

# Somatic mutations in Parkinson disease are enriched in synaptic and neuronal processes

Irene Lobon<sup>1</sup>, Manuel Solís-Moruno<sup>1,7</sup>, David Juan<sup>1</sup>, Ashraf Muhaisen<sup>2,3</sup>, Federico Abascal<sup>4</sup>, Paula Esteller-Cucala<sup>1</sup>, Raquel García-Pérez<sup>1</sup>, Maria Josep Martí<sup>3,5</sup>, Eduardo Tolosa<sup>3,5</sup>, Jesús Ávila<sup>3,6</sup>, Raheleh Rahbari<sup>4</sup>, Ferran Casals<sup>7</sup>, Tomas Marques-Bonet<sup>1,8,9,10</sup>, Eduardo Soriano<sup>2,3</sup>

1 Institute of Evolutionary Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003 Barcelona, Spain.

2 Department of Cell Biology, Physiology and Immunology, and Institute of Neurosciences, University of Barcelona, 08028, Barcelona, Spain

3 Centre for Networked Biomedical Research on Neurodegenerative Diseases (CIBERNED), 28031, Madrid, Spain

4 Cancer, Ageing and Somatic Mutation (CASM), Wellcome Sanger Institute, Cambridge, UK

5 Department of Neurology, Hospital Clínic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona (UB), Barcelona 08036, Spain

6 Centro de Biología Molecular Severo Ochoa, 28049, Madrid, Spain

7 Genomics Core Facility, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, 08003, Barcelona, Spain

8 Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010, Barcelona, Spain

9 CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldori i Reixac 4, 08028 Barcelona, Spain

10 Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain

## Abstract

The role of somatic mutations in complex diseases, including neurodevelopmental and neurodegenerative disorders, is becoming increasingly clear. To explore their relevance in sporadic Parkinson disease, we performed whole-exome sequencing in blood and four brain regions of ten patients. We identified 59 candidate somatic single nucleotide variants (sSNVs) through sensitive calling and extensive filtering. We validated 27 of them with amplicon-based deep sequencing, with a 70% validation rate for the highest-confidence variants. Most of the sSNVs were exclusively called in blood but were also found in the brain tissues with the ultra-deep amplicon sequencing, demonstrating the strength of multi-tissue sampling designs. We could confirm between 0 and 6 sSNVs per patient and generally those with a shorter lifespan carried more variants. Remarkably, the validated sSNVs are enriched in genes with synaptic functions that are co-expressed with genes previously associated with Parkinson disease.

## Introduction

Somatic mutations appear during development and tissue maintenance, making every individual a mosaic of cells with slightly different genomes. Early mutations occurring before gastrulation are shared by tissues of different germ layer origin (Bae et al. 2018; Lodato et al. 2015) and can cause disease (Mensa-Vilaró et al. 2019; Kluge et al. 2019). Mutations occurring later in development or

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

the cells carrying them, limited by each tissue's dynamics (Lee-Six et al. 2019; Moore et al. 2020). Notably, this type of somatic mutations are the cause of cancer (Nowell 1976), but have also been found in healthy skin (Martincorena et al. 2015; Abyzov et al. 2017), esophagus (Martincorena et al. 2018), colon (Lee-Six et al. 2019), liver (Brunner et al. 2019), endometrium (Moore et al. 2020) or lung (Yoshida et al. 2020). Further, organ-exclusive somatic mutations in the MTOR pathway are involved not only in cancer (Guertin and Sabatini 2007) but also in neurodevelopmental disorders such as hemimegalencephaly (Poduri et al. 2012) and focal cortical dysplasia (Lim et al. 2015). Other studies linking somatic mutations to disease did not determine whether the variants were of early or late origin, as in a case of congenital arrhythmia where the causal mutation was only tested in cardiomyocytes and lymphocytes and found in both cell types (Priest et al. 2016).

Before birth, each human neuron carries 200 to 400 somatic single nucleotide variants (sSNVs) (Bae et al. 2018) and about 300 to 900 sSNVs can be detected by early infancy (Lodato et al. 2018). Those with a frequency higher than ~2% can also be detected in tissues that originate from a different germ layer, suggesting that brain development does not heavily rely on clonal expansion. Other types of somatic variation have also been found in neurons: retrotransposon mobilization is common and disproportionately impacts protein-coding loci (Baillie et al. 2011) and over three copy number variants (CNVs), mainly losses, can be found in each cell (Cai et al. 2014). Strikingly, while the number of CNVs decreases in brain with age (Chronister et al. 2019), sSNV load increases (Hoang et al. 2016; Lodato et al. 2018) posing questions about their relevance in neuronal diversification, plasticity and dysfunction.

The somatic variant hypothesis for neurodegenerative diseases states that unexplained sporadic cases could be caused by somatic mutations, presumably in the same genes affected in familial cases (Pamphlett 2004). Supporting this theory, age-related sSNVs accumulate faster in neurodegeneration (Lodato et al. 2018). However, not only these later-acquired mutations seem relevant, but also earlier somatic mutations may contribute to phenotypes and diseases. For example, individuals with Autism spectrum disorder have a higher burden of somatic mutations than their unaffected siblings, measured in blood (Dou et al. 2017). In Alzheimer, targeted sequencing of blood samples showed that somatic variants in autosomal dominant genes (such as *APP*) can explain ~2% of cases (Nicolas et al. 2018). Another study on blood and hippocampus exomes from 52 Alzheimer patients showed that over a fourth carried somatic mutations affecting pathways known to contribute to tau hyperphosphorylation (Park et al. 2019), demonstrating the power of exome analysis.

The link between Parkinson disease (PD) and somatic mutations is not as clear. Notably, only about 10% of cases can be attributed to monogenic forms (Lesage and Brice 2009) and at most 30% of patients have an affected first-degree relative (Rocca et al. 2004). A study on 511 sporadic cases tested multiple brain regions with a sensitivity limit at 5% of variant allele frequency (VAF) and did not find any somatic variants in *SNCA*, the main causative gene in early onset familial PD (Proukakis et al. 2014). On the other hand, patients showed high levels of heteroplasmic mitochondrial DNA deletions (Bender et al. 2006) and more *SNCA* somatic copy number gains in substantia nigra neurons compared to controls, which positively correlated with age of onset (Mokretar et al. 2018).

To explore the potential link between somatic SNVs and Parkinson disease, we sequenced the exome of five different tissues from ten sporadic Parkinson patients – substantia nigra, striatum, neocortex, cerebellum and blood – at an average coverage of 60X. We developed and implemented a filtering

approach based on both single tissue and joint information which we used to identify 59 candidate sSNVs. Further amplicon-based deep sequencing confirmed 27 of them, with an average of 3 sSNVs per individual (range 0-6). The confirmed sSNVs were enriched in synaptic and axonal processes and patients with more sSNVs tended to die earlier, suggesting a potential role of these variants in the disease. Interestingly, over 76% of the variants validated in multiple brain tissues were only called in blood in the exome data, demonstrating that studying other more accessible and unaffected tissues may well serve for the identification of variants with lower frequency in diseased organs.

## Results

### Dataset

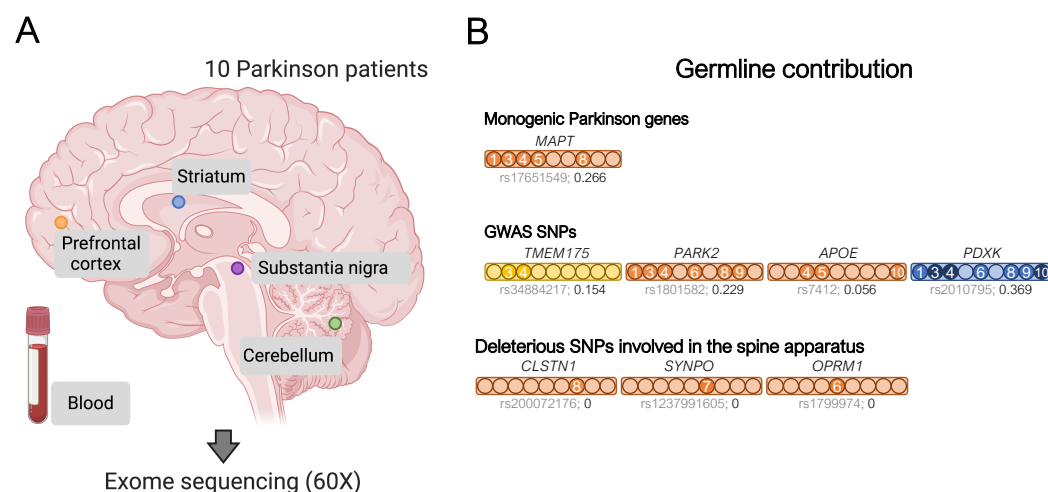
We sequenced the whole exome of five different tissues from ten sporadic Parkinson patients at an average coverage of 60X. Blood was obtained from stored vials while the cerebellum, neocortex, substantia nigra and striatum samples were collected during autopsies (Figure 1A). Individuals' median age was 81 at the time of death, they had varying ages at disease onset and both sexes were represented (Supplementary table 1). As expected, germline SNV calling showed a good clustering of tissue samples by individual except for two samples from individual DV2, which appear to be contaminated (Supplementary figure 1). Accordingly, this sample was excluded from subsequent analyses but used for noise profiling.

