays for *SNCA*, we “spiked” DNA samples with known *SNCA* copy number variants into control DNA at variable ratios, to create “artificial mosaics” for gains. For array-CGH, the detection threshold was 30% heterozygous mosaicism (leading to a 15% increase in genomic material, equivalent to copy number 2.3) (Supplementary fig. 2). We analysed DNA from at least one region of 26 Parkinson's disease and 12 control brains, including nigra in most (Supplementary Tables 1, 2). For *SNCA* digital PCR, copy number 2.2 appeared to be the detection threshold (Supplementary fig. 3a). We used it to analyse DNA from the nigra of 36 Parkinson's disease, 2 MSA, and 11 controls (Supplementary fig. 3b; Supplementary Table 1). In total, 38 Parkinson's nigras were analysed with at least one of these methods, and 20 with both. We found no evidence of high level mosaicism.

### **FISH allows direct visualization of *SNCA* copy number**

To detect any lower level gains, we first performed pilot FISH work using bacterial artificial chromosome probes, on unselected cell smears from substantia nigra. These suggested occasional gains, but did not allow definitive assessment (Supplementary Note 2). To improve on this, we used a customised *SNCA* “SureFISH” probe, made of individual labelled oligonucleotides 120 base-pairs each, tiling 50 kilobases of the *SNCA* locus (Fig. 1A), emitting a red signal. These recently developed probes avoid repetitive sequences and other sources of error in bacterial artificial chromosomes, and provide excellent quality signals without noise (Sugita *et al.*, 2017; Takeda *et al.*, 2017). We used the smallest size recommended, as larger probes might lead to nearby signals merging (Bushman *et al.*, 2015). We expected that a gain of *SNCA* could involve most of the ~12 Mb fragile site around it (Fungtammasan *et al.*, 2012), as inherited copy number variants can be several megabases long (Kara *et al.*, 2014; Ross *et al.*, 2008), but not extend beyond it. We therefore used catalogue green-labelled probes, located on chromosome 4 but away from the fragile site, as reference. To verify the *SNCA* probe, we performed FISH on fibroblasts with a heterozygous *SNCA* triplication and a control, and

cingulate gyrus from fixed brain with a heterozygous *SNCA* duplication. Additional *SNCA* signals were detected where appropriate (Fig. 1B). Multiple copies were sometimes hard to distinguish, appearing as larger apparently compound or elongated signals (arrowheads in Fig. 1B).

### **FISH shows low level gains of *SNCA* in nigra, commoner in Parkinson's disease dopaminergic neurons than controls**

We obtained successful hybridisation, with no significant background, in a total of 40 Parkinson's disease, 25 control, and 5 MSA nigras. The age of death, and onset and duration of disease, are shown in Supplementary fig. 4 and Supplementary table 1. We counted on average 142.1 neuromelanin-positive and 305.6 neuromelanin-negative cells per slide (Supplementary fig. 5). We observed a low proportion of nuclei with clear unique *SNCA* gains, with two copies of the reference, in both cell types, in many disease and control nigras, with very rare cells showing more than 3 *SNCA* copies (Fig. 2; Supplementary table 1). Comparing the proportions of brains with any evidence of gains, in each cell type and overall, between Parkinson's disease and controls, showed that Parkinson's disease nigras were far more likely to be positive for neuromelanin-positive cells ( $p=0.0036$ ), and overall ( $p=0.0052$ ), but this did not reach significance for neuromelanin-negative cells alone ( $p=0.1144$ ) (Table 1). We next compared the percentage of nuclei with *SNCA* gains in each nigra and cell type. Although the proportion of nuclei with gains was small (mean 0.78% in Parkinson's neuromelanin-positive cells), there was a significant difference across all groups ( $p=0.032$ ) (Fig. 3). Pairwise comparisons between Parkinson's disease and control for each cell type showed a significantly higher proportion in Parkinson's disease for neuromelanin-positive (corrected  $p=0.017$ ), but not neuromelanin-negative cells ( $p=0.202$ ). We reviewed the histology report of the control with the highest gains in dopaminergic neurons (C4; 1.87%), and noted incidental  $\alpha$ -synuclein pathology, with rare Lewy neurites in the substantia nigra, but no Lewy bodies. The few cells with more than 3 *SNCA* copies had 4, except for one which had 5. There were more in neuromelanin-positive than neuromelanin-negative cells overall, but this was not significant (7 / 9,209 and 5 / 20,172 respectively,  $p=0.088$ ). Within neuromelanin-positive cells, slightly more were in Parkinson's than controls, but this was not significant either (5 / 5,677 and 2 / 3,532 respectively,  $p=0.715$ ).

### **More complex patterns than straightforward gains are occasionally seen**

We also noted very rare nuclei with other patterns (Supplementary Table 3; Supplementary fig. 6). Loss of one or both probes, seen in 0.53% and 0.12% of all neuromelanin-positive and

neuromelanin-negative cells respectively ( $p < 0.001$ ), could represent a real loss, or a sectioning artefact. Apparent gains of both *SNCA* and reference, seen in 0.14% and 0.16% of all neuromelanin-positive and neuromelanin-negative cells respectively ( $p = 0.017$ ), could be gains of the entire chromosome 4, or at least its long arm, or could be due to overlapping nuclei. Three of these cells, all in Parkinson's disease, had  $>5$  *SNCA* copies, with the highest number being 8 (one neuromelanin-positive; Supplementary fig. 6A). There were also occasional cells with other patterns, e.g. 3 *SNCA* / 1 reference. It is hard to attribute this to sectioning, and the reverse pattern was never seen. Out of eleven cells with this pattern (including three neuromelanin-positive), ten were in disease (nine Parkinson's, one MSA), although not significantly commoner in Parkinson's compared to control ( $p = 0.135$ ). Three of these were from the same Parkinson's disease nigra (case PD35; one neuromelanin-positive; Supplementary fig. 6C). To ensure that we were not biasing our results by excluding cells with losses, or gains of both *SNCA* and reference, we also performed the following supplementary analyses. We calculated the mean *SNCA* copy number in cells with 2 or more copies of the reference (i.e. including nuclei with possible chromosomal gains, but excluding any which might have been partially sectioned). We also analysed the ratio of the total number of *SNCA* and reference signals, across all cells, including any with losses, as sectioning artefacts should have an equal chance of affecting *SNCA* or the reference locus. The results were essentially the same (Supplementary fig. 7), eliminating these rare nuclei as the cause of the differences observed.

### **The level of *SNCA* gains in dopaminergic neurons is negatively correlated with Parkinson's disease onset age**

If somatic mutations contribute to Parkinson's disease, their level may be higher in earlier onset (Frank, 2010; Proukakis *et al.*, 2013). Spearman correlation analysis of percentage of neuromelanin-positive cells with *SNCA* gain in Parkinson's disease showed a significant negative correlation with onset age ( $r = -0.391$ ,  $p = 0.013$ ), but not disease duration, or age of death (table 2). There was no significant correlation in neuromelanin-negative cells. We also found no correlation with age of death in controls ( $r = 0.002$ ,  $p = 0.995$ ). The only two Parkinson's disease cases with no gains in either cell type had the latest onset (case PD34; 72) and the third-longest disease duration (case PD40; 30 years). The only individual with onset under 50 who had no *SNCA* gains in neuromelanin-positive cells was the *PARK2* missense variant carrier (case PD3), with a 31-year disease duration. To investigate *PARK2* further, we

performed array-CGH on nigral DNA, which had not been done before in this case; this revealed one, and possibly two, *PARK2* intronic deletions (Supplementary fig. 8).

### **Parkinson's cases with tremor at onset are less likely to have gains**

We compared clinical features of cases with and without gains in neuromelanin-positive cells (Supplementary table 1; Supplementary table 4). All cases presenting without tremor or asymmetry had gains, while asymmetric cases, which mostly also had tremor, sometimes did not. This difference was significant for tremor ( $p=0.035$ ) but not for asymmetry ( $p=0.156$ ). Hallucinations were reported in the majority of both groups ( $p=0.447$ ). The only cases with no gains in either cell type (PD34 and PD40) had asymmetry and tremor, and no hallucinations.

### **Nigra from two MSA brains shows the highest gains**

We studied five MSA nigras, all with nigral neuronal synuclein pathology (nigrostriatal or mixed pattern), but did not include them in the statistical analysis due to their small number. We noted gains in neuromelanin-negative cells in all, and in four also in neuromelanin-positive. Indeed, the highest percentage of cells with *SNCA* gains were neuromelanin-positive in two MSA cases, studied in different experiments (case MSA1- 2.76, case MSA2- 2.48) (Fig. 4A,B; Supplementary table 1).

### **There are no significant pathological correlates of gains in dopaminergic neurons**

We compared Parkinson's cases staged by the McKeith and Braak criteria (Supplementary table 1) and found no significant differences in gains (Supplementary table 4). With the McKeith criteria, the median percentage of neuromelanin-positive gains in each case was 0.64 in limbic and 0.73 in neocortical cases (Mann-Whitney  $p=0.187$ ). When using the Braak criteria, the values were 0.32 in stage 5 and 0.73 in stage 6 ( $p=0.157$ ). The proportions of cases with any gains was also not significantly different, although higher in the more advanced cases ( $p=0.316$  for McKeith and 0.213 for Braak). The pathology of three cases with gains (PD22, PD35, PD39) and one without (PD34) was reviewed blindly, with no obvious differences. Review of the *PARK2* carrier (case PD3) did not reveal unusual features. Blinded review of four MSA cases showed no distinct pathological features in the ones with more gains, although the two cases with the highest proportions of dopaminergic cells with gains had the highest ventrolateral neuronal loss (Supplementary Table 5).

### **Chromosome 4 paint indicates gains are on chromosome 4**

We hypothesised that *SNCA* gains were likely to be in tandem, on the same chromosome, rather than translocations. We stained a nigra section from case MSA1 with *SNCA* and chromosome paint (unblinded). We confirmed the rate of *SNCA* gains by extensive counting (neuromelanin-positive: 6/196, 3.06%; neuromelanin-negative: 28/3431, 0.82%). We recorded whether *SNCA* signals appeared to colocalise with chromosome 4. In normal cells, 97.5% of *SNCA* signals did (figure 4C); the remaining 2.5% may reflect the limitation of this method. In cells with gains, 95.1% of *SNCA* signals were on chromosome 4 (18/19 in neuromelanin-positive, 89/93 in neuromelanin-negative). This is consistent with most, if not all, the *SNCA* gains being on chromosome 4.

### **FISH with new probes and dye-swap confirms results**

To ensure that the *SNCA* excess signals were not probe artefacts, we obtained a custom-designed green *SNCA* probe, and verified it on *SNCA* triplication fibroblasts (Supplementary fig. 9). We used two different red reference probes, one also targeting a gene on chromosome 4 (*PDGFRA*), 1 megabase away from the *FIPL1* reference, and one on chromosome 7 (*IZKF1*) (figure 1A). We performed three separate experiments using the green *SNCA* probe and the new references, on a total of seven nigra slides. Two experiments included slides hybridised with the original *SNCA* red / reference green probes, blinded for the probe type / colour. We found that the pattern was now essentially reversed, with green gains, and almost no red gains (Fig. 5A; details in Supplementary Table 6). Case MSA1 (analysed blinded to the probes) showed *SNCA* gains again. Nigra from case MSA4, which had not been analysed before, showed remarkably similar gains when tested with green and red *SNCA* in a fully blinded experiment: in neuromelanin-positive cells, 0.66% with red *SNCA*, 0.64% with green; in neuromelanin-negative cells, 1.25% with red *SNCA*, 0.96% with green (Fig. 5B). Finally, the total number of signals with the green *SNCA* probe now showed a clear excess of the green probe (ratio 1.0045 in neuromelanin-positive, and 1.0018 in neuromelanin-negative cells).