### Germline variants are associated with the disease

The subject individuals were categorized as sporadic Parkinson cases because no affected first-degree relatives were known. This assessment did not include genetic analyses, so we evaluated their germline variants before considering their somatic mutational landscape. We used three different strategies to prioritize germline single nucleotide changes. First, we identified germline SNVs in genes linked to Parkinson in OMIM ([www.omim.org](http://www.omim.org)) that were deleterious as indicated by a CADD score > 15 (Rentzsch et al. 2019). Only one variant met these criteria, rs17651549, a missense mutation in the *MAPT* gene, which encodes the tau protein, predicted as deleterious by multiple methods and at a highly conserved position in vertebrates (Supplementary table 2). This variant has been previously linked to PD by different means: multivariate family-based association tests (Wang, Mullersman, and Liu 2010), pathway analysis (Song and Lee 2013) and targeted resequencing (Spataro et al. 2015). However, in a contradictory haplotype association analysis it provided a reduced risk for PD (J. Li et al. 2018). This SNP is not infrequent in Europe; particularly, the frequency of the alternative allele in the 1000GP IBS population (Iberian populations in Spain) is 0.27 and similarly, it was 0.28 in our Spanish cohort (Figure 1B).

Then, we made use of GWAS studies, with greater statistical power, to recognize Parkinson-associated germline variants (41 variants, Supplementary table 3) (Tan et al. 2010; Do et al. 2011; International Parkinson Disease Genomics Consortium 2011; International Parkinson Disease Genomics Consortium and Wellcome Trust Case Control Consortium 2 2011; Lill et al. 2012; Nalls et al. 2014a; Chen et al. 2016) and found four in our cohort. DV3 and DV4 carried a splice acceptor variant at *TMEM175*, rs34884217. This variant, predicted to affect nonsense-mediated decay, has been associated with Parkinson (Nalls et al. 2014; Heckman et al. 2017) and *TMEM175* deficiency is linked to the increase of  $\alpha$ -synuclein aggregation (Jinn et al. 2017), indicating a possible causal

link. The missense mutation of *PARK2* showed the highest odds ratio for Parkinson disease in Europeans in a meta-analysis (Ramakrishnan et al. 2016). The other two SNVs have been associated with the disease (Huang, Chen, and Poole 2004; Williams-Gray et al. 2009; Elstner et al. 2009), but conflicting results have also been reported (Federoff et al. 2012; Guella et al. 2010).



**Figure 1. A. Experimental design.** DNA was obtained from blood and four brain regions from ten sporadic Parkinson disease patients. Each sample's exome was captured and sequenced. **B. Germline SNVs related to Parkinson.** A bar is shown for each germline variant and carrier individuals are indicated by filled circles (DV2 is not included). Darker circles indicate a homozygous SNV. The affected gene is indicated on top and the dbSNP identifier and the frequency of the alternative allele in the IBS population are shown below each bar. Bar color indicates the type of variant: missense (red), splice acceptor (yellow) or intronic (blue).

In addition, all germline SNVs were filtered by deleteriousness (CADD > 15 and SIFT prediction) and frequency in the 1000GP European population (< 0.1). The resulting 214 variants affected 207 genes, with each patient carrying a median of 27 variants (range 20-32). An overrepresentation enrichment analysis (Zhang, Kirov, and Snoddy 2005) found 7 molecular function and cellular component terms significantly enriched in these genes (FDR ≤ 0.05, Supplementary figure 2). Remarkably, the term with the highest enrichment ratio was *spine apparatus*, a derivate of the smooth endoplasmic reticulum generally present in dendritic spines that seems to participate in spine remodeling in Parkinson disease models (Smith, Villalba, and Raju 2009). *Kinesin binding* and *motor activity* were also among the significantly enriched terms, with 7 out of the 9 individuals carrying a deleterious mutation in genes driving these 3 associations. These genes have been previously linked to PD in different studies. As an example, *CLSTN1* overlapped a significantly hypomethylated CpG (Chuang et al. 2017), was differentially expressed (Kong et al. 2018) and carried a missense mutation (Yemni et al. 2019) in PD cases. Several dynein and kinesin proteins also appear to be relevant in PD. All of this highlights the complexity of Parkinson disease, where common and rare variants affect multiple pathways that seem to contribute to the phenotype. Abundant data is needed to uncover these associations, so the evaluation of somatic mutations has the potential to contribute to this effort.

## Somatic variant calling

Somatic variant calling was first developed for cancer and standard approaches are based on paired tumor and normal samples comparison (Krøigård et al. 2016; Xu 2018). Growing evidence shows that low frequency somatic mutations are present in multiple tissues, even from different germ layers (Bae et al. 2018; Lodato et al. 2015), so paired comparisons are inadequate in scenarios with no or limited clonal expansion. Requiring two supporting reads in our 60X data would imply a mean VAF of

over 3%. These variants could have originated before gastrulation, be present in multiple tissues and require joint sample calling. On the other hand, blood clonal expansion, especially prevalent in old age (Jaiswal et al. 2014), will result in variants private to this tissue, requiring single sample calling. In standard paired samples calling, shared calls are essentially discarded, which filters out germline variants and recurrent artefacts. The main obstacle unpaired single or joint sample calling have is precisely differentiating these confounding factors from true somatic variants.

Germline callers are optimized to discard variants that do not fit with set VAF expectations. Low frequency variants, such as those we are interested in, are considered sequencing noise or contamination and discarded. For this reason, we called variants in all our samples using HaplotypeCaller with `--ploidy 10`, which increases its sensitivity to lower frequency variants. We also used VarScan 2 with lax parameters (`--min-coverage 1 --min-reads2 1 --p-value 1 --min-var-freq 0.000001`). The resulting call set is mostly composed of germline heterozygous SNVs, calls within CNVs, recurrent sequencing errors and other artefacts. Hence, we developed a filtering strategy to identify high confidence somatic SNVs, COSMOS (Combined Or Single sample MOSaicism detection, <https://github.com/ilobon/COSMOS>). COSMOS can be used to annotate the relevant information needed to filter somatic candidates on standard VCFs as well as to filter calls directly. Filtering can be performed with a single sample approach, a joint filtering approach using multiple samples from the same individual, or both.

## COSMOS

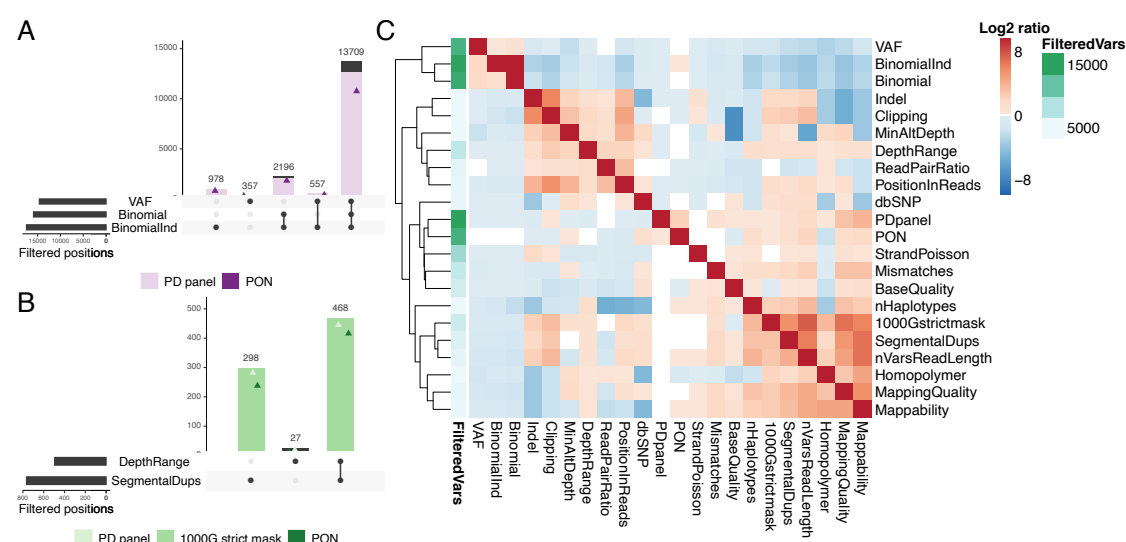
Manual inspection of the calls obtained from HaplotypeCaller and VarScan 2 allowed the identification of multiple sources of artefacts, for which we devised filtering approaches. The main rationale was that even if our patients had the same disease, and could carry mutations in the same pathways or even genes, the probability they would bear mutations in the exact same position was negligible in just 10 individuals. Hence, variants present in multiple individuals are presumably artefacts.

We discarded all off-target calls because they have higher strand imbalances (69.4% of off-target calls fail the Poisson test vs. 39.5% of on-target calls) and read pair imbalances (21.8% vs 4.4% failing the read pair ratio filter). The most frequent confounding factors for on-target calls were germline heterozygous SNVs, CNV regions, unresolved regions of the genome and regions that are more difficult to align (indels, homopolymers).

Germline heterozygous calls supposed at least 93.1% of our on-target calls. They can be easily distinguished when their VAF is close to or higher than 50%. However, germline calls' VAFs are more over dispersed the lower the coverage (Supplementary figure 3) and VAFs as low as 18% can result from germline variants sequenced at 60X (and even 10% for 20X, our lower bound depth). We used a binomial test (83.2% of on-target calls) and a VAF upper limit (76.5% of on-target calls had VAF > 40%) to identify the most obvious cases, with a large overlap between the two sets of calls (Figure 2A). These variants are common to all tissues from the same individual, so when multiple samples are available, requiring that all pass the binomial test greatly increases the detection power, following an exponential distribution (Supplementary figure 4). We found an additional 5.5% of variants with this approach (first bar in Figure 2A). However, an alternative way of flagging most of these calls is by a large number of called individuals. By using the other 45 samples from the remaining 9 PD individuals, which results in a batch and population-specific panel, we can identify 92.1% of them.

Alternatively, an external panel of normals (PON) consisting of 428 whole genomes sequenced at the Sanger Institute tagged 82.6% of them (Figure 2A).

Regions with CNVs can create artefactual somatic variants when reads are collapsed –mapped into a single copy of the reference genome. CNV callers can be used to identify the highest confidence regions, but more sensitive approaches are needed to obtain high-confidence somatic variant calls. We excluded regions with extremely low or high coverage (keeping 75% of the distribution) or overlapping the WGAC segmental duplications track, which filtered out 60.4% of the remaining on-target non-germline calls. 97.5% of these calls overlap the 1000GP strict mask, which identifies regions of the genome with recurrently higher/lower coverage or lower mapping quality, showing its suitability as an alternative. Again, most calls failing our CNV filters were called in multiple individuals (94.1% in our PD panel and 84% in the independent PON) (Figure 2B). Other features useful to identify these artefacts are a high number of variants within read length distance, an imbalance in the number of mismatches in reads carrying each allele or in the proportion of clipped reads and a low mappability (Supplementary figure 5).



**Figure 2. Relevance of COSMOS filters in our Parkinson exome data. A. Germline SNVs.** Intersection of the main filters that identify germline heterozygous point mutations: VAF (high variant allele frequency), Binomial (non-significant binomial test for allele depths) and BinomialInd (non-significant binomial test in at least one sample from the individual). For each intersection, the number of variants also found in the other individuals of our dataset (PD panel, filled in lavender) and an external panel (PON, purple triangle) are shown. **B. CNV filters.** Intersection of DepthRange (most extreme depth variants) and SegmentalDups (segmental duplications WGAC track) with PD panel and PON (light and dark green triangles, respectively) and variants present in the 1000GP strict mask filled in green. **C. Relationship between filters.** Log2 ratio between the number of variants failing both the row and the column filter and those failing the row filter only. Red indicates higher log2 ratios and hence higher co-appearance of filters. The annotation column indicates the total number of positions failing each of the row filters.

Unresolved regions of the genome or highly variable duplications can result in collapsed mapping with no significant increment of coverage because only reads spanning the homologous region and carrying few variants will be mapped. These are easily identified by a biased position of the alternative allele in the reads, clustered at read ends. We defined a position-in-reads bias score (PIR) to address this. Other features targeting these artefacts are allele clipping imbalances, a high number of variants in the vicinity, more than three haplotypes, allele imbalances for mapping quality and read pair imbalances (Figure 2C). Local alignment around indels and homopolymers is challenging and technical and biological noise are difficult to distinguish, so we discarded close-by variants to increase



our call set confidence (Methods). Finally, variants with too few reads supporting the alternative allele cannot be distinguished from random noise and their imbalances cannot be evaluated, so a hard cut-off was applied (minimum of two and three reads supporting the alternative allele for joint and single sample calling respectively).

### **Most somatic validated variants are present in both blood and brain**

A total of 59 variants passed COSMOS filtering, 7 of which were called exclusively by VarScan. After manual inspection, we classified variants in four tiers of confidence (examples in supplementary figures 6 and 7). Tier 1 had 17 high confidence variants, tier 2 consisted of the 7 VarScan calls, tier 3 contained 29 lower confidence candidate sSNVs and tier 4 included 5 putative false positives (Supplementary table 4 and Methods). A median of 7 candidates were called in each individual (range 1-11) and perhaps unsurprisingly given the limited coverage, 91.5% of all the candidate variants were called in just one tissue, mostly blood (61.1% of single tissue calls) (Figure 3A).

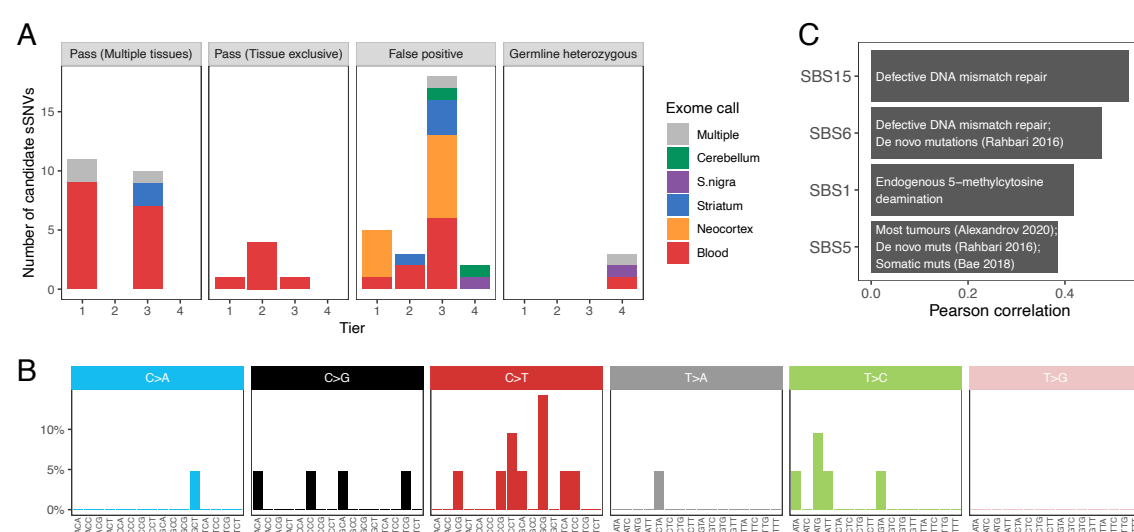
We evaluated 58 out of the 59 variants with the rhAmpSeq targeted amplicon sequencing system, which allows the multiplexed amplification of multiple genomic positions. This way, we sequenced data from all positions and all patients at a mean coverage of over 18,000X so that non-called individuals could be used as negative controls. Four of the samples could not be amplified enough and had low coverages (mean  $\leq 2,300$  X, supplementary figure 8). To consider a variant as a validated sSNV in a given sample, we required that (1) it was the second most common allele in that sample and (2) its VAF was higher than the mean + 2 standard deviations of all other individuals' samples VAFs with sufficient coverage (defined as not lower than the mean - 1 standard deviation). We additionally classified variants passing these filters in multiple tissues of an individual and with VAFs > 30% in all of them as germline heterozygous variants (Supplementary material).

A total of 27 somatic SNVs were validated using these criteria. Validation rates were 70.6% for tier 1 (12 out of 17), 57.1% for tier 2 (4 out of 7), 37.9% for tier 3 (11 out of 29) and 0% for tier 4 (0 out of the 5 negative control variants). Per individual, a median of 4 variants were validated and the median validation rate was 50% (mean of 40.6%, ranging from 0% to 83%) as all calls from 3 individuals (DV4, DV5 and DV9) were false positives. DV9 had just one candidate sSNV call but the other two had 7 and 6, respectively, demonstrating moderate interindividual variability. Only 6 variants were present in a single tissue, which was always blood (Figure 3A, second panel). Interestingly, 76.2% (16 out of 21) of variants validated in multiple tissues had been called exclusively in blood in the exome data (Figure 3A, first panel), as they have a higher frequency in this tissue (mean difference of 10.8%, mean VAF in blood 11.5% and 0.7% in the other tissues). This was not a consequence of our calling method, as 68.8% of these variants (11 out of 16) had no read supporting the alternative allele in any brain sample but were then validated in at least 2 brain regions. This probably results from their random amplification in tissue maintenance of blood but could also be a consequence of depletion in the central nervous system.

### **The brain somatic SNVs spectrum is most similar to the expected mutational signatures**

The discovery that different mutagenic agents –such as UV light, carcinogens or intrinsic cell processes– produce distinct substitution patterns in a context dependent manner led to the development of mutational signature analysis (Alexandrov et al. 2013). To obtain the mutational

spectrum of Parkinson disease, we combined the 21 sSNVs validated in at least one brain sample and classified them by substitution and trinucleotide context (Figure 3B). As the number of variants was insufficient for mutational signature deconvolution, we calculated its Pearson correlation with the COSMIC single base signatures. Since most of the identified variants are present in multiple tissues and are therefore of early origin, we expected to find a high similarity to signatures SBS1 and SBS5. Indeed, we found both among the four moderate correlations (Pearson's  $r > 0.3$ , Figure 3C). SBS1 is an ubiquitous signature that results from the spontaneous deamination of methylated cytosines. SBS5 can be detected not only in most cancer samples (Alexandrov et al. 2020) but also in *de novo* mutations (Rahbari et al. 2016), somatic mutations (Bae et al. 2018) and in population level variants (Mathieson and Reich 2017). Interestingly, SBS6 was also found to be highly correlated with the *de novo* mutational spectrum (Rahbari et al. 2016). The other signatures we found to be similar to the Parkinson spectrum, SBS15, has been proposed to be caused by defective DNA repair.



**Figure 3. A. Validation of candidate variants.** Number of mutations validated (Pass) in multiple or a single tissue, or found to be false positives or germline heterozygous variants with amplicon sequencing data. Variants are distributed by calling-confidence tier and colors indicate the tissue in which the variant was originally called in the exome data. **B. Mutational spectrum of somatic SNVs in Parkinson brains.** Variants present in the brain of Parkinson patients, segregated by substitution and trinucleotide context. **C. Correlation with COSMIC signatures.** Moderate Pearson correlations ( $r > 0.3$ ) between the spectrum of brain somatic variants and the single base signatures (SBS) from COSMIC. Text describes the etiology or studies relevant for each signature.