### **Reference probes show minimal or no gains**

To ensure that gains were not simply a common feature of all probes, we calculated the proportion of cells of each type showing unique reference gains, which was  $\leq 0.1\%$  in neuromelanin-positive cells for all three probes used (Supplementary Table 8). The green *FIPL1* reference probe had gains in 0.10% of neuromelanin-positive cells and 0.046% of neuromelanin-negative. The two red reference probes showed no gains in neuromelanin-



positive cells; in neuromelanin-negative cells, *PDGFRA* showed no gains, and *IKZF1* had gains 0.23%.

### **Gains may also occur outside nigra**

In case MSA1, we also found gains in the putamen (5/476 cells; 1.05%), but not in frontal cortex (230 cells) or occipital cortex (408 cells) (Supplementary fig. 10A). In the cerebellum, only one of 428 cells had a unique *SNCA* gain, although differentiating nuclei was difficult due to the tightly packed granule cells. We also studied the frontal cortex from four Parkinson's disease brains in unblinded experiments, using NeuN immunohistochemistry to identify neurons (Supplementary table 1). We observed three neurons with *SNCA* gains after counting ~150 in each (0.49%; Supplementary fig. 10b), with no apparent glial gains.

## Discussion

We investigated somatic *SNCA* gains in sporadic synucleinopathies by three different methods. To analyse DNA extracted from tissue homogenates, predominantly substantia nigra, we used digital droplet PCR and customised array comparative genomic hybridisation. We excluded high-level *SNCA* copy number variant mosaicism (~20-30% of cells having an extra copy). These results could be consistent with a low level of somatic gains, or their presence only in certain cell types. We thus next investigated *SNCA* copy number in individual brain cells by FISH, focusing on the substantia nigra, where neuromelanin staining allows identification of dopaminergic neurons (Kordower *et al.*, 2013; Nichterwitz *et al.*, 2016). We demonstrated low level gains of *SNCA*, commoner in nigral dopaminergic neurons of Parkinson's disease than controls, negatively correlated with onset age, validated by probe dye-swapping, with the highest levels in two MSA cases. As FISH may underestimate copy number by not resolving nearby gains (Bushman *et al.*, 2015), which we indeed noticed in the brain and fibroblast samples with known inherited copy number variants (Fig. 2; Supplementary fig. 9), we may have underestimated gains. Indeed, we noted several cells where larger *SNCA* signals might have been due to this (Fig. 2C, 5B), but were conservative in our calls, requiring the presence of a spatially distinct signal on the microscope, and not calling additional copies based on signal size. We also noted occasional cells with more than 3 *SNCA* copies, and cells with increased numbers of *SNCA* and reference signals, which may have had up to 8 *SNCA* copies, which would further increase the alpha-synuclein load.

The gains in Parkinson's disease nigral dopaminergic neurons showed significant negative correlation with onset age, as expected if somatic mutations have a role in disease development (Frank, 2010). Review of clinical data revealed an intriguing association of lack of tremor at presentation with presence of gains, suggesting that cases with tremor at onset may be less likely to have *SNCA* mosaicism. We are, by definition, only able to assess the cells which are still alive at the time of the individual's death, and cannot determine any possible gains in cells which preferentially died earlier. Neuromelanin staining may help detect more surviving dopaminergic neurons, as loss of tyrosine hydroxylase may occur before loss of melanised neurons (Kordower *et al.*, 2013). Even so, and despite trying to analyse cases with relatively short disease duration, it should be noted that half the dopaminergic neurons may be lost before a diagnosis is made, with an estimated 33-80% melanised neurons lost in the first decade (Kordower *et al.*, 2013), and our cases were all advanced pathologically (mostly Braak stage

6). The true burden of somatic *SNCA* gains in early stage of Parkinson's disease, or before onset, could therefore be much higher. We did not find significant pathological correlations with the presence or level of mosaicism, but our study was not designed for this, and more cases, particularly early stage, would be needed. We note that the highest levels of neuronal gains were in two MSA cases, and all five studied had gains in at least neuromelanin-negative cells. We did select MSA with clear neuronal nigral synuclein pathology, and cannot draw firm conclusions based on these few cases, but this clearly merits further study.

Notably, the only two Parkinson's disease cases without gains in either cell type had either a long duration (30 years) or late onset (age 72), suggesting that our selection of short duration and early onset was appropriate. Gains in dopaminergic neurons were absent in only one of ten individuals with onset under 50, who also carried a *PARK2* heterozygous missense variant, and at least one *PARK2* intronic deletion (Supplementary fig. 8). The lack of *SNCA* gains could be due to neurons carrying them being lost due to the very long disease duration (31 years, the joint longest). We wonder whether parkin dysfunction might have exacerbated this, or conversely whether *SNCA* gains are not present in cases with *PARK2* mutations. As the pathology of this case was typical Parkinson's, however, and the significance of a single *PARK2* coding variant, and intronic *PARK2* copy number variants, remain unclear (Wang *et al.*, 2013), this may be a coincidence. *PARK2* is also in a fragile site, and somatic deletions have been reported in glioblastoma (Veeriah *et al.*, 2010), but our array-CGH did not reveal other deletions.

We have been careful to base our analysis on cells which had two copies of the reference, and at least two of *SNCA*, to minimise any sectioning artefacts. The very low proportion of cells with losses (Supplementary Table 4) suggests that we did not often have partial nuclei. When we analysed data in different ways, also including cells with gains of both probes, or any losses, our key finding of gains in dopaminergic neurons being commoner in Parkinson's compared to controls was still apparent (Supplementary fig. 7). Although sectioning artefacts are possible, gains of both *SNCA* and reference could indicate a somatic whole chromosome gain, which would also increase the alpha-synuclein load. Mosaic aneuploidy has been consistently reported in normal neurons in other brain regions, albeit at varying levels (Iourov *et al.*, 2009; Knouse *et al.*, 2014; Rehen *et al.*, 2005). Evidence is increasing for DNA replication in specific neuronal populations, in development or in neurodegeneration (Frade *et al.*, 2015), and polyploidy has been reported in Parkinson's disease nigral neurons (Höglinger

*et al.*, 2007). With regard to the few cells with gains of *SNCA* and loss of reference, possibly commoner in disease, the possibility of a somatic complex chromosomal rearrangement, as reported in mouse olfactory neurons (Hazen *et al.*, 2016), is a plausible but unproven explanation.

We only found minimal evidence of gains with three different reference probes. The main chromosome 4 reference probe, targeting *FIPL1*, had gains in 0.1% and <0.05% in neuromelanin-positive and neuromelanin-negative cells respectively. These may represent true gains, or the false positive rate of our technique. No gains were seen with the other chromosome 4 probe. The relative lack of gains with the chromosome 4 probes used as reference is consistent with gains restricted to the fragile site around *SNCA*. The third reference probe, targeting a chromosome 7 locus, had gains only in 0.23% of neuromelanin-negative cells. This probe detects a locus which can be deleted in haematopoietic malignancies (Schwab *et al.*, 2013), although we are not aware of any somatic gains.

The somewhat surprising result of low level gains even in some controls may indicate genomic instability of *SNCA* in at least nigral dopaminergic neurons or their precursors, consistent with its presence in a fragile site (Rozier *et al.*, 2004). Interestingly, the control with the highest gains in dopaminergic neurons had rare Lewy neurites in the substantia nigra. It is impossible to know if she had lived past the age of 82 whether she would have developed the picture of incidental Lewy body disease, or even clinical Parkinson's disease. This highlights the difficulty in selecting controls in an age-related condition, which could have been even more problematic if we had selected younger ones. We should note that this is not the first report of somatic *SNCA* gains in normal human neurons, as a gain including *SNCA* was present in a frontal neuron from a young control (FCTX225 in McConnell *et al.*, 2013).

Our study was mostly limited to one section from each nigra, and cells were selected based on neuromelanin positivity, rather than their precise anatomical location. Further work should seek to assess topographically defined nigral neuronal populations, differentiating between the ventrolateral and dorsomedial regions, correlate gains with alpha-synuclein mRNA and protein levels, and immunologically characterise neuromelanin-negative cells with gains. We have not performed a definitive assessment of other brain regions, or peripheral neuroectodermal structures, where pathology is often found (Beach *et al.*, 2010). Limited analysis of NeuN-labelled frontal neurons, and one MSA putamen and cerebellum, suggested some gains. We did

not detect any significant pathological correlates of gains, and the lower prevalence of gains in cases with tremor has to be viewed as preliminary. More widespread regional analysis, and detailed correlation with clinical and pathological features, would be of great interest.

Further molecular characterisation of *SNCA* gains, and detection of low-level copy number variants in other relevant genes like *PARK2*, may require single cell whole genome sequencing, which is advancing rapidly (Gawad *et al.*, 2016). Large copy number variants can be detected, although uneven amplification, which could lead to false positives, can only be minimised with novel protocols (Chen *et al.*, 2017; Zahn *et al.*, 2017), untested on neurons. There are no sequencing data on nuclear DNA from single nigral neurons reported to our knowledge. One related finding, however, is the evidence for DNA content variation in neurons. Frontal neurons have increased DNA content compared to cerebellar and to lymphocytes (Westra *et al.*, 2010). This is particularly prominent in Alzheimer neurons, with some of the increase due to extra copies of *APP*, although the majority of the additional DNA remained unexplained (Bushman *et al.*, 2015). DNA content of nigral neurons was increased in Lewy body diseases compared to control and to other regions, without the source of extra DNA being known (Yang *et al.*, 2014), but not in MSA (Yang *et al.*, 2017).

*SNCA* gains may arise in very early development, where rapid expansion of neuronal precursors takes place, DNA replication stress may predispose to copy number variants in fragile sites, such as the ones where *SNCA* and *PARK2* reside, and DNA breaks occur. The amount of alpha-synuclein, an important pre-synaptic protein, influences the development of a subset of nigral dopaminergic neurons (Garcia-Reitboeck *et al.*, 2013), so developing neurons with losses might be at a selective disadvantage, although we cannot predict the effect of gains. Gains may additionally, or alternatively, arise in post-mitotic neurons, for example by aberrant cell cycle re-entry as claimed in Alzheimer disease (Yang *et al.*, 2001), with mitotic activation also reported in Parkinson's nigral neurons (Höglinger *et al.*, 2007). In this case, they could be a result of, rather than contributor to, the disease. Another question is whether *SNCA* gains in controls might make individual neurons more vulnerable to ageing, as reported for hyperploidy in Alzheimer's (Arendt *et al.*, 2010), and for increased DNA content in ageing (Fischer *et al.*, 2012). Although we used mostly old controls, we found no correlation of gains with age of death.

In conclusion, we present evidence of low level somatic *SNCA* gains in brain, predominantly in nigral dopaminergic neurons, where they are commoner in Parkinson's disease than controls, negatively correlated with onset age, and possibly commonest in some MSA cases. Our findings are analogous to Alzheimer's disease, where rare inherited *APP* gains are pathogenic, and somatic gains commoner in disease frontal neurons (Bushman *et al.*, 2015). Gains may arise in early embryonic development, in which case they might be detectable in peripheral neural-crest derived structures (Proukakis *et al.*, 2013). They could have a role in disease initiation, as "seeds" from where pathology propagates, or progression, by increasing sensitivity of neurons to intrinsic or extrinsic harmful processes. They may also result from aberrant DNA synthesis in neurodegeneration. Progress in single cell sequencing will help confirm the level of *SNCA* gains, define their breakpoints, and shed light on the timing and mechanism of their origin.

### **Acknowledgements**

We thank Professor John Hardy for very helpful discussions. We are grateful to the Queen Square and Parkinson's UK brain banks for providing the samples for this study, and the South West England Brain Bank and Professor Seth Love for providing sections from the *SNCA* duplication brain, and to all the patients and controls who donated their brains to research. The Queen Square Brain Bank is supported by the Reta Lila Weston Institute for Neurological Studies and the Medical Research Council UK. The Parkinson's UK Tissue Bank is funded by Parkinson's UK, a charity registered in England and Wales (258197) and in Scotland (SC037554). We would also like to thank Diana Brazma and Colin Grace for help with array CGH.

### **Funding**

This work was supported by the Michael J Fox Foundation for Parkinson's Research. TL is supported by an Alzheimer's Research UK senior fellowship. HL is supported by a CBD Solutions Grant.