### Variant allele frequency of sSNV can be used to cluster the tissues

We validated sSNVs in 6 out of the 9 subject individuals, ranging from 1 to 6 variants per patient, a modest yet sensible number for an exome analysis. All of DV10's sSNVs were only detected in blood but the other 21 variants (from DV1, DV3, DV6, DV7 and DV8) were detected in at least two brain tissues. We clustered the tissues based on their VAFs at these sSNVs. Because each individual has just a few variants, we pooled them together to analyze general tissue dynamics (Figure 4A). As expected from its clonal expansion, blood is the most distant tissue, with brain tissues being more closely clustered. Remarkably, striatum and substantia nigra, both affected in Parkinson disease (Bernheimer et al. 1973), cluster together, which could be caused by their closer developmental origin and/or more similar physiology. The presence of 3 variants with higher frequencies in all brain tissues than blood suggests that our findings are not the result of blood contamination in the other tissues. Importantly, blood was not called in the exome sequencing data for these variants, and the amplicon sequencing confirmed the VAF distribution in the different tissues.



## Patients with more deleterious sSNVs in brain tend to die earlier

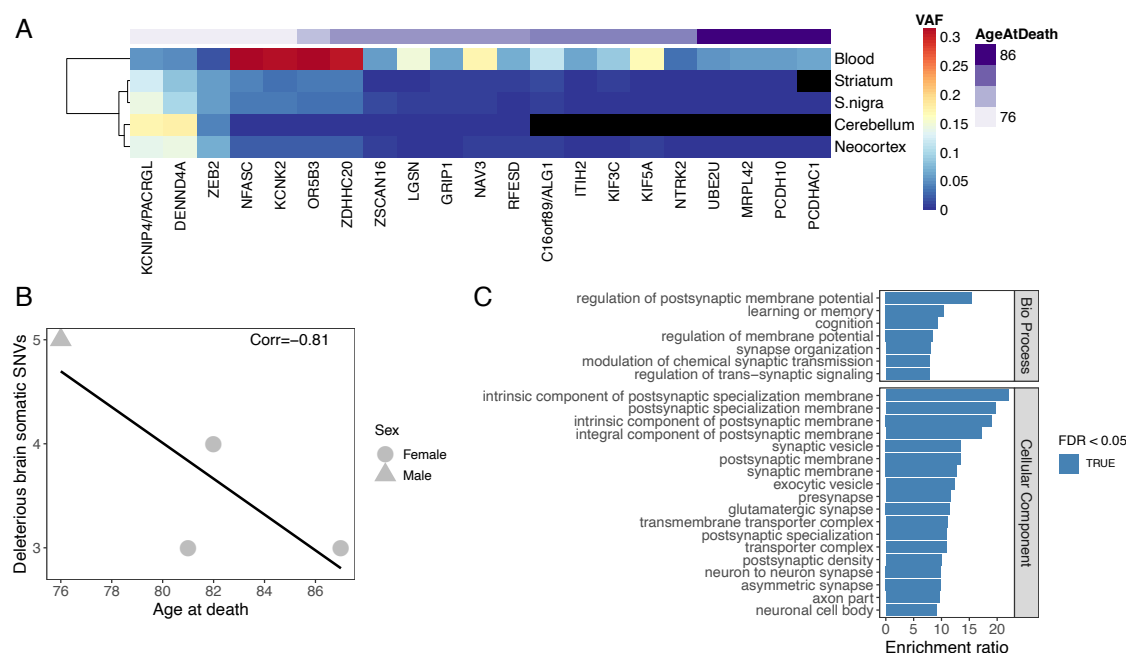
We then tested the relationship between the number of potentially deleterious brain variants and individuals' age at death. We excluded synonymous sSNVs (4 variants, including the only sSNV validated in DV8's brain, supplementary table 4) and intronic variants not affecting donor or acceptor sites (2 variants), leaving 11 nonsynonymous sSNVs, 3 variants affecting splicing consensus sites and 1 variant in a UTR region, as these are more probable to affect age at death. The age of onset was unknown for one of the four individuals included in this analysis, DV1. As age of onset is highly correlated with age at death (Pearson's  $r = 0.9$ ), we used it to evaluate the correlation between age and the number of sSNVs. Indeed, counts of deleterious somatic variants identified in brain were negatively correlated with age at death (Figure 4B, Pearson's  $r = -0.81$ ) but unsurprisingly from this small data set, the trend was not significant ( $P$ -value = 0.19).

## Genes carrying sSNVs in Parkinson are enriched in synaptic processes functions

Out of the 21 sSNVs found in brain, 4 nonsynonymous variants were predicted to be deleterious by SIFT. Remarkably, three of the genes carrying these variants are related to processes relevant in the brain. *GRIP1* is involved in transmission across chemical synapses and the regulation of neuron projection arborization (Geiger et al. 2014). *KCNK2* (or TREK-1) encodes a voltage-independent potassium channel essential in securing saltatory conduction at high frequency on myelinated afferent nerves (Kanda et al. 2019). *UBE2U*, a ubiquitination enzyme, is a candidate regulator of chromatin responses at double strand breaks (Guo et al. 2017) which are of fundamental relevance for gene expression in the brain (Madabhushi et al. 2015). Furthermore, ubiquitination dysfunction has been linked to Parkinson (Geisler et al. 2014) and Alzheimer (Gómez-Ramos et al. 2015). The other variant is in *DENND4A*, a secondary guanine nucleotide exchange factor that activates Rab-10, participating in the insulin-regulated glucose transporter GLUT4 translocation to cell membranes (Sano et al. 2011).

We explored the biological pathways potentially affected by the identified somatic mutations with an ORA. For this analysis we only included the 15 sSNVs that were nonsynonymous changes, that could affect splicing consensus sites or in UTR regions and were validated in a brain tissue. Top enriched terms included "glial cell projection", "axonogenesis" and similar processes (Supplementary figure 9). However, as expected from such a small set of genes, none of the enrichments were significant ( $FDR \leq 0.05$ ). To gain power and retrieve the GO terms that better describe the functional context of the validated sSNVs, we expanded the gene set by adding genes with significantly high co-expression scores (above 900) as reported in the STRING database (Szklarczyk et al. 2019), resulting in a total of 177 genes. Performing an enrichment analysis with the expanded gene set, we found significant results including many synaptic terms (top 25 in Figure 4C), showing that our validated sSNVs affect genes that are tightly connected to the protein networks associated to these functions. Besides the Gene Ontology databases, to test a more direct relationship of our expanded gene list with phenotypes and diseases, we also included the Human Phenotype Ontology and the GALD4U disease databases. Remarkably, chorea, dyskinesia and tremor were among the significantly enriched phenotypes and neurodegenerative diseases and Parkinson disease among the GALD4U database enriched terms (Supplementary Table 5). To further investigate the tissue specificity of the genes harboring the 15 validated sSNVs, we performed a tissue enrichment using

TissueEnrich (Jain and Geetu 2019). In accordance with the functional enrichment, cerebral cortex was the only significantly enriched tissue (adjusted P-value = 0.003 and Supplementary figure 10).



**Figure 4. A. Clustering of tissues by VAF.** Frequency of the 21 somatic variants found in brain was used to cluster the tissues. Genes are shown and variants are ordered by individual, with age at death shown on top. Black tiles indicate the variant did not pass all validation criteria in amplicon sequencing, but could still have support in the tissue. **B. Age correlation.** Correlation between each patient's number of potentially deleterious variants (nonsynonymous, in splicing consensus sites or in UTRs) found in brain and age at death. The line shows a non-significant fitted linear model ( $p$ -value = 0.19). **C. Functional enrichments of extended gene set.** Enrichment ratio of the top 25 terms by FDR of an overrepresentation enrichment analysis of the co-expression network extended gene set. The considered databases were geneontology Biological Process, Cellular Component, Molecular Function; Human Phenotype Ontology and disease GLAD4U.

To further explore the consequences of the validated variants, we examined their effect on protein structures and found two interesting cases. All tissues from patient DV7 carried a mutation (p.Gly998Glu) in *KIF5A* (Uniprot Q12840), a kinesin heavy chain protein. Using Genome3D (Lewis et al. 2015), we found a DomSerf (Buchan et al. 2013) structural prediction for the region containing this amino acid (913-1032, confidence 100). The affected residue is in the surface of this small globular C-terminal domain (Supplementary figure 11, left). This region interacts with kinesin adaptor proteins such as TRAK1 and TRAK2, which mediate cargo binding (Hirokawa and Noda 2008), suggesting the possible relevance of this sSNV. All DV1's tissues but cerebellum carried a substitution in *UBE2U* (Q5VVX9), p.Pro96Ala, which is included in a DomSerf structure of residues 2-157 (confidence 100) and appears to be close to the active site. Although it seems improbable that the structure of the active site changes as a consequence of this mutation (Supplementary figure 11, right), its flexibility or orientation may be affected.

## Discussion

The role of somatic mutations in a variety of diseases is becoming increasingly clear (Gleeson et al. 2000; Messiaen et al. 2011; Poduri et al. 2012; Priest et al. 2016; Bar et al. 2017; Nicolas et al. 2018; Mensa-Vilaró et al. 2019), including neurodegenerative (Park et al. 2018) and neurodevelopmental disorders (Dou et al. 2017). Here, we explored the presence of somatic mutations in a cohort of 10

Parkinson disease patients and validated 27 somatic SNVs whose functions are enriched in synaptic processes, suggesting a potential role of somatic variants in Parkinson disease.

sSNV calling from sequencing data outside the framework of paired samples still poses a challenge. Here we present COSMOS, an approach to accurately identify somatic mutations, that is highly customizable to each experiment's requirements, COSMOS (<https://github.com/ilobon/COSMOS>). We highlight the relevance of a panel of individuals to help identify germline variants and recurrent artefacts, which at the sensitivity a reasonable coverage provides, are far more common than somatic mutations. We also demonstrated that having multiple samples from the same individual can help increase accuracy and sensitivity limits, as previously suggested (Kim et al. 2019). Although this advantage can be offset by a panel of individuals for germline variant and recurrent artefacts identification, relevant mutations can be depleted in the context of neurodegeneration, so using unaffected tissues might be key for their discovery. Indeed, some of the somatic variants we validated did not have enough support in the brain tissues, and it was only the information from blood that allowed their detection. This could result from their random amplification in tissue maintenance of blood, especially in older aged individuals who show higher levels of clonal hematopoiesis (Watson et al. 2020) but could also be a consequence of depletion in the central nervous system. Including a wider variety of tissues in future studies could clarify this scenario.

A possible confounding factor in this type of analysis is blood contamination, which would result in the presence of the same somatic mutations at lower frequencies in other tissues. Our results show that 3 of the validated variants have consistently higher VAFs in the brain tissues than in blood. This demonstrates that blood cells are not a subset of the brain bulk samples. It would be interesting to include other tissues in this type of analyses, such as epithelia, which shares the advantage of the random amplification and absence of the putative depletion without this risk of contamination. Nonetheless there is positive selection of sSNVs in normal skin (Martincorena et al. 2015) and the extent to which this can be detrimental for their use as an outgroup tissue remains unknown.

The main caveat of our study is the lack of control individuals, for which paired tissue samples of similar quality are difficult to obtain. However, we took advantage of the remaining 45 samples and used them as a background panel to remove false positives with high statistical power. Also, all validated sSNVs were present in blood, discarding that this is an effect of using brain samples. Finally, their mutational spectrum agrees with our expectations. Extending this experiment to more tissues and individuals will be key to discover the fine role of somatic mutations in Parkinson disease. In addition to somatic mutations, we found that our cohort carried germline variants linked to the disease, hinting at the possibility that the presence of somatic mutations might contribute as an additive factor that increases an individual's susceptibility or disease severity.

We discarded variants when multiple individuals had reads supporting the same allele. This is because the probability of recurrent errors is much higher than the somatic mutation rate, especially given the number of divisions we can disentangle with bulk medium-coverage sequencing data. However, in the same way germline variants are recurrent because they appear in more fragile or tolerated regions of the genome, somatic mutations will surely be subject to the same processes. Furthermore, variants might be selected for or against in different tissues, as shown for epithelia (Martincorena et al. 2015, 2018; Lee-Six et al. 2019; Yoshida et al. 2020; Moore et al. 2020). Using more sensitive sequencing techniques, such as duplex sequencing (Schmitt et al. 2012), will be

fundamental to discover such somatic recurrent events. Finally, exome sequencing data, although more direct in its variant interpretation, is limited to a narrow set of regions of the genome. This limited our power to observe different cell lineages and their presence at each tissue or perform mutational signature deconvolution. Furthermore, the decreasing cost of sequencing and recent advances in contact maps (Lu et al. 2020) make finding and interpreting non-coding variants more accessible.

The somatic mutations we found in Parkinson patients seem to contribute to the disease, as they affect genes involved in neuronal and axonal pathways and interact with genes associated with the disease. Together with previous studies, our results suggest an exciting new research path for the study of diseases, especially complex diseases such as neurodegenerative disorders. This new evidence supports that not only germline point mutations, copy number variants, mitochondrial variants and the environment are relevant, but also somatic mutations can contribute to Parkinson disease, probably by affecting the same key pathways. The study of somatic mutations in larger cohorts can help to identify the relevant molecular routes, helping us understand the disease and finding potential therapy targets.

## Methods

### Exome sequencing

Tissue samples from cerebellum, neocortex, striatum and substantia nigra were collected at patients' autopsies, 6 to 18 hours after death, and kept frozen at -80 °C at the HCB-IDIBAPS biobank. Blood samples from the same individuals were obtained from stored vials. DNA extractions were carried out with the Qiagen DNeasy Blood & Tissue Kit. Genomic DNA samples were randomly fragmented into 150-200 bp length sequences. Adapters were ligated and the resulting templates were purified with AgencourtAMPure SPRI beads. Libraries were amplified by ligation-mediated polymerase chain reaction (LM-PCR). The Exon Focus SureSelect kit from Agilent was used to capture the exome and paired-end 100 bp sequencing was performed on an Illumina HiSeq2000 platform.

### Bam processing

The resulting FASTQ files were mapped with BWA v0.7.8 mem (Heng Li 2013) to the human hs37d5 assembly. Lane-specific read groups were added with Picard Tools v1.95 (Broad Institute 2013). AddOrReplaceReadGroups and bams were merged by sample with samtools v1.9 (H. Li et al. 2009). Read duplicates were removed with Picard Tools v1.95 MarkDuplicates REMOVE\_DUPLICATES=true. Base quality score recalibration and indel realignment were applied following GATK's best practices (DePristo et al. 2011) with GATK v3.6 (McKenna et al. 2010). Secondary alignments were also excluded with samtools view -F 256.

### Germline variants

Germline variants were called with GATK v3.6. First, GVCFs were obtained for each sample independently with -T HaplotypeCaller -emitRefConfidence GVCF. Then, all samples were genotyped together with -T GenotypeGVCFs and a standard hard filter was applied with -T VariantFiltration --filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 ||

ReadPosRankSum < -8.0". A PCA of the hard-filtered genotypes was performed with EIGENSOFT v7.2.1 (Price et al. 2006). Information on the called variants was annotated with SnpEff 4.3t and SnpSift 4.3t (Cingolani et al. 2012) and dbNSFP was used to add population frequencies, effect prediction and conservation scores. Overrepresentation enrichment analysis was performed with WebGestalt (Zhang, Kirov, and Snoddy 2005) using the Gene Ontology (GO) database (Carbon et al. 2009) for molecular functions and genome protein-coding genes as background.

## Somatic variant calling

Somatic variants were called using HaplotypeCaller and VarScan 2 with lax parameters. Then, an extensive filtering strategy was applied to recover high confidence somatic SNVs. For HaplotypeCaller somatic variant calling, GVCFs per sample were obtained with GATK -T HaplotypeCaller -ploidy 10 -A StrandAlleleCountsBySample --emitRefConfidence GVCF. Then, all GVCFs were genotyped together per chromosome with -T GenotypeGVCFs -L chr -ploidy 10 -A StrandAlleleCountsBySample to obtain somatic SNV and indel calls. For VarScan 2 (Koboldt et al. 2012) somatic variant calling, mpileup files were obtained with samtools mpileup per individual. Then, single nucleotide variants were called with VarScan v2.3.2 mpileup2snp with lax parameters: --min-coverage 1 --min-reads2 1 --p-value 1 --min-var-freq 0.000001 --output-vcf. Indels were called with mpileup2indel with the same parameters.

Depth of coverage files were obtained with GATK v3.6 DepthOfCoverage and GC content per target was calculated with GCContentByInterval. Then, CNVs were called jointly for all samples with XHMM v1.0 (Fromer et al. 2012) with standard parameters following its recommended best practices. Short tandem repeats in hs37d5 were determined with Tandem Repeats Finder v.4.09 (Benson 1999) with parameters 2 7 7 80 10 12 500 -h. Homopolymers were then extracted based on their homogeneous repeat motif. The 1000GP strict mask FASTA files were obtained from the 1000GP FTP ([ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible\\_genome\\_masks/](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_masks/)) and were transformed into a BED file. BED files for WGAC segmental duplications, common dbSNP SNPs and mappability for 100mers for hg19 – which shares coordinates with hs37d5 – were obtained from the UCSC table browser (Karolchik et al. 2004). This information was annotated in the VCF files with BCFtools (H. Li et al. 2009). A panel of normals (PON) containing allele counts at each genomic position along 428 individuals sequenced at the Sanger Institute was used to identify recurrent errors.

## Somatic variant filtering using COSMOS

We first defined a non-callable set of positions, including off-target calls as well as those overlapping with the 1000GP strict mask, the WCAG track, mappability lower than 1, by homopolymers or within 5 base pairs of indels. Additional sources of artefacts were identified in the lax call sets by manual inspection of the raw data with IGV (Robinson et al. 2011), oftentimes evidenced by recurrent patterns in different individuals. To filter calls we wrote a python module, COSMOS (Combined Or Single sample MOSaicism detection), which is available in <https://github.com/ilobon/COSMOS>. The first step is to annotate all the information necessary to classify true and false calls. Then, variants are filtered according to the user indicated features and thresholds. Two different filtering approaches can be used. When multiple samples – tissues or replicates – from the same individual are available, their combined information can be used to filter calls more accurately. This is, since true and false variants have partially overlapping values in the determining features, even true somatic mutations

can fail some tests. Hence, we can take advantage of the multiple samples, and assuming the variants are present in more than one of them, require that at least  $n$  samples pass each filter, allowing a different combination of passing samples at each feature (--combined TRUE --nSamplesPerInd  $n$ ). COSMOS can also be used to filter each sample individually (--combined FALSE), or output the result of both approaches (--combined BOTH --nSamplesPerInd  $n$ ).

To filter our Parkinson exome data, we used both approaches, requiring at each position either a single sample passing all filters or any combination of four out of the five tissues passing at every filter. The parameters we used were `-c BOTH -ns 4 -ad 2 -adss 3 -vaf 0.5 -dp1 20 -dp2 100 -sr 2 -pr 4 -sp 0.05 -b 0.05 -nrl 4 -hap 4 -cnv NO -pir 4,4 -vafq 0.4 -clip 0.9 -mq 0.05 -mm 0.05 -pon 0.05`. In short, this requires at least 2 reads supporting the alternative allele, or 3 for the single sample approach; a variant allele frequency lower than 0.5; a depth between 20 and 100; a strand count ratio <2; a pair count ratio <4; non-significant p-values for a Poisson test of strand counts; a significant binomial test of allele counts; at most 4 variants within read length; less than 4 haplotypes; absence of XHMM CNV call; PIR score of 4, meaning that position in reads of each allele is not biased; a variant allele frequency from direct bam read counts <0.4; an allele clipping ratio difference <0.9; non-significant Mann-Whitney tests for mapping quality and number of mismatches per allele and a significant beta-binomial test for allele counts compared to the PON. More features and different thresholds can be used for filtering depending on each dataset characteristics such as depth.