## Tables

	NM +ve cells			NM -ve cells			All cells		
	Gain	No gain	% gain	Gain	No gain	% gain	Gain	No gain	% gain
<b>Parkinson's</b>	31	9	77.5	29	11	72.5	38	2	95
<b>Control</b>	10	15	40	13	12	52	17	8	68
<b>p</b>	<b>0.0036</b>			0.1144			<b>0.0052</b>		

**Table 1. Summary of Parkinson's disease and control nigras that had any cells with a *SNCA* gain.** Results are shown by cell type and overall, with the associated p value for each (Fisher's exact test).

Cell type	Onset age		Disease duration		Age of death	
	p	r	p	r	p	r
<b>Neuromelanin-positive</b>	0.013	-0.39	0.062	0.30	0.255	-0.18
<b>Neuromelanin-negative</b>	0.921	-0.02	0.344	-0.15	0.198	-0.21

**Table 2. Correlation of proportion of cells with *SNCA* gains in Parkinson's disease with onset age, disease duration, and age of death.** Results are shown by cell type.

## Figure legends

### Figure 1. SureFISH probe selection and validation.

The targets of the *SNCA* and reference probes, together with the probe colour and target size in each case, the fragile site around *SNCA*, and the relevant distances, are shown.

(A) Initial probes, used for majority of experiments, and their localisation on chromosome 4. (B) Further probes used for validation as outlined in results. Chr4 (above) and chr7 (below). Validation of the red *SNCA* probe was performed on (C) fibroblasts with a triplication, and (D) brain with a duplication (cingulate gyrus shown). Arrows point to clearly resolved copies of *SNCA*. Arrowheads point to examples of large signals representing two copies, which could not be fully resolved.

Figure 2. Examples of cells with normal copies (A), and *SNCA* gain (B, C). Neuromelanin-positive cells are above, and neuromelanin-negative below. Scale bar = 3µm. Arrows in (C) indicate elongated *SNCA* signals, which could represent additional copies which cannot be resolved by this method.

Figure 3. Percentage of nuclei with *SNCA* gain in each Parkinson's disease (PD) and control nigra, per cell type. Mean and 95% CI shown. NM+ indicates cells with neuromelanin, and NM- without.

### Figure 4. *SNCA* gains in MSA nigra samples.

(A) The percentage of cells of each type with unique *SNCA* gains in each MSA nigra. The two results from each nigra are joined by a line.

(B) An example of a neuromelanin-positive cell with a gain.

(C) A cell stained with *SNCA* and chromosome 4 paint, showing 4 copies of *SNCA* (arrowed), all localising to chromosome 4. Scale bar = 3 µm.

### Figure 5. Dye swap studies in MSA nigra.

The probes were blinded in both cases. Further details are given in Supplementary Table 6. (A) *SNCA* green, on previously studied case MSA1. Neuromelanin-positive and neuromelanin-negative cells with gains are shown with both reference probes. (B) *SNCA* red and green, on case MSA4 nigra, never studied before. Experiment blinded to probes, diagnosis, and the fact these slides were from the same nigra. Neuromelanin-positive and neuromelanin-negative cells



with *SNCA* gains revealed by both probes are shown. The arrow points to an unusually large *SNCA* signal, in a nucleus with at least one extra copy.

For Peer Review

## References

- Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, Tomasini L, et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* 2012; 492: 438–42.
- Arendt T, Brückner MK, Mosch B, Losche A. Selective cell death of hyperploid neurons in Alzheimer's disease. *Am J Pathol* 2010; 177: 15–20.
- Arlt MF, Mülle JG, Schaibley VM, Ragland RL, Durkin SG, Warren ST, et al. Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet* 2009; 84: 339–50.
- Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 2011; 479: 534–537.
- Beach TG, Adler CH, Sue LI, Vedders L, Lue L, White Iii CL, et al. Multi-organ distribution of phosphorylated alpha-synuclein histopathology in subjects with Lewy body disorders. *Acta Neuropathol* 2010; 119: 689–702.
- Beck JA, Poulter M, Campbell TA, Uphill JB, Adamson G, Geddes JF, et al. Somatic and germline mosaicism in sporadic early-onset Alzheimer's disease. *Hum Mol Genet* 2004; 13: 1219–1224.
- Braak H, del Tredici K, Rüb U, de Vos R, Jansen Steur ENH, et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* 2003; 24: 197–211.
- Bushman DM, Kaeser GE, Siddoway B, Westra JW, Rivera RR, Rehen SK, et al. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *Elife* 2015; 4: e05116.
- Cai X, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, Poduri A, et al. Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain. *Cell Rep.* 2014; 8: 1280–9.
- Chen C, Xing D, Tan L, Li H, Zhou G, Huang L, et al. Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* 2017; 356: 189–194.
- Chun J, Schatz DG. Rearranging Views on Neurogenesis: Neuronal Death in the Absence of DNA End-Joining Proteins. *Neuron* 1999; 22: 7–10.
- Cykowski MD, Coon EA, Powell SZ, Jenkins SM, Benarroch EE, Low PA, et al. Expanding the spectrum of neuronal pathology in multiple system atrophy. *Brain* 2015; 138: 2293–2309.
- Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, et al. Cell Lineage Analysis in Human Brain Using Endogenous Retroelements. *Neuron* 2015; 85: 49–59.
- Federoff M, Price TR, Sailer A, Scholz S, Hernandez D, Nicolas A, et al. Genome-wide estimate of the heritability of Multiple System Atrophy. *Parkinsonism Relat Disord* 2016; 22: 35–41.
- Fischer H-G, Morawski M, Brückner MK, Mittag A, Tarnok A, Arendt T. Changes in neuronal DNA content variation in the human brain during aging. *Aging Cell* 2012; 11: 628–33.
- Forsberg LA, Absher D, Dumanski JP. Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. *J Med Genet* 2013; 50: 1–10.
- Frade JM, Ovejero-Benito MC. Neuronal cell cycle: the neuron itself and its circumstances. *Cell cycle* 2015; 14: 712–720.
- Frank SA. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc Natl Acad Sci U S A* 2010; 107 Suppl: 1725–1730.
- Frigerio CS, Lau P, Troakes C, Deramecourt V, Gele P, Van Loo P, et al. On the

- identification of low allele frequency mosaic mutations in the brains of Alzheimer disease patients. *Alzheimers Dement* 2015.
- Fujishiro H, Imamura AY, Lin W-L, Uchikado H, Mark MH, Golbe LI, et al. Diversity of pathological features other than Lewy bodies in familial Parkinson's disease due to SNCA mutations. *Am J Neurodegener. Dis.* 2013; 2: 266–75.
- Fungtammasan A, Walsh E, Chiaromonte F, Eckert K a, Makova KD. A genome-wide analysis of common fragile sites: what features determine chromosomal instability in the human genome? *Genome Res* 2012; 22: 993–1005.
- Garcia-Reitboeck P, Anichtchik O, Dalley JW, Ninkina N, Tofaris GK, Buchman VL, et al. Endogenous alpha-synuclein influences the number of dopaminergic neurons in mouse substantia nigra. *Exp Neurol* 2013; 248: 541–5.
- Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Gene.* 2016;
- Gole J, Gore A, Richards A, Chiu Y-J, Fung H-L, Bushman D, et al. Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat Biotechnol* 2013; 31: 1126–32.
- Hazen JL, Faust GG, Rodriguez AR, Ferguson WC, Shumilina S, Clark RA, et al. The Complete Genome Sequences, Unique Mutational Spectra, and Developmental Potency of Adult Neurons Revealed by Cloning. *Neuron* 2016; 89: 1223–1236.
- Höglinger GU, Breunig JJ, Depboylu C, Rouaux C, Michel PP, Alvarez-Fischer D, et al. The pRb/E2F cell-cycle pathway mediates cell death in Parkinson's disease. *Proc Natl Acad Sci USA* 2007; 104: 3585–90.
- Iourov IY, Vorsanova SG, Liehr T, Yurov YB. Aneuploidy in the normal, Alzheimer's disease and ataxia-telangiectasia brain: differential expression and pathological meaning. *Neurobiol Dis* 2009; 34: 212–220.
- Kara E, Kiely AP, Proukakis C, Giffin N, Love S, Hehir J, et al. A 6.4 Mb duplication of the  $\alpha$ -synuclein locus causing frontotemporal dementia and Parkinsonism: phenotype-genotype correlations. *JAMA Neurol* 2014; 71: 1162–71.
- Kasten M, Klein C. The many faces of alpha-synuclein mutations. *Mov Disord* 2013; 28: 697–701.
- Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simón-Sánchez J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum Mol Genet* 2012; 21: 4996–5009.
- Kiely AP, Asi YT, Kara E, Limousin P, Ling H, Lewis P, et al.  $\alpha$ -Synucleinopathy associated with G51D SNCA mutation: a link between Parkinson's disease and multiple system atrophy? *Acta Neuropathol* 2013; 125: 753–69.
- Kluwe L. Digital PCR for discriminating mosaic deletions and for determining proportion of tumor cells in specimen. *Eur J Hum Genet* 2016; 24: 1644–1648.
- Knouse KA, Wu J, Whittaker CA, Amon A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proc Natl Acad Sci USA* 2014; 111: 13409–14.
- Kordower JH, Olanow CW, Dodiya HB, Chu Y, Beach TG, Adler CH, et al. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. *Brain* 2013; 136: 2419–2431.
- Leija-Salazar M, Piette CL, Proukakis C. Invited Review: Somatic mutations in neurodegeneration. *Neuropathol Appl Neurobiol* 2018; doi: 10.1111/nan.12465. [Epub]
- Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 2015; 350: 94–98.
- Lupski JR. Genetics. Genome mosaicism--one human, multiple genomes. *Science* 2013; 341: 358–9.

- Macosko EZ, McCarroll SA. Exploring the variation within. *Nat Genet* 2012; 44: 614–616.
- McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, et al. Mosaic copy number variation in human neurons. *Science* 2013; 342: 632–7.
- McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: Third report of the DLB consortium. *Neurology* 2005; 65:1863-1872.
- Mkrtchyan H, Gross M, Hinreiner S, Polytiko A, Manvelyan M, Mrasek K, et al. Early embryonic chromosome instability results in stable mosaic pattern in human tissues. *PLoS One* 2010; 5: e9591.
- Mullin S, Schapira A. The genetics of Parkinson's disease. *Br Med Bull* 2015; 114: 39–52.
- Nacheva E, Mokretar K, Soenmez A, Pittman AM, Grace C, Valli R, et al. DNA isolation protocol effects on nuclear DNA analysis by microarrays, droplet digital PCR, and whole genome sequencing, and on mitochondrial DNA copy number estimation. *PLoS One* 2017; 12: e0180467.
- Nichterwitz S, Chen G, Aguila Benitez J, Yilmaz M, Storrval H, Cao M, et al. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat Commun* 2016; 7: 12139.
- O'Huallachain M, Karczewski KJ, Weissman SM, Urban AE, Snyder MP. Extensive genetic variation in somatic human tissues. *Proc Natl Acad Sci USA* 2012; 109: 18018–23.
- Oliveira VC, Carrara RC V, Simoes DLC, Saggiaro FP, Carlotti CG, Covas DT, et al. Sudan Black B treatment reduces autofluorescence and improves resolution of in situ hybridization specific fluorescent signals of brain sections. *Histol Histopathol* 2010; 25: 1017–24.
- Petrucci S, Ginevrino M, Valente EM. Phenotypic spectrum of alpha-synuclein mutations: New insights from patients and cellular models. *Parkinsonism Relat Disord* 2016; 22: S16–S20.
- Proukakis C, Houlden H, Schapira AH. Somatic alpha-synuclein mutations in Parkinson's disease: hypothesis and preliminary data. *Mov Disord* 2013; 28: 705–12.
- Proukakis C, Shoaee M, Morris J, Brier T, Kara E, Sheerin U-M, et al. Analysis of Parkinson's disease brain-derived DNA for alpha-synuclein coding somatic mutations. *Mov Disord* 2014; 29: 1060–4.
- Rehen SK, Yung YC, McCreight MP, Kaushal D, Yang AH, Almeida BS V, et al. Constitutional aneuploidy in the normal human brain. *J Neurosci* 2005; 25: 2176–80.
- Ross OA, Braithwaite AT, Skipper LM, Kachergus J, Hulihan MM, Middleton FA, et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. *Ann Neurol* 2008; 63: 743–750.
- Rozier L, El-Achkar E, Apiou F, Debatisse M. Characterization of a conserved aphidicolin-sensitive common fragile site at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. *Oncogene* 2004; 23: 6872–6880.
- Schwab CJ, Chilton L, Morrison H, Jones L, Al-Shehhi H, Erhorn A, et al. Genes commonly deleted in childhood B-cell precursor acute lymphoblastic leukemia: association with cytogenetics and clinical features. *Haematologica* 2013; 98: 1081–1088.
- Schwer B, Wei P-C, Chang AN, Kao J, Du Z, Meyers RM, et al. Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc Natl Acad Sci USA*. 2016: 201525564.
- Sugita S, Aoyama T, Ito Y, Asanuma H, Sugawara T, Segawa K, et al. Diagnostic utility of automated SureFISH (Dako Omnis) in the diagnosis of musculoskeletal translocation-related sarcomas. *Pathol Int* 2017; 67: 510–513.
- Takeda M, Kasai T, Shimizu S, Kitaichi M, Kojima K, Nagoya A, et al. Assessment of ALK gene rearrangement in lung cancer using a new rapid automated SureFISH (Dako Omnis) assay. *J Clin Pathol* 2017; 70: 712–714.