### Amplicon-based deep sequencing

To validate the high confidence variants, we performed amplicon-based deep sequencing (ADS) by using rhAmpSeq technology (IDT, Coralville, USA). This is a multiplexed strategy that amplifies all selected positions in a unified reaction. We then sequenced the amplified material from each sample in a MiSeq v3 run obtaining paired end 300bp reads to a mean coverage ~18,000X. Data were processed the same way as the exome files. Sequencing data at each position in the samples from other individuals were used as negative controls.



- Abyzov, Alexej, Livia Tomasini, Bo Zhou, Nikolaos Vasmataz, Gianfilippo Coppola, Mariangela Amenduni, Reenal Pattani, et al. 2017. "One Thousand Somatic SNVs per Skin Fibroblast Cell Set Baseline of Mosaic Mutational Load with Patterns That Suggest Proliferative Origin." *Genome Research* 27 (4): 512–23. <https://doi.org/10.1101/gr.215517.116>.
- Alexandrov, Ludmil B., Serena Nik-Zainal, David C. Wedge, Peter J. Campbell, and Michael R. Stratton. 2013. "Deciphering Signatures of Mutational Processes Operative in Human Cancer." *Cell Reports* 3 (1): 246–59. <https://doi.org/10.1016/J.CELREP.2012.12.008>.
- Alexandrov, Ludmil B., Jaegil Kim, Nicholas J. Haradhvala, Mi Ni Huang, Alvin Wei Tian Ng, Yang Wu, Arnoud Boot, et al. 2020. "The Repertoire of Mutational Signatures in Human Cancer." *Nature* 578 (7793): 94–101. <https://doi.org/10.1038/s41586-020-1943-3>.
- Bae, Taejeong, Livia Tomasini, Jessica Mariani, Bo Zhou, Tanmoy Roychowdhury, Daniel Franjic, Mihovil Pletikos, et al. 2018. "Different Mutational Rates and Mechanisms in Human Cells at Pregastrulation and Neurogenesis." *Science (New York, N.Y.)* 359 (6375): 550–55. <https://doi.org/10.1126/science.aan8690>.
- Baillie, J. Kenneth, Mark W. Barnett, Kyle R. Upton, Daniel J. Gerhardt, Todd A. Richmond, Fioravante De Sapio, Paul M. Brennan, et al. 2011. "Somatic Retrotransposition Alters the Genetic Landscape of the Human Brain." *Nature* 479 (7374): 534–37. <https://doi.org/10.1038/nature10531>.
- Bar, Daniel Z, Martin F Arlt, Joan F Brazier, Wendy E Norris, Susan E Campbell, Peter Chines, Delphine Larrieu, et al. 2017. "A Novel Somatic Mutation Achieves Partial Rescue in a Child with Hutchinson-Gilford Progeria Syndrome." *Journal of Medical Genetics* 54 (3): 212–16. <https://doi.org/10.1136/jmedgenet-2016-104295>.
- Bender, Andreas, Kim J Krishnan, Christopher M Morris, Geoffrey A Taylor, Amy K Reeve, Robert H Perry, Evelyn Jaros, et al. 2006. "High Levels of Mitochondrial DNA Deletions in Substantia Nigra Neurons in Aging and Parkinson Disease." *Nature Genetics* 38 (5): 515–17. <https://doi.org/10.1038/ng1769>.
- Benson, G. 1999. "Tandem Repeats Finder: A Program to Analyze DNA Sequences." *Nucleic Acids Research* 27 (2): 573–80. <https://doi.org/10.1093/nar/27.2.573>.
- Bernheimer, H., W. Birkmayer, O. Hornykiewicz, K. Jellinger, and F. Seitelberger. 1973. "Brain Dopamine and the Syndromes of Parkinson and Huntington Clinical, Morphological and Neurochemical Correlations." *Journal of the Neurological Sciences* 20 (4): 415–55. [https://doi.org/10.1016/0022-510X\(73\)90175-5](https://doi.org/10.1016/0022-510X(73)90175-5).
- Broad Institute. 2013. "Picard: A Set of Command Line Tools (in Java) for Manipulating High-Throughput Sequencing (HTS) Data and Formats Such as SAM/BAM/CRAM and VCF." 2013. <https://broadinstitute.github.io/picard/>.
- Brunner, Simon F., Nicola D. Roberts, Luke A. Wylie, Luiza Moore, Sarah J. Aitken, Susan E. Davies, Mathijs A. Sanders, et al. 2019. "Somatic Mutations and Clonal Dynamics in Healthy and Cirrhotic Human Liver." *Nature* 574 (7779): 538–42. <https://doi.org/10.1038/s41586-019-1670-9>.
- Buchan, Daniel W.A., Federico Minneci, Tim C.O. Nugent, Kevin Bryson, and David T. Jones. 2013. "Scalable Web Services for the PSIPRED Protein Analysis Workbench." *Nucleic Acids Research* 41 (Web Server issue). <https://doi.org/10.1093/nar/gkt381>.
- Cai, Xuyu, Gilad D. Evrony, Hillel S. Lehmann, Princess C. Elhosary, Bhaven K. Mehta, Annapurna Poduri, and Christopher A. Walsh. 2014. "Single-Cell, Genome-Wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain." *Cell Reports* 8 (5): 1280–89. <https://doi.org/10.1016/J.CELREP.2014.07.043>.
- Carbon, Seth, Amelia Ireland, Christopher J. Mungall, ShengQiang Shu, Brad Marshall, and Suzanna Lewis. 2009. "AmiGO: Online Access to Ontology and Annotation Data." *Bioinformatics* 25 (2): 288–89. <https://doi.org/10.1093/bioinformatics/btn615>.
- Chen, Chiung-Mei, Yi-Chun Chen, Mu-Chun Chiang, Hon-Chung Fung, Kuo-Hsuan Chang, Guey-Jen Lee-Chen, and Yih-Ru Wu. 2016. "Association of GCH1 and MIR4697, but Not SIPA1L2 and VPS13C Polymorphisms, with Parkinson's Disease in Taiwan." *Neurobiology of Aging* 39 (March): 221.e1-221.e5. <https://doi.org/10.1016/J.NEUROBIOLAGING.2015.12.016>.
- Chronister, William D., Ian E. Burbulis, Margaret B. Wierman, Matthew J. Wolpert, Mark F. Haakenson, Aiden C.B. Smith, Joel E. Kleinman, et al. 2019. "Neurons with Complex Karyotypes Are Rare in Aged Human Neocortex." *Cell Reports* 26 (4): 825-835.e7. <https://doi.org/10.1016/J.CELREP.2018.12.107>.
- Chuang, Yu-Hsuan, Kimberly C. Paul, Jeff M. Bronstein, Yvette Bordelon, Steve Horvath, and Beate