- Thomas C a, Paquola ACM, Muotri AR. LINE-1 retrotransposition in the nervous system. *Annu Rev Cell Dev Biol* 2012; 28: 555–73.
- Valli R, Marletta C, Pressato B, Montalbano G, Lo Curto F, Pasquali F, et al. Comparative genomic hybridization on microarray (a-CGH) in constitutional and acquired mosaicism may detect as low as 8% abnormal cells. *Mol Cytogenet* 2011; 4: 13.
- Veeriah S, Taylor BS, Meng S, Fang F, Yilmaz E, Vivanco I, et al. Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies. *Nature Genet* 2010; 42: 77-82.
- Wang L, Nuytemans K, Bademci G, Jauregui C, Martin ER, Scott WK, et al. High-resolution survey in familial Parkinson disease genes reveals multiple independent copy number variation events in PARK2. *Hum Mutat* 2013; 34: 1071–4.
- Wei P-C, Chang AN, Kao J, Du Z, Meyers RM, Alt FW, et al. Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* 2016; 164: 644–655.
- Westra JW, Rivera RR, Bushman DM, Yung YC, Peterson SE, Barral S, et al. Neuronal DNA content variation (DCV) with regional and individual differences in the human brain. *J Comp Neurol* 2010; 518: 3981–4000.
- Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol Aging* 2011; 32: 1923.
- Yang Y, Geldmacher DS, Herrup K. DNA replication precedes neuronal cell death in Alzheimer's disease. *J Neurosci* 2001; 21: 2661–8.
- Yang Y, Shepherd C, Halliday G. Aneuploidy in Lewy body diseases. *Neurobiol Aging* 2015. 36: 1253-60.
- Yang Y, Shepherd CE, Halliday GM. Increased aneuploidy is not a universal feature across  $\alpha$ -synucleinopathies. *Mov Disord* 2017; 32: 475–476.
- Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, Shah SP, et al. Scalable whole-genome single-cell library preparation without preamplification. *Nat Methods* 2017; 14: 167–173.
- Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet* 2015; 16: 172–183.

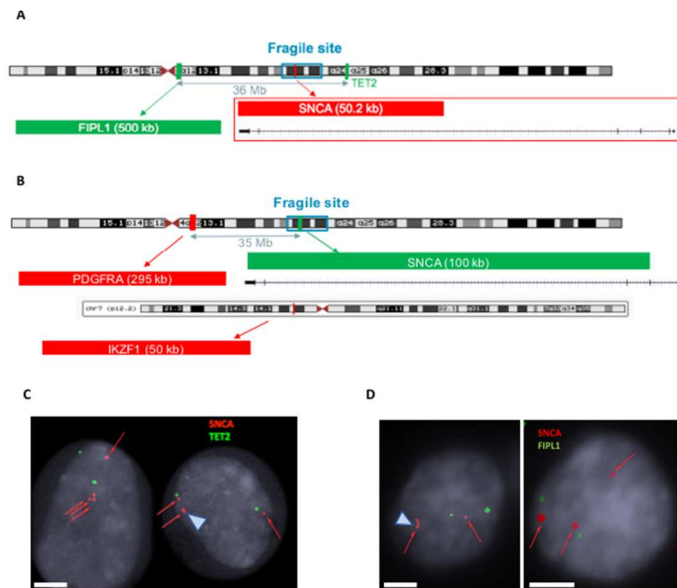


Figure 1. SureFISH probe selection and validation. The targets of the SNCA and reference probes, together with the probe colour and target size in each case, the fragile site around SNCA, and the relevant distances, are shown. (A) Initial probes, used for majority of experiments, and their localisation on chromosome 4. (B) Further probes used for validation as outlined in results. Chr4 (above) and chr7 (below). Validation of the red SNCA probe was performed on (C) fibroblasts with a triplication, and (D) brain with a duplication (cingulate gyrus shown). Arrows point to clearly resolved copies of SNCA. Arrowheads point to examples of large signals representing two copies, which could not be fully resolved.

190x275mm (96 x 96 DPI)

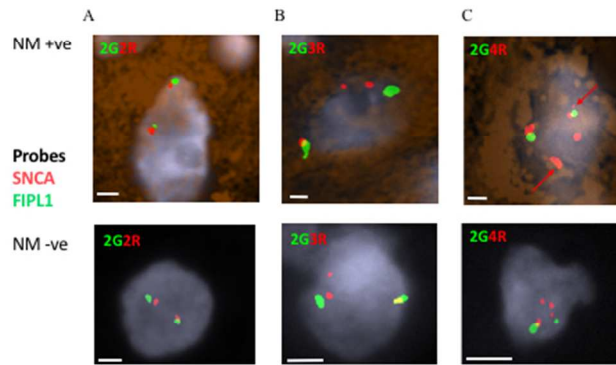


Figure 2. Examples of cells with normal copies (A), and SNCA gain (B, C). Neuromelanin-positive cells are above, and neuromelanin-negative below. Scale bar = 3 $\mu$ m. Arrows in (C) indicate elongated SNCA signals, which could represent additional copies which cannot be resolved by this method.

338x190mm (54 x 54 DPI)

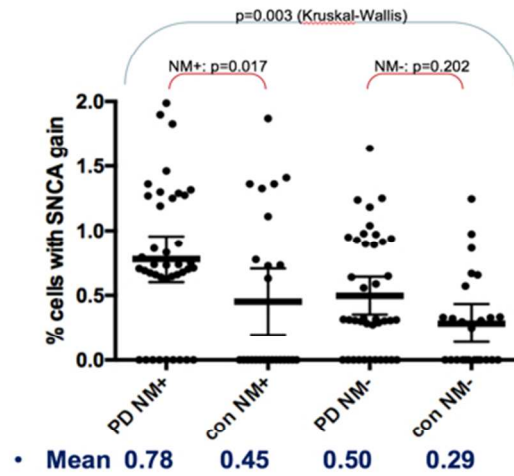


Figure 3. Percentage of nuclei with SNCA gain in each Parkinson's disease (PD) and control nigra, per cell type. Mean and 95% CI shown. NM+ indicates cells with neuromelanin, and NM- without.

338x190mm (54 x 54 DPI)



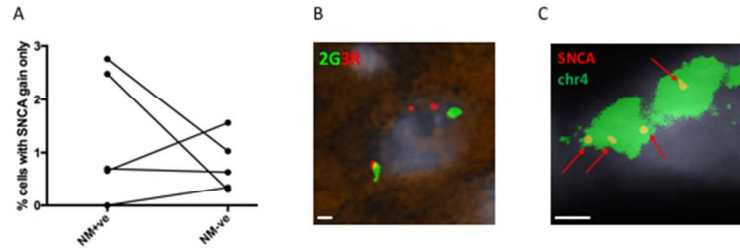


Figure 4. SNCA gains in MSA nigra samples.

(A) The percentage of cells of each type with unique SNCA gains in each MSA nigra. The two results from each nigra are joined by a line.

(B) An example of a neuromelanin-positive cell with a gain.

(C) A cell stained with SNCA and chromosome 4 paint, showing 4 copies of SNCA (arrowed), all localising to chromosome 4. Scale bar = 3  $\mu$ m.

338x190mm (54 x 54 DPI)

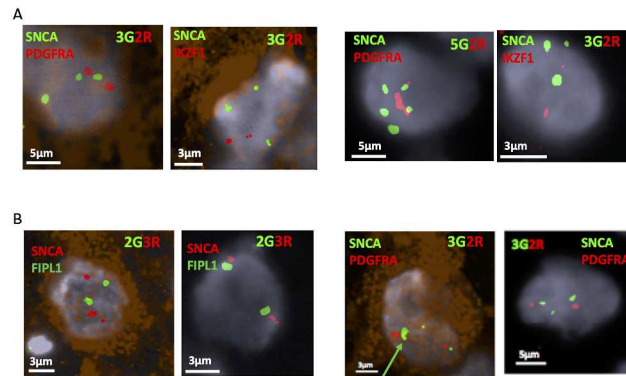


Figure 5. Dye swap studies in MSA nigra.

The probes were blinded in both cases. Further details are given in Supplementary Table 6. (A) SNCA green, on previously studied case MSA1. Neuromelanin-positive and neuromelanin-negative cells with gains are shown with both reference probes. (B) SNCA red and green, on case MSA4 nigra, never studied before. Experiment blinded to probes, diagnosis, and the fact these slides were from the same nigra. Neuromelanin-positive and neuromelanin-negative cells with SNCA gains revealed by both probes are shown. The arrow points to an unusually large SNCA signal, in a nucleus with at least one extra copy.

297x209mm (300 x 300 DPI)

**Supplementary note 1. Protocol optimization for FISH on brain sections.** Optimisation of the FISH protocol was performed mostly on cingulate gyrus sections, and using only a reference probe where SureFISH used, to avoid consuming valuable nigra and custom designed probes. The following parameters were tested:

1. Probe type: BAC probe use on sections (before SureFISH probes delivered) - tried on fixed 10 and 20  $\mu\text{m}$  sections, with Dako Target Retrieval Solution (Agilent) and citric acid pre-treatment for antigen retrieval. Poor images, so only SureFISH used henceforth.
2. Sudan black B (to reduce autofluorescence) - 0.1% for 20 minutes<sup>1</sup> and 0.2% for 10 minutes<sup>2</sup> in fixed sections. The former was marginally clearer so was used henceforth.
3. For fixed sections- antigen retrieval optimization was performed comparing Dako buffer and in-house made citric acid solution. Incubation for 2 hours at 80 degrees, and microwave heating for 5, 10 and 20 minutes, were tried. Optimal was citric solution for 20 minutes with microwave heating.
4. Section type - frozen v fixed (Dako pre-treated). The signals appeared brighter and cleaner in frozen, and only frozen sections were used henceforth (The exception was a brain with known duplication, used as positive control, where fixed only available).
5. Section thickness – 10 v 20  $\mu\text{m}$ . The thicker sections appeared to have multiple overlaying cells in some cases, and in nigra the neuromelanin frequently obscured the nucleus, so only 10  $\mu\text{m}$  sections used.
6. Denaturation of probes - separate or combined. Clearer images with separate denaturation.
7. Pepsin treatment - 10 mins v 30 mins. No clear difference, 30 minutes used.
8. Hybridisation time – overnight to 96 hours. At least 48 hours required, 72-96 hours optimal.
9. Reference probe- We first tested was *TET2* used it throughout the validation stages, but the signal was faint and *FIPIL* was obtained. This showed slightly clearer signals, and was used as our main reference, for all nigra experiments. Note that *TET2* was the reference probe used in the validation of the *SNCA* probe on fibroblasts and duplication brain as shown in figure 1.

## Supplementary note 2. Preliminary FISH analysis using BAC probes.

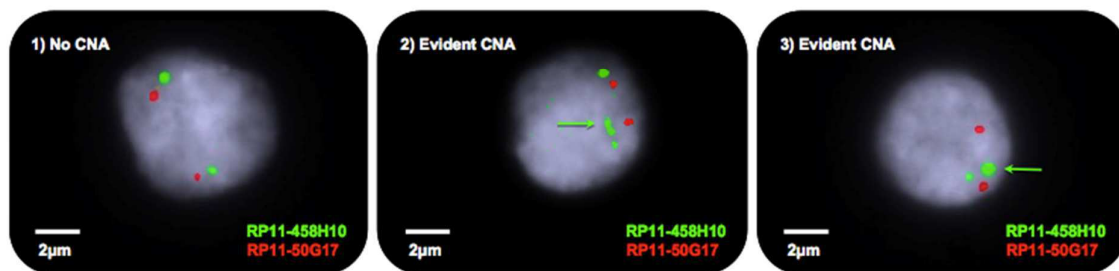
### Materials and methods

For preliminary FISH analysis, we obtained bacterial artificial chromosome (BAC) clones from BACPAC Resources Center (Children's Hospital Oakland Research Institute, Oakland, CA, USA) covering *SNCA* (RP11-458H10; chr4: 90,630,567 – 90,799,235), and a control region on the same chromosome, almost 50 Mb away (RP11-50G17; chr4: 42,226,848 - 42,404,980). BACs were grown on LB broth with chloramphenicol, and DNA extracted using QIAprep Spin Miniprep Kit (Qiagen). BAC DNA was labelled by standard nick translation<sup>3</sup>. Substantia nigra suspensions were made by first mincing approximately 50mg tissue on a disposable glass slide. Each minced sample was scooped into an Eppendorf tube containing 10ml of RPMI Medium (Life Technologies). Tubes were centrifuged for 10 min at 1500g, supernatants were removed and pellets were resuspended in 6ml of 0.075M KCl pre-heated to 37°C. Samples were then incubated for 40 min at 37°C, after which 1ml of ice-cold Carnoy's fixative (3:1 methanol : acetic acid) was added and tubes were centrifuged again for 10 min at 1500G, added drop-by-drop, with continuous vortexing of the tube. Supernatants were carefully removed and each pellet was resuspended in 5ml of ice-cold fixative. Tubes were subsequently kept at 4°C overnight and centrifuged for 5 min at 1200g. Supernatants were again removed and pellets were mixed with fresh fixative until samples had a milky appearance. A single droplet from each sample was then applied onto wet sterile glass slides. Prepared slides were tilted to allow sample spreading before 2-3 drops of fixative were similarly applied to each slide. Slides were then aged for 24 hours at room temperature. FISH was done as before<sup>3</sup>. Slides were incubated in 2×SSC, dehydrated through an ethanol series and air dried, prior to applying RNase and then pepsin, and a 3% formaldehyde / 0.01M MgCl<sub>2</sub> solution. They were then washed and again dehydrated through an ethanol series and air dried. Slides were then denatured in 70% formamide at 72°C for 85 sec, dehydrated through an ethanol series and air dried again. The hybridization buffer working solution comprised a 1:1:8:40:50 ratio of 1M pH7.4 Trizma Base, 10% Tween-20 stock, 2×SSC, 25% Dextran Sulfate stock, and ≥99% Formamide. The probes were then mixed as required, and for each slide, 3μl of each probe, and 3μl of the hybridization buffer (1:1:8:40:50 ratio of 1M pH7.4 Trizma Base, 10% Tween-20 stock, 2×SSC, 25% Dextran Sulfate stock, and ≥99% Formamide), were mixed and heated to 73°C for 5 minutes. The mixture was then applied to the dried slide and covered with a cover slip which was hybridised in a dark humidified 37°C chamber for 48-72 hours. Cover slips were removed and

slides were washed in a 0.4×SSC / 0.3% IGEPAL (Sigma-Aldrich) (73°C ~1 min with shaking), immersed in a second identical solution for 5 min, followed by two consecutive washes in 37°C 2×SSC / 0.1% IGEPAL solutions for 5 min each. DAPI was then applied to each slide. Images were reviewed on the same Olympus microscope.

## Results and discussion

In total, the nigra was analysed in 11 Parkinson's disease and 5 control nigras, in a blinded fashion. Although most cells were normal (1 in figure below), very occasional cells with apparent *SNCA* copy number alterations (CNAs) which were always gains were seen (2, arrow, likely two extra copies), and several more with suspicious larger *SNCA* signals (3, arrow).



The difficulties with this approach included the inability to determine cell type, the background noise which is a common problem due to BAC repetitive sequences, and the difficulty assessing the significance of large signals, which could represent tandem gains unresolvable by this technique, as discussed<sup>4</sup>. We considered these results as strongly suggestive of low level mosaicism, but did not attempt formal quantitation due to the limitations, and changed to using SureFISH on sections, as discussed in the main manuscript.

References for supplementary note (note: 1 and 4 are also cited in main manuscript)

1. Oliveira VC, Carrara RC V, Simoes DLC, Saggiaro FP, Carlotti CG, Covas DT, et al. Sudan Black B treatment reduces autofluorescence and improves resolution of in situ hybridization specific fluorescent signals of brain sections. *Histol Histopathol* 2010;25:1017–24.
2. Mizielska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P, et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol.* 2013;126:845–57.

3. Virgili A, Nacheva EP. Genomic amplification of BCR/ABL1 and a region downstream of ABL1 in chronic myeloid leukaemia: a FISH mapping study of CML patients and cell lines. *Mol Cytogenet.* 2010;3:15.
4. Bushman DM, Kaeser GE, Siddoway B, Westra JW, Rivera RR, Rehen SK, et al. Genomic mosaicism with increased amyloid precursor protein ( APP ) gene copy number in single neurons from sporadic Alzheimer's disease brains. *Elife.* 2015;4:e05116.

For Peer Review

Sample	Clinical and demographic features						Pathology stage		Targeted array CGH			SNCA digital PCR in SN	SNCA FISH			
	age at death	onset age	duration	Tremor	Asymmetry	Hallucinations	Braak	McKeith	CER	FC	SN		Nigral neurons	Nigral other	FC	Dye swap
	Parkinson's disease															
PD1	79	48	31	+	+	+	6	N				x	1.25	1.64		
PD2	65	44	21	+	+	+	6	N				x	0.70	0.95		x
PD3	75	44	31	+	+	-	5	L			*	x	0	0.31		
PD4	72	64	8	+	+	-	6	N				x	0	0.89		
PD5	65	56	9	-	-	+	5	L				x	1.19	0.94		
PD6	60	53	7	-	-	-	6	L					0.79	0.29		
PD7	81	51	30	+	+	+	6	N				x	1.27	0		
PD8	56	48	8	-	-	+	6	N				x	0.75	0.91		
PD9	64	44	20	-	-		6	L					0.83	0.93		
PD10	62	55	7	-	+	+	6	N				x	0.71	0.59		
PD11	71	49	22	+	+	+	6	N				x	1.36	1.04		
PD12	75	65	10	+	+	-	5	L				x	0.64	0.28		
PD13	61	56	5	-	-	+	6	N				x	0.67	0.56		
PD14	65	59	6	+	+	-	6	N				x	1.30	0		
PD15	61	39	22	+	+	+	6	N					0.86	0.30		
PD16	72	66	6	+	+	+	6	N	x	x	x	x	0	1.24		
PD17	61	59	2	+	+	-	6	N		x	x	x	0.65	0.65		x
PD18	63	54	9			+	6	N	x	x	x	x	0	0.31		
PD19	69	65	4	+	+	+	6	N	x	x	x	x	0.74	0.30		
PD20	75	50	25	+	+	+	6	N	x	x	x	x	1.46	0		
PD21	78	68	10	+	+	+	6	N	x			x	0	0.29		
PD22	74	64	10	+	+	+	6	N	x	x	x	x	1.99	0.32	0	
PD23	76	69	7	+	+	+	6	N	x		x	x	0.66	0		
PD24	76	52	24	+	+	-	6	N	x			x	0.70	0		
PD25	73	67	6	+		+	6	L	x	x	x	x	0	1.25		
PD26	76	71	5	-	-	+	6	N	x		x		0.68	0.90		

PD27	69	50	19	+	+	-	6	N	x		x	x	1.83	0.32		
PD28	71	64	7	-	-	-	6	N	x	x	x	x	1.27	1.18		
PD29	72	64	8	-	-	+	6	N		x		x	1.29	0.32		
PD30	71	61	10	+	+	+	6	N	x	x	x	x	0.73	0.31		
PD31	68	61	7	-	+	-	6	N	x	x	x	x	0.69	0.97		
PD32	78	68	10	+	+	+	6	N		x	x	x	1.32	0.64		
PD33	76	68	8	+	+	+	6	N	x	x	x	x	0	0.33		
PD34	78	72	6	+	+	-	6	N		x	x	x	0	0	0.65	
PD35	76	71	5	-	-	+	6	N	x	x	x	x	0.66	0	1.26	
PD36	68	54	14	-	-	+	6	N	x	x	x	x	0.74	0		
PD37	78	72	6	-	+		6	N	x	x	x					
PD38	70	59	11	+	+	+	6	N	x	x	x	x	0.63	0.98		
PD39	69	49	20	-	+	+	6	N	x	x	x	x	1.90	0	0	x
PD40	83	53	30	+	+	-	5	L	x		x	x	0	0		
PD41	66	53	13	+	+	-	6	N	x	x	x	x	0.90	0		

**MSAMSA**

MSA1	55	49	6										2.76	1.03	0	x
MSA2	68	65	3									x	2.48	0.30		
MSA3	71	66	5										0.70	0.63		
MSA4	76	73	3										0.66	1.56		x
MSA5	62	56	6									x	0	0.35		

**Controls**

C1	94											x	1.36	0		
C2	89												0	0		
C3	69											x	0	0.33		
C4	82											x	1.87	0		
C5	84												0	0		
C6	87											x	0	0.57		
C7	80												1.11	0.86		
C8	87								x		x	x				
C9	86										x		0	0.31		



C10	92			x		x		0	0.25		
C11	93							1.41	1.25		
C12	87							1.32	0		
C13	87			x		x					
C14	76					x	x	0	0		
C15	92			x	x		x				
C16	78			x	x	x	x				
C17	89					x	x	0.63	0.32		
C18	84					x	x	0	0		
C19	82			x	x						
C20	90			x	x		x	0	0		
C21	89					x	x	0.78	0.30		
C22	89						x	0	0		
C23	58							0.73	0.66		
C24	75						x	0	0		
C25	69						x	0	0		
C26	63						x	1.36	0.34		
C27	71						x	0	0.97		
C28	76						x	0.72	0.67		
C29	76							0	0.33		
C30	59							0	0		

**Supplementary table 1. Summary of brains used, experiments performed on each, and FISH results.**

The age of death is given for all, and the age of disease onset and disease duration for Parkinson's and MSA. The presence or absence of tremor and asymmetry at onset, and hallucinations at any stage, in Parkinson's disease cases, are denoted by "+" or "-" (where data available). The pathological stage is given according to Braak and McKeith (N=neocortical, L=limbic). Pathology information on the MSA cases is in Supplementary table 5.

Case MSA2 had a diagnosis of essential tremor, responsive to alcohol and propranolol, from age 15.