- Ritz. 2017. "Parkinson's Disease Is Associated with DNA Methylation Levels in Human Blood and Saliva." *Genome Medicine* 9 (1): 76. <https://doi.org/10.1186/s13073-017-0466-5>.
- Cingolani, Pablo, Viral M Patel, Melissa Coon, Tung Nguyen, Susan J Land, Douglas M Ruden, and Xiangyi Lu. 2012. "Using Drosophila Melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift." *Frontiers in Genetics* 3: 35. <https://doi.org/10.3389/fgene.2012.00035>.
- DePristo, Mark A, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, Christopher Hartl, Anthony A Philippakis, et al. 2011. "A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data." *Nature Genetics* 43 (5): 491–98. <https://doi.org/10.1038/ng.806>.
- Do, Chuong B., Joyce Y. Tung, Elizabeth Dorfman, Amy K. Kiefer, Emily M. Drabant, Uta Francke, Joanna L. Mountain, et al. 2011. "Web-Based Genome-Wide Association Study Identifies Two Novel Loci and a Substantial Genetic Component for Parkinson's Disease." Edited by Greg Gibson. *PLoS Genetics* 7 (6): e1002141. <https://doi.org/10.1371/journal.pgen.1002141>.
- Dou, Yanmei, Xiaoxu Yang, Ziyi Li, Sheng Wang, Zheng Zhang, Adam Yongxin Ye, Linlin Yan, et al. 2017. "Postzygotic Single-Nucleotide Mosaicism Contributes to the Etiology of Autism Spectrum Disorder and Autistic Traits and the Origin of Mutations." *Human Mutation* 38 (8): 1002–13. <https://doi.org/10.1002/humu.23255>.
- Elstner, Matthias, Christopher M. Morris, Katharina Heim, Peter Lichtner, Andreas Bender, Divya Mehta, Claudia Schulte, et al. 2009. "Single-Cell Expression Profiling of Dopaminergic Neurons Combined with Association Analysis Identifies Pyridoxal Kinase as Parkinson's Disease Gene." *Annals of Neurology* 66 (6): 792–98. <https://doi.org/10.1002/ana.21780>.
- Federoff, Monica, Belen Jimenez-Rolando, Michael A. Nalls, and Andrew B. Singleton. 2012. "A Large Study Reveals No Association between APOE and Parkinson's Disease." *Neurobiology of Disease* 46 (2): 389–92. <https://doi.org/10.1016/J.NBD.2012.02.002>.
- Fromer, Menachem, Jennifer L. Moran, Kimberly Chambert, Eric Banks, Sarah E. Bergen, Douglas M. Ruderfer, Robert E. Handsaker, et al. 2012. "Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth." *The American Journal of Human Genetics* 91 (4): 597–607. <https://doi.org/10.1016/J.AJHG.2012.08.005>.
- Geiger, Julia C., Joanna Lipka, Inmaculada Segura, Susanne Hoyer, Max A. Schlager, Phebe S. Wulf, Stefan Weinges, Jeroen Demmers, Casper C. Hoogenraad, and Amparo Acker-Palmer. 2014. "The GRIP1/14-3-3 Pathway Coordinates Cargo Trafficking and Dendrite Development." *Developmental Cell* 28 (4): 381–93. <https://doi.org/10.1016/j.devcel.2014.01.018>.
- Geisler, Sven, Stefanie Vollmer, Sonia Golombek, and Philipp J. Kahle. 2014. "The Ubiquitin-Conjugating Enzymes UBE2N, UBE2L3 and UBE2D2/3 Are Essential for Parkin-Dependent Mitophagy." *Journal of Cell Science* 127 (15): 3280–93. <https://doi.org/10.1242/jcs.146035>.
- Gleeson, J G, S Minnerath, R I Kuzniecky, W B Dobyns, I D Young, M E Ross, and C A Walsh. 2000. "Somatic and Germline Mosaic Mutations in the Doublecortin Gene Are Associated with Variable Phenotypes." *American Journal of Human Genetics* 67 (3): 574–81. <https://doi.org/10.1086/303043>.
- Gómez-Ramos, A., P. Podlesniy, E. Soriano, and J. Avila. 2015. "Distinct X-Chromosome SNVs from Some Sporadic AD Samples." *Scientific Reports* 5 (1): 1–11. <https://doi.org/10.1038/srep18012>.
- Guella, Ilaria, Rosanna Asselta, Silvana Tesei, Michela Zini, Gianni Pezzoli, and Stefano Duga. 2010. "The PDXK Rs2010795 Variant Is Not Associated with Parkinson Disease in Italy." *Annals of Neurology* 67 (3): 411–12. <https://doi.org/10.1002/ana.21964>.
- Guertin, David A., and David M. Sabatini. 2007. "Defining the Role of MTOR in Cancer." *Cancer Cell*. Cell Press. <https://doi.org/10.1016/j.ccr.2007.05.008>.
- Guo, Yingying, Liwei An, Hoi Man Ng, Shirley M.H. Sy, and Michael S.Y. Huen. 2017. "An E2-Guided E3 Screen Identifies the RNF17-UBE2U Pair as Regulator of the Radiosensitivity, Immunodeficiency, Dysmorphic Features, and Learning Difficulties (RIDDLE) Syndrome Protein RNF168." *Journal of Biological Chemistry* 292 (3): 967–78. <https://doi.org/10.1074/jbc.M116.758854>.
- Heckman, Michael G., Koji Kasanuki, Nancy N. Diehl, Shunsuke Koga, Alexandra Soto, Melissa E. Murray, Dennis W. Dickson, and Owen A. Ross. 2017. "Parkinson's Disease Susceptibility Variants and Severity of Lewy Body Pathology." *Parkinsonism & Related Disorders* 44 (November): 79–84. <https://doi.org/10.1016/J.PARKRELDIS.2017.09.009>.

- Hirokawa, Nobutaka, and Yasuko Noda. 2008. "Intracellular Transport and Kinesin Superfamily Proteins, KIFs: Structure, Function, and Dynamics." *Physiological Reviews*. American Physiological Society. <https://doi.org/10.1152/physrev.00023.2007>.
- Hoang, Margaret L., Isaac Kinde, Cristian Tomasetti, K. Wyatt McMahon, Thomas A. Rosenquist, Arthur P. Grollman, Kenneth W. Kinzler, Bert Vogelstein, and Nickolas Papadopoulos. 2016. "Genome-Wide Quantification of Rare Somatic Mutations in Normal Human Tissues Using Massively Parallel Sequencing." *Proceedings of the National Academy of Sciences of the United States of America* 113 (35): 9846–51. <https://doi.org/10.1073/pnas.1607794113>.
- Huang, X., P. C. Chen, and C. Poole. 2004. "APOE- 2 Allele Associated with Higher Prevalence of Sporadic Parkinson Disease." *Neurology* 62 (12): 2198–2202. <https://doi.org/10.1212/01.WNL.0000130159.28215.6A>.
- International Parkinson Disease Genomics Consortium. 2011. "Imputation of Sequence Variants for Identification of Genetic Risks for Parkinson's Disease: A Meta-Analysis of Genome-Wide Association Studies." *The Lancet* 377 (9766): 641–49. [https://doi.org/10.1016/S0140-6736\(10\)62345-8](https://doi.org/10.1016/S0140-6736(10)62345-8).
- International Parkinson Disease Genomics Consortium, and Wellcome Trust Case Control Consortium 2. 2011. "A Two-Stage Meta-Analysis Identifies Several New Loci for Parkinson's Disease." Edited by Greg Gibson. *PLoS Genetics* 7 (6): e1002142. <https://doi.org/10.1371/journal.pgen.1002142>.
- Jain, Ashish, and Tuteja Geetu. 2019. "TissueEnrich: Tissue-Specific Gene Enrichment Analysis." *Bioinformatics*. 2019. <https://academic.oup.com/bioinformatics/article-abstract/35/11/1966/5140218>.
- Jaiswal, Siddhartha, Pierre Fontanillas, Jason Flannick, Alisa Manning, Peter V. Grauman, Brenton G. Mar, R. Coleman Lindsley, et al. 2014. "Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes." *New England Journal of Medicine* 371 (26): 2488–98. <https://doi.org/10.1056/NEJMoa1408617>.
- Jinn, Sarah, Robert E. Drolet, Paige E. Cramer, Andus Hon-Kit Wong, Dawn M. Toolan, Cheryl A. Gretzula, Bhavya Voleti, et al. 2017. "TMEM175 Deficiency Impairs Lysosomal and Mitochondrial Function and Increases  $\alpha$ -Synuclein Aggregation." *Proceedings of the National Academy of Sciences* 114 (9): 2389–94. <https://doi.org/10.1073/PNAS.1616332114>.
- Kanda, Hirosato, Jennifer Ling, Sotatsu Tonomura, Koichi Noguchi, Sadis Matalon, and Jianguo G. Gu. 2019. "TREK-1 and TRAAK Are Principal K<sup>+</sup> Channels at the Nodes of Ranvier for Rapid Action Potential Conduction on Mammalian Myelinated Afferent Nerves." *Neuron* 104 (5): 960-971.e7. <https://doi.org/10.1016/j.neuron.2019.08.042>.
- Karolchik, D., Angela S Hinrichs, Terrence S Furey, Krishna M Roskin, Charles W Sugnet, David Haussler, and W James Kent. 2004. "The UCSC Table Browser Data Retrieval Tool." *Nucleic Acids Research* 32 (90001): 493D – 496. <https://doi.org/10.1093/nar/gkh103>.
- Kim, Junho, Dachan Kim, Jae Seok Lim, Ju Heon Maeng, Hyeonju Son, Hoon-Chul Kang, Hojung Nam, Jeong Ho Lee, and Sangwoo Kim. 2019. "The Use of Technical Replication for Detection of Low-Level Somatic Mutations in next-Generation Sequencing." *Nature Communications* 10 (1): 1047. <https://doi.org/10.1038/s41467-019-09026-y>.
- Kluge, Michelle L, Evan Graber, Kathryn Foley, Lynnette V Hansen, Heidi L Sellers, Dragana Milosevic, Kendall W Cradic, and Stefan K Grebe. 2019. "Monozygotic Twins Discordant for Congenital Adrenal Hyperplasia Due to Mosaicism." *European Journal of Endocrinology*, December. <https://doi.org/10.1530/EJE-19-0249>.
- Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson. 2012. "VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing." *Genome Research* 22 (3): 568–76. <https://doi.org/10.1101/gr.129684.111>.
- Kong, Ping, Ping Lei, Shishuang Zhang, Dai Li, Jing Zhao, and Benshu Zhang. 2018. "Integrated Microarray Analysis Provided a New Insight of the Pathogenesis of Parkinson's Disease." *Neuroscience Letters* 662 (January): 51–58. <https://doi.org/10.1016/J.NEULET.2017.09.051>.
- Krøigård, Anne Bruun, Mads Thomassen, Anne-Vibeke Lænkholm, Torben A. Kruse, and Martin Jakob Larsen. 2016. "Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data." Edited by I. King Jordan. *PLOS ONE* 11 (3): e0151664. <https://doi.org/10.1371/journal.pone.0151664>.
- Lee-Six, Henry, Sigurgeir Olafsson, Peter Ellis, Robert J. Osborne, Mathijs A. Sanders, Luiza Moore, Nikitas Georgakopoulos, et al. 2019. "The Landscape of Somatic Mutation in Normal