For array CGH, brain regions are listed (CER= cerebellum, FC= frontal cortex, SN=substantia nigra). The \* symbol indicates that this was performed later in the study (see "results"). "x" indicates that an experiment was done, and a blank box that it was not. For FISH, the % of nuclei with unique *SNCA* gains is given. The nigra results are divided into neuromelanin-positive and neuromelanin-negative. For FC FISH, the results for Parkinson's disease refer to neurons (positive for NeuN). Case MSA1 also underwent FISH of occipital cortex, putamen, cerebellum, and of nigra combined with chr4 paint. The "dye-swap" column indicates nigra samples also analysed with a green *SNCA* probe and red reference(s) (see Supplementary table 6).

Status	Brain regions analysed				
	All three	CER, SN	CER, FC	FC, SN	One only
PD	16	3	-	3	4 (3 FC, 1 CER)
Control	3	1	3	-	5 (SN)

**Supplementary Table 2. Samples which underwent successful aCGH analysis.**

These consisted of up to 3 brain regions of 26 PD cases and 12 controls: cerebellum (CER: n=30; 22 PD, 8 control), frontal cortex (FC: n=27; 20 PD, 7 control), and substantia nigra (SN: n=32; 22 PD, 10 control).

Diagnosis	Cell type	% cells with any loss	% cells with gains of both
Parkinson's	Neuromelanin-positive	0.58	0.12
	Neuromelanin-negative	0.15	0.18
Control	Neuromelanin-positive	0.43	0.11
	Neuromelanin-negative	0.09	0.09
MSA	Neuromelanin-positive	0.67	0.40
	Neuromelanin-negative	0.13	0.32

**Supplementary Table 3. Percentage of cells with any losses (SNCA or reference), or gains of both probes.** The MSA figures are included for completeness, although they were not compared statistically.

	Pathological stage				Clinical features					
	McKeith		Braak		Tremor		Asymmetry		Hallucinations	
	Limbic	Neocortical	5	6	Present	Absent	Present	Absent	Present	Absent
<b>Gains</b>	4	27	2	29	18	13	21	10	21	9
<b>No gains</b>	3	6	2	7	8	0	7	0	5	4
Total	7	33	4	36	26	13	28	10	26	13
% with gains	57.1	81.8	50.0	80.6	69.2	100	75	100	80.8	69.2

**Supplementary Table 4. Summary of pathological and clinical features of Parkinson's disease cases with and without *SNCA* gains in dopaminergic neurons.** Individual details are given in Supplementary Table 1.

Case	Pathological type	GCI	NCI	Neuronal loss
MSA1	SND	3	2	3 ventrolateral, 2 elsewhere
MSA2	mixed	3	1	3 ventrolateral, 2 elsewhere
MSA3	mixed	2	1	2
MSA5	SND	3	2	3 medial, 2 elsewhere

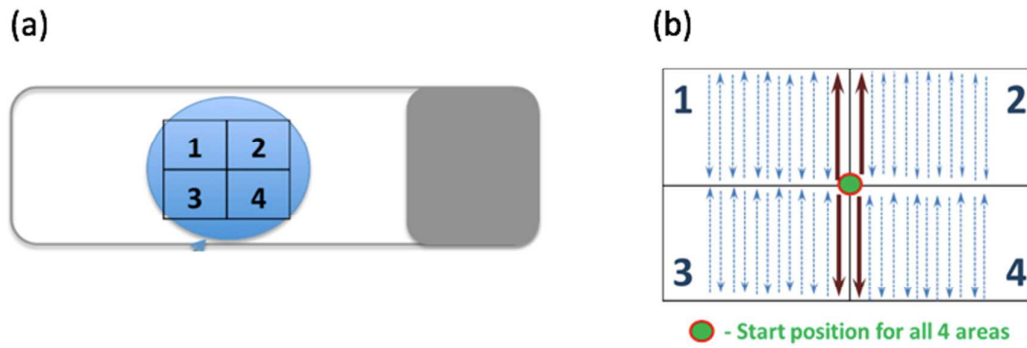
**Supplementary Table 5. Pathological features in substantia nigra of four MSA cases studied by FISH, reviewed blindly.** SND= striatonigral. GCI= glial cytoplasmic inclusions, NCI= neuronal cytoplasmic inclusions. One case was not available. All features graded 1 (mild) to 3 (severe). The cases with the highest proportions of dopaminergic cells with *SNCA* gains (1 and 2) had the highest ventrolateral neuronal loss, but no clear correlations could be made. Slides from case MSA4 were not available.

	<b>Red probe</b>	<b>Green probe</b>	<b>% of NM-positive cells with gains</b>		<b>% of NM-negative cells with gains</b>	
<b>Experiment 1</b>			<b>SNCA</b>	<b>Reference</b>	<b>SNCA</b>	<b>Reference</b>
<b>PD17</b>	<b>PDGFRA</b>	<b>SNCA</b>	<b>0.65</b>	0	<b>0.32</b>	0.00
	<b>IK2F1</b>	<b>SNCA</b>	0	0	<b>0.32</b>	<b>0.63</b>
	<b>SNCA</b>	<b>FIPL1</b>	<b>0.65</b>	0	<b>0.65</b>	0
<b>Experiment 2</b>						
<b>MSA1</b>	<b>PDGFRA</b>	<b>SNCA</b>	<b>0.76</b>	0.00	<b>0.93</b>	0.00
	<b>IK2F1</b>	<b>SNCA</b>	<b>1.32</b>	0.00	<b>0.33</b>	0.00
<b>MSA3</b>	<b>SNCA</b>	<b>FIPL1</b>	<b>0.70</b>	<b>0.70</b>	<b>0.63</b>	<b>0.32</b>
<b>MSA4</b>	<b>SNCA</b>	<b>FIPL1</b>	<b>0.66</b>	0.00	<b>1.25</b>	0.00
	<b>PDGFRA</b>	<b>SNCA</b>	<b>0.64</b>	0.00	<b>0.96</b>	0.00
<b>Experiment 3</b>						
<b>PD39</b>	<b>IKZF1</b>	<b>SNCA</b>	0	0	<b>0.31</b>	<b>0</b>
<b>PD2</b>	<b>IKZF1</b>	<b>SNCA</b>	<b>1.31</b>	0	<b>0.31</b>	<b>0.31</b>

**Supplementary Table 6. Full details of dye-swap FISH experiments.** The probes were blinded in experiments 1 and 2. All samples were from nigra, and the slides were blinded. Slides from experiment 2 are shown in Fig. 5 (A – MSA1, B- MSA 4). NM= neuromelanin. Note that the results for the NM-positive gains with red *SNCA* are also given in Supplementary table 1, but are included here as the experiments were performed together.

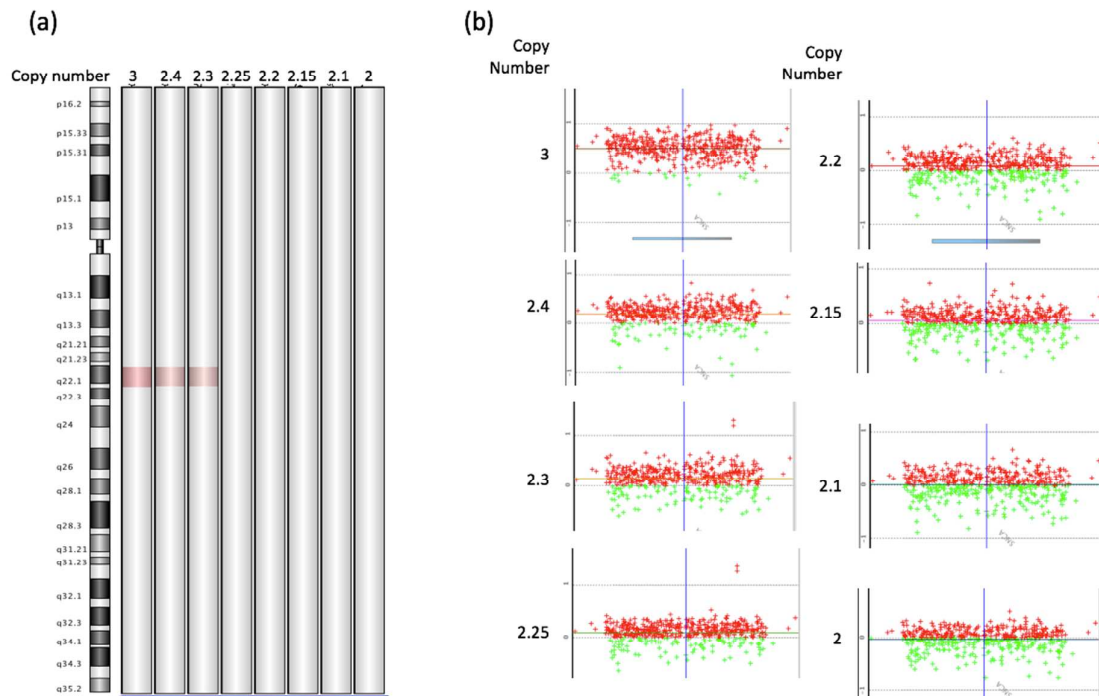
Probe target and colour	chr	Neuromelanin-positive			Neuromelanin-negative		
		Total counted	Cells with gains	% gains	Total counted	Cells with gains	% gains
<b>FIPL1</b>	4	9998	10	0.100	21696	10	0.046
<b>PDGFRA</b>	4	440	0	0	962	0	0
<b>IKZF1</b>	7	563	0	0	1282	3	0.234

**Supplementary Table 7. Gains of all reference probes in substantia nigra.** These are cells which had 2 copies of *SNCA*, and 3 of the reference, regardless of the colour used.



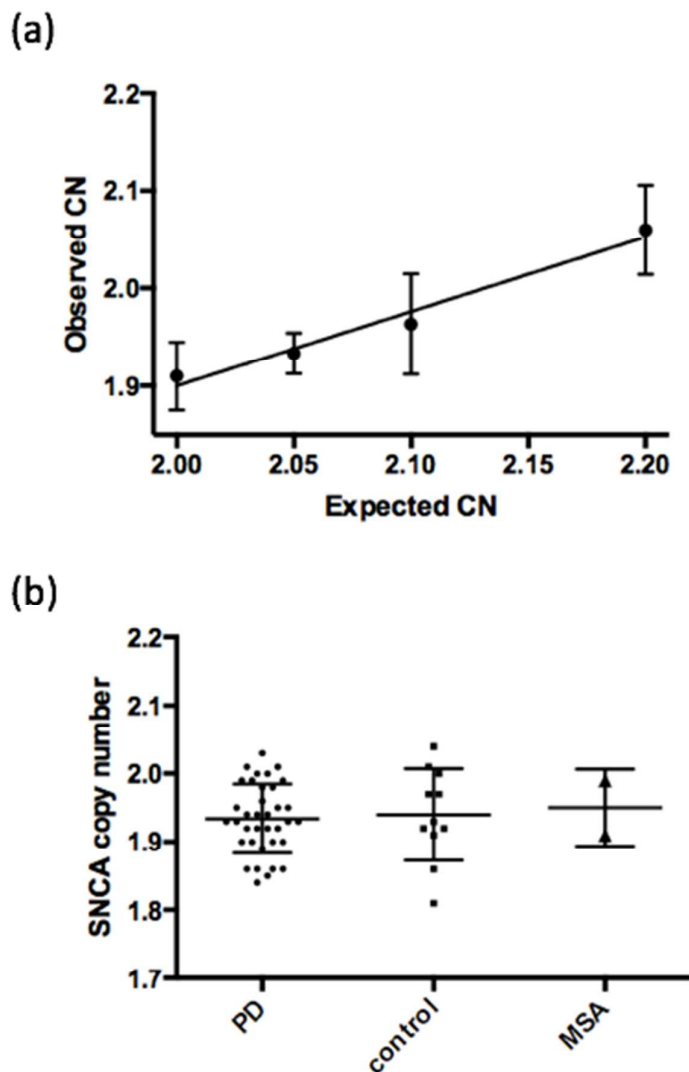
**Supplementary Figure 1. Counting scheme for FISH of substantia nigra.**

The slide was mentally divided in quadrants (a). Provided neuromelanin-positive cells were seen in a quadrant, at least 30 were counted in it by scanning fields in the same direction (b), aiming for a total of 150 per slide. At least 50 of the more numerous neuromelanin-negative cells from the same fields were scored from each quadrant, aiming for 300 per slide in total. Pictures of at least 5 normal cells of each type from each slide, and all abnormal cells, were taken, in multiple focal planes if needed to capture all signals. These were reviewed by the PI, but the copy number was determined by counting on the microscope to avoid any bias. No processing of images was done, other than cropping.



**Supplementary Figure 2. Evaluation of sensitivity of array-CGH design for *SNCA* CNV mosaicism.** “Artificial mosaics” were created to simulate mosaic (non-integer) copy number states by mixing DNA with 3 *SNCA* copies (heterozygous duplication) and sex-matched Agilent reference DNA. The result for each copy number tested is shown, along with the duplication and reference sample. Note that the results of the duplication sample only (copy number 3) were included in a previous publication where this was characterized (Kara et al, 2014), but are included here for completeness.

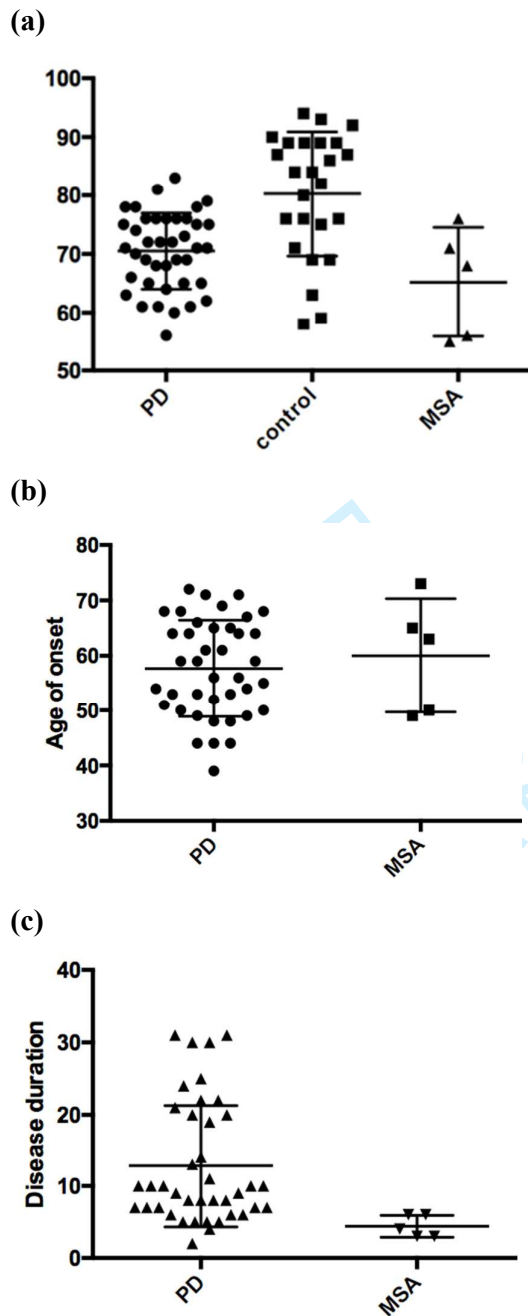
- (a) Output of Agilent Genomic Workbench 7.0, showing chromosome 4, with region of gain in red where detected. The colour intensity corresponds to the copy number.
- (b) Derivative Log Ratio of individual probes over and adjacent to *SNCA*, with moving average line shown. In the 2.25 and 2.2 copy number, visual inspection suggests a possible gain, although below the threshold. We did not, however, see even such borderline results in by careful visual inspection of all our aCGH brain data.



**Supplementary Fig. 3. Droplet digital PCR calculation of *SNCA* copy number (CN). Mean and 95% CI shown.**

- (a) Linear regression of observed *SNCA* copy number (CN) versus expected in control (CN 2.00) and artificial mosaics. Experiments done in triplicates.  $R^2=0.73$ .
- (b) *SNCA* copy number determined by ddPCR in DNA from nigra. T-test (Parkinson's disease v control)  $p=0.759$

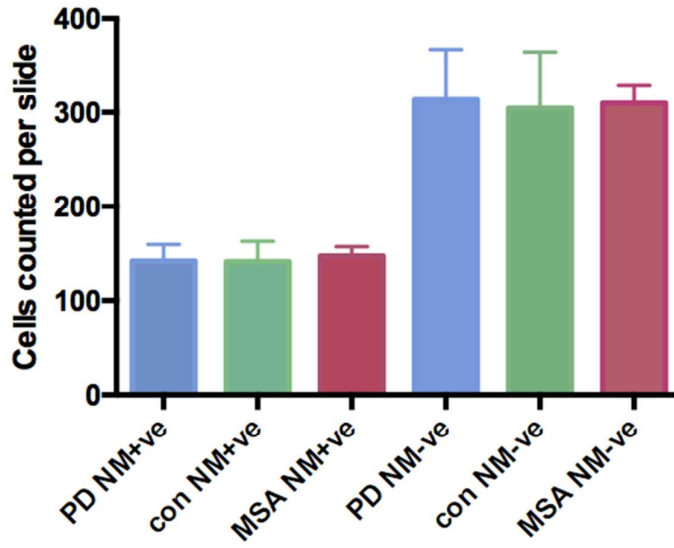




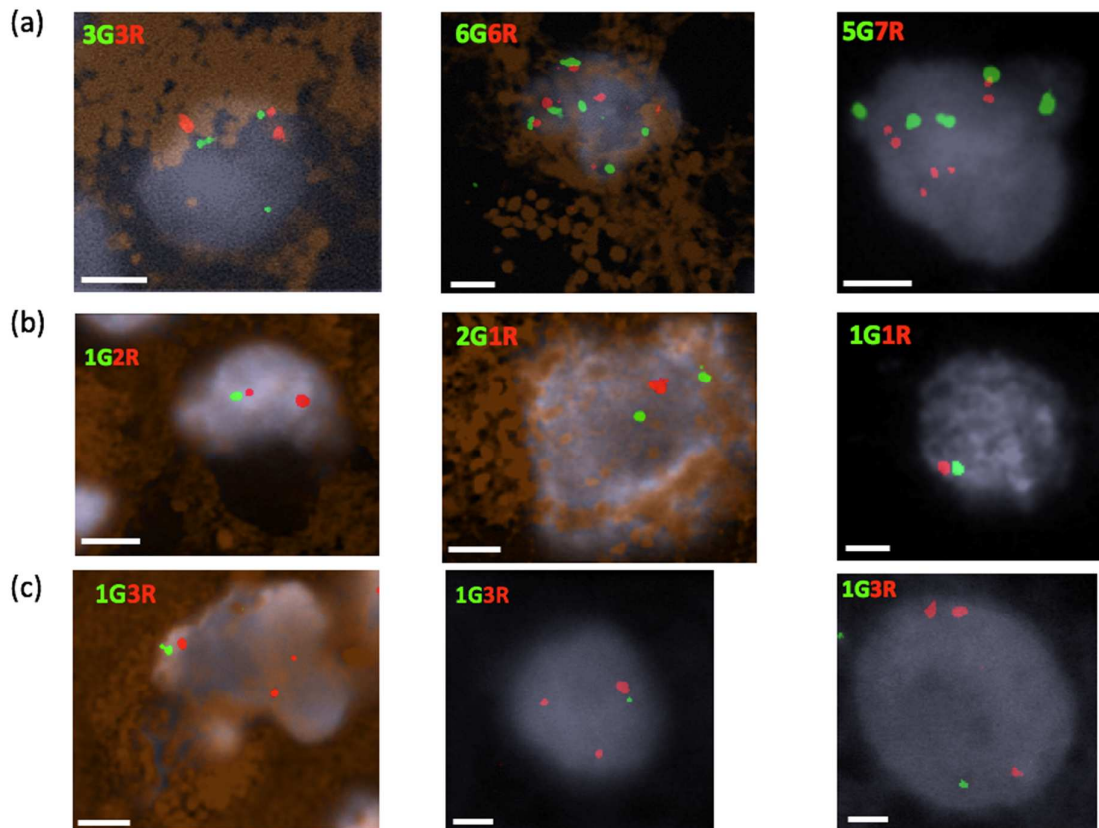
**Supplementary Figure 4. Demographics of all individuals with FISH results from nigra.**

Mean and SD shown for each group. These are a subset of the total samples used in the project (Supplementary Table 1). All but three of the Parkinson's disease (PD) nigras, and around half of the others, had been analysed by array CGH and / or ddPCR (Supplementary Table 1).

- (a) Ages of death of all.
- (b) Disease onset age for Parkinson's and MSA.
- (c) Disease duration for Parkinson's and MSA.



**Supplementary Figure 5. Cells of each type counted.** Mean and SD shown. NM+ve = neuromelanin-positive, NM-ve = neuromelanin-negative, con=control.



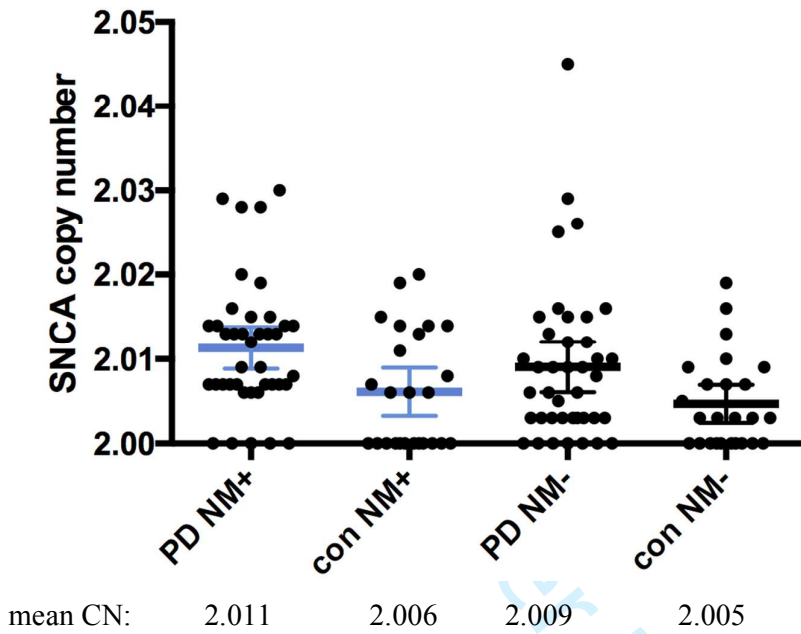
**Supplementary Figure 6. Cells with gains of both probes, or any losses.** Scale bar = 3  $\mu$ m.

(a) Cells with multiple copies of both probes (two neuromelanin-positive and one neuromelanin-negative). Note that the third cell has an excess of *SNCA* signals, suggesting that a simple sectioning artefact with partial overlying nuclei is unlikely to account for this appearance. Cells with 3 or more copies of both probes usually had the same number of copies of *SNCA* and reference. This could be due to more than one nucleus, but a few had excess *SNCA* copies (3/16 neuromelanin-positive, 5/34 neuromelanin-negative), while the opposite was never seen.

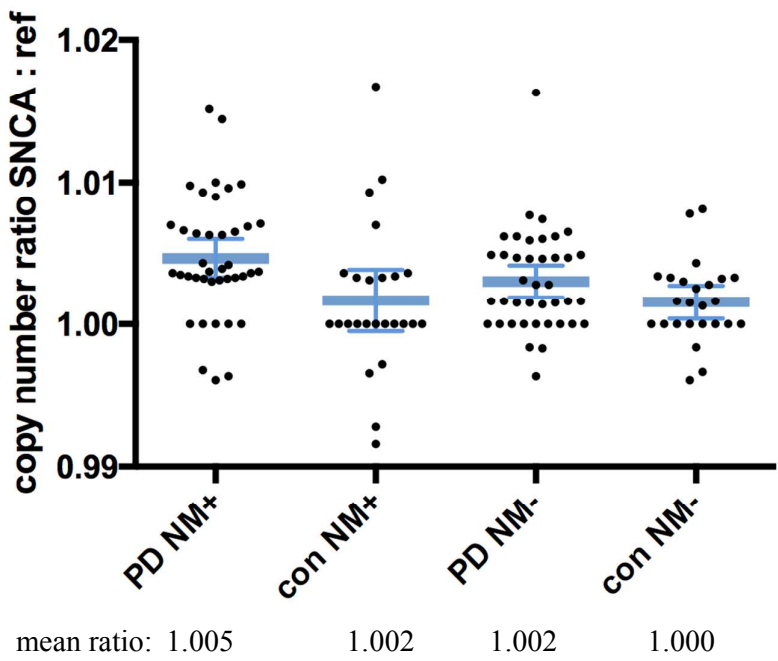
(b) Cells with loss of a copy of reference, *SNCA*, or both (two neuromelanin-positive and one neuromelanin-negative). In neuromelanin-positive cells, 18 had lost *SNCA*, and 11 the reference ( $p=0.265$ ). In neuromelanin-negative cells, the numbers were 7 and 6, respectively.

(c) Cells with 3 *SNCA* copies and one reference, from the same PD nigra (case PD35; one neuromelanin-positive, two neuromelanin-negative, with apparently large nuclei). This case had one of the shortest disease durations (5 years). There were only two other neuromelanin-positive cells with this pattern in our entire study, in different PD cases who also had very short disease duration (5 and 6 years)

A



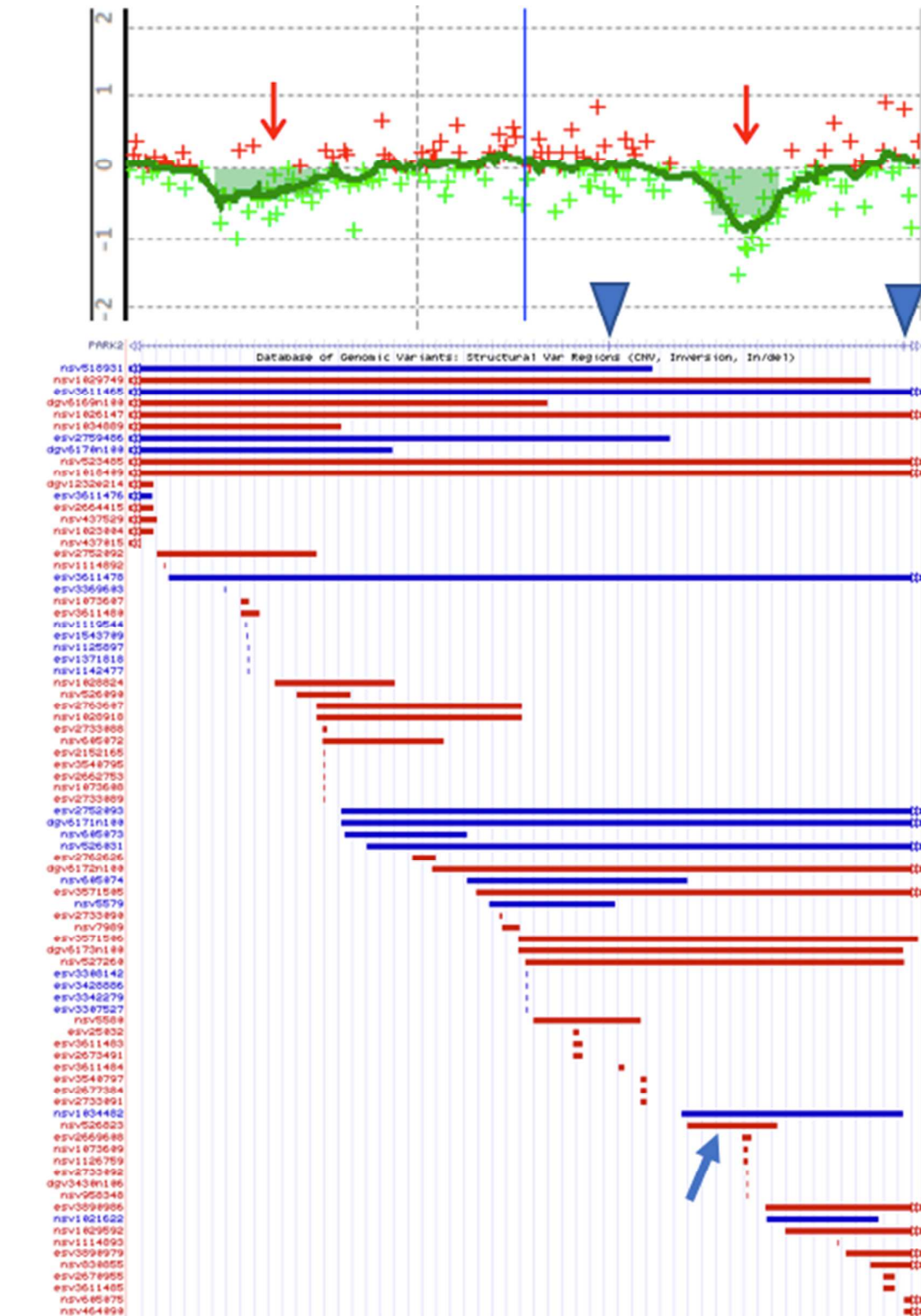
B



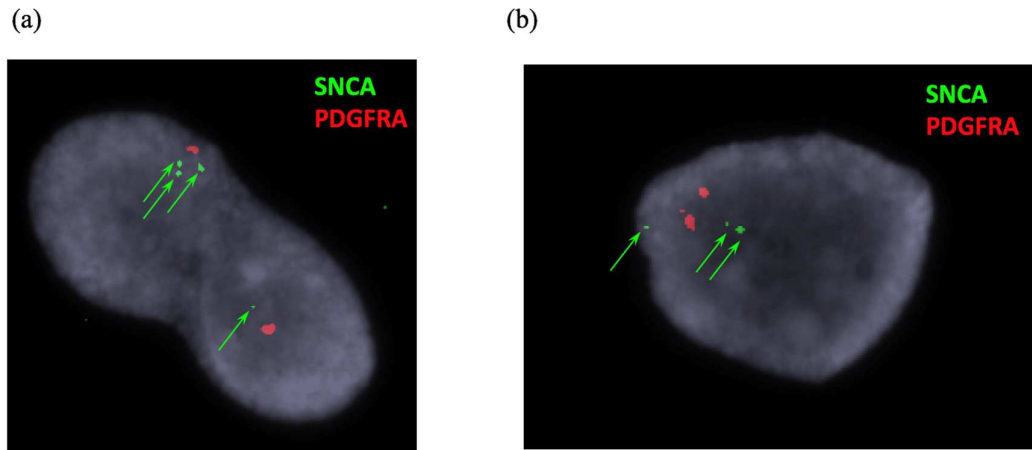
**Supplementary Figure 7. Additional analyses of Parkinson's and control in each nigra / cell type.** These were performed as discussed in the text to determine whether the clear excess of gains in neuromelanin-positive (NM+ve) cells in Parkinson's disease (PD) relative to controls (con) is still present if cells with gains of both probes, or losses of any, are also included. The mean and 95% confidence intervals are shown for each.

**(A) Mean *SNCA* copy number (CN).** This was calculated including all cells which have no losses, including any with gains of reference as well. The total number of *SNCA* signals in each nucleus was added up for each nigra / cell type. This measurement was highly correlated with the main outcome measurement, the percentage of cells with *SNCA* gains and normal reference on which figure 4 is based, as expected, for each sample ( $r=0.73$  for neuromelanin-positive,  $0.76$  for neuromelanin-negative;  $p<0.0001$  for both). Analysis by Kruskal-Wallis ANOVA showed a highly significant difference overall ( $p=0.002$ ), with significant pairwise difference between PD and control for neuromelanin-positive, but not neuromelanin-negative, cells ( $p=0.01$  and  $0.0834$  respectively after correction).

**(B) Mean ratio of *SNCA*: reference probe number.** These results also showed significant overall difference (Kruskal-Wallis ANOVA  $p=0.0017$ ), with pairwise comparisons of Parkinson' disease and control showing once again significant difference for neuromelanin-positive, but not neuromelanin-negative, cells (corrected  $p=0.0029$  and  $0.2193$ , respectively).

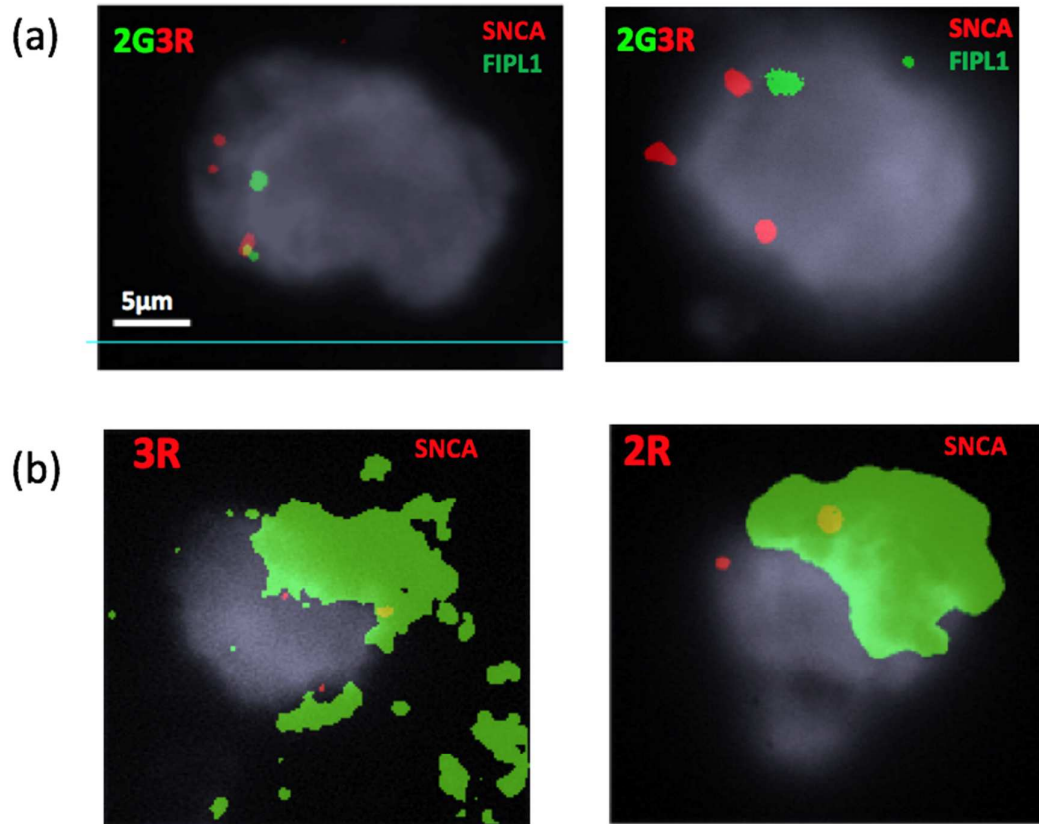


**Supplementary Figure 8. Array CGH analysis of *PARK2* variant carrier shows at least one intronic deletion.** The derivative log ratios for all probes over a 217kb region are shown, together with the moving average line (chr6: 162,263,234-162,480,518). Arrowheads point to coding exons in this region. Red arrows indicating deletions called by the software (shaded green regions). Known copy number variants in the region are shown below (red, loss; blue, gain) from the DGV track of the UCSC genome browser. The right loss appears definite, and corresponds to a known 24.5 kb copy number variant (nsv526823; blue arrow). The left loss is novel, but may be a false positive, as not all probes within it indicate a loss.



**Supplementary Figure 9. Validation of *SNCA* green probe on fibroblasts with triplication.**

Arrows point to the signals, in a cell where the copies were clearly differentiated (a), and one where they were not (b). The signals were less bright than the *SNCA* red probe, despite this being twice as large.



**Supplementary Figure 10. FISH in other brain regions.**

**(a) MSA1.** A cell with *SNCA* gain in putamen (left) and cerebellum (right)

**(b) Frontal cortical neurons** (identified by green NeuN stain) with a gain (left) and normal (right).