- Colorectal Epithelial Cells." *Nature* 574 (7779): 532–37. <https://doi.org/10.1038/s41586-019-1672-7>.
- Lesage, Suzanne, and Alexis Brice. 2009. "Parkinson's Disease: From Monogenic Forms to Genetic Susceptibility Factors." *Human Molecular Genetics* 18 (Review Issue 1): R48–R59. <https://doi.org/10.1093/hmg/ddp012>.
- Lewis, Tony E., Ian Sillitoe, Antonina Andreeva, Tom L. Blundell, Daniel W.A. Buchan, Cyrus Chothia, Domenico Cozzetto, et al. 2015. "Genome3D: Exploiting Structure to Help Users Understand Their Sequences." *Nucleic Acids Research* 43 (D1): D382–86. <https://doi.org/10.1093/nar/gku973>.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM," March. <http://arxiv.org/abs/1303.3997>.
- Li, Jiao, Jennifer A. Ruskey, Isabelle Arnulf, Yves Dauvilliers, Michele T.M. Hu, Birgit Högl, Claire S. Leblond, et al. 2018. "Full Sequencing and Haplotype Analysis of *MAPT* in Parkinson's Disease and Rapid Eye Movement Sleep Behavior Disorder." *Movement Disorders* 33 (6): 1016–20. <https://doi.org/10.1002/mds.27385>.
- Lill, Christina M., Johannes T. Roehr, Matthew B. McQueen, Fotini K. Kavvoura, Sachin Bagade, Brit-Maren M. Schjeide, Leif M. Schjeide, et al. 2012. "Comprehensive Research Synopsis and Systematic Meta-Analyses in Parkinson's Disease Genetics: The PDGene Database." Edited by Amanda J. Myers. *PLoS Genetics* 8 (3): e1002548. <https://doi.org/10.1371/journal.pgen.1002548>.
- Lim, Jae Seok, Woo-il Kim, Hoon-Chul Kang, Se Hoon Kim, Ah Hyung Park, Eun Kyung Park, Young-Wook Cho, et al. 2015. "Brain Somatic Mutations in MTOR Cause Focal Cortical Dysplasia Type II Leading to Intractable Epilepsy." *Nature Medicine* 21 (4): 395–400. <https://doi.org/10.1038/nm.3824>.
- Lodato, Michael A, Rachel E Rodin, Craig L Bohrsen, Michael E Coulter, Alison R Barton, Minseok Kwon, Maxwell A Sherman, et al. 2018. "Aging and Neurodegeneration Are Associated with Increased Mutations in Single Human Neurons." *Science (New York, N.Y.)* 359 (6375): 555–59. <https://doi.org/10.1126/science.aao4426>.
- Lodato, Michael A, Mollie B Woodworth, Semin Lee, Gilad D Evrony, Bhaven K Mehta, Amir Karger, Soohyun Lee, et al. 2015. "Somatic Mutation in Single Human Neurons Tracks Developmental and Transcriptional History." *Science (New York, N.Y.)* 350 (6256): 94–98. <https://doi.org/10.1126/science.aab1785>.
- Lu, Leina, Xiaoxiao Liu, Wei-Kai Huang, Paola Giusti-Rodríguez, Jian Cui, Shanshan Zhang, Wanying Xu, et al. 2020. "Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-Coding Genome in Neural Development and Diseases." *Molecular Cell* 0 (0). <https://doi.org/10.1016/j.molcel.2020.06.007>.
- Madabhushi, Ram, Fan Gao, Andreas R Pfenning, Ling Pan, Satoko Yamakawa, Jinsoo Seo, Richard Rueda, et al. 2015. "Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes." *Cell* 161 (7): 1592–1605. <https://doi.org/10.1016/j.cell.2015.05.032>.
- Martincorena, Iñigo, Joanna C Fowler, Agnieszka Wabik, Andrew R J Lawson, Federico Abascal, Michael W J Hall, Alex Cagan, et al. 2018. "Somatic Mutant Clones Colonize the Human Esophagus with Age." *Science (New York, N.Y.)* 362 (6417): 911–17. <https://doi.org/10.1126/science.aau3879>.
- Martincorena, Iñigo, Amit Roshan, Moritz Gerstung, Peter Ellis, Peter Van Loo, Stuart McLaren, David C Wedge, et al. 2015. "Tumor Evolution. High Burden and Pervasive Positive Selection of Somatic Mutations in Normal Human Skin." *Science (New York, N.Y.)* 348 (6237): 880–86. <https://doi.org/10.1126/science.aaa6806>.
- Mathieson, Iain, and David Reich. 2017. "Differences in the Rare Variant Spectrum among Human Populations." Edited by Santhosh Girirajan. *PLOS Genetics* 13 (2): e1006581. <https://doi.org/10.1371/journal.pgen.1006581>.
- McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernysky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303. <https://doi.org/10.1101/gr.107524.110>.



- Mensa-Vilaró, Anna, María Bravo García-Morato, Oscar de la Calle-Martin, Clara Franco-Jarava, María Teresa Martínez-Saavedra, Luis I. González-Granado, Eva González-Roca, et al. 2019. "Unexpected Relevant Role of Gene Mosaicism in Patients with Primary Immunodeficiency Diseases." *Journal of Allergy and Clinical Immunology* 143 (1): 359–68. <https://doi.org/10.1016/J.JACI.2018.09.009>.
- Messiaen, Ludwine, Julia Vogt, Kathrin Bengesser, Chuanhua Fu, Fady Mikhail, Eduard Serra, Carles Garcia-Linares, David N. Cooper, Conxi Lazaro, and Hildegard Kehrer-Sawatzki. 2011. "Mosaic Type-1 NF1 Microdeletions as a Cause of Both Generalized and Segmental Neurofibromatosis Type-1 (NF1)." *Human Mutation* 32 (2): 213–19. <https://doi.org/10.1002/humu.21418>.
- Mokretar, Katya, Daniel Pease, Jan-Willem Taanman, Aynur Soenmez, Ayesha Ejaz, Tammaryn Lashley, Helen Ling, et al. 2018. "Somatic Copy Number Gains of  $\alpha$ -Synuclein (SNCA) in Parkinson's Disease and Multiple System Atrophy Brains." *Brain* 141 (8): 2419–31. <https://doi.org/10.1093/brain/aww157>.
- Moore, Luiza, Daniel Leongamornlert, Tim H.H. Coorens, Mathijs A. Sanders, Peter Ellis, Stefan C. Dentre, Kevin J. Dawson, et al. 2020. "The Mutational Landscape of Normal Human Endometrial Epithelium." *Nature* 580 (7805): 640–46. <https://doi.org/10.1038/s41586-020-2214-z>.
- Nalls, Mike A, Nathan Pankratz, Christina M Lill, Chuong B Do, Dena G Hernandez, Mohamad Saad, Anita L DeStefano, et al. 2014. "Large-Scale Meta-Analysis of Genome-Wide Association Data Identifies Six New Risk Loci for Parkinson's Disease." *Nature Genetics* 46 (9): 989–93. <https://doi.org/10.1038/ng.3043>.
- Nicolas, Gaël, Rocio Acuña-Hidalgo, Michael J. Keogh, Olivier Quenez, Marloes Steehouwer, Stefan Lelieveld, Stéphane Rousseau, et al. 2018. "Somatic Variants in Autosomal Dominant Genes Are a Rare Cause of Sporadic Alzheimer's Disease." *Alzheimer's & Dementia* 14 (12): 1632–39. <https://doi.org/10.1016/J.JALZ.2018.06.3056>.
- Nowell, Peter C. 1976. "The Clonal Evolution of Tumor Cell Populations." *Science* 194 (4260): 23–28. <https://doi.org/10.1126/science.959840>.
- Pamphlett, Roger. 2004. "Somatic Mutation: A Cause of Sporadic Neurodegenerative Diseases?" *Medical Hypotheses* 62 (5): 679–82. <https://doi.org/10.1016/J.MEHY.2003.11.023>.
- Park, Jun Sung, Junehawk Lee, Eun Sun Jung, Myeong Heui Kim, Il Bin Kim, Hyeonju Son, Sangwoo Kim, et al. 2019. "Brain Somatic Mutations Observed in Alzheimer's Disease Associated with Aging and Dysregulation of Tau Phosphorylation." *Nature Communications* 10 (1): 1–12. <https://doi.org/10.1038/s41467-019-11000-7>.
- Poduri, Annapurna, Gilad D. Evrony, Xuyu Cai, Princess Christina Elhosary, Rameen Beroukhim, Maria K. Lehtinen, L. Benjamin Hills, et al. 2012. "Somatic Activation of AKT3 Causes Hemispheric Developmental Brain Malformations." *Neuron* 74 (1): 41–48. <https://doi.org/10.1016/J.NEURON.2012.03.010>.
- Price, Alkes L, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics* 38 (8): 904–9. <https://doi.org/10.1038/ng1847>.
- Priest, James Rush, Charles Gawad, Kristopher M Kahlig, Joseph K Yu, Thomas O'Hara, Patrick M Boyle, Sridharan Rajamani, et al. 2016. "Early Somatic Mosaicism Is a Rare Cause of Long-QT Syndrome." *Proceedings of the National Academy of Sciences of the United States of America* 113 (41): 11555–60. <https://doi.org/10.1073/pnas.1607187113>.
- Proukakis, Christos, Maryiam Shoaee, James Morris, Timothy Brier, Eleanna Kara, Una-Marie Sheerin, Gavin Charlesworth, et al. 2014. "Analysis of Parkinson's Disease Brain-Derived DNA for Alpha-Synuclein Coding Somatic Mutations." *Movement Disorders* 29 (8): 1060–64. <https://doi.org/10.1002/mds.25883>.
- Rahbari, Raheleh, Arthur Wuster, Sarah J Lindsay, Robert J Hardwick, Ludmil B Alexandrov, Saeed Al Turki, Anna Dominiczak, et al. 2016. "Timing, Rates and Spectra of Human Germline Mutation." *Nature Genetics* 48 (2): 126–33. <https://doi.org/10.1038/ng.3469>.
- Ramakrishnan, V., T. Alu Alphonsa, RS Akram Husain, Shiek SSJ Ahmed, K. Subramaniyan, and Suresh Kumar. 2016. "Association of Rs1801582 and Rs1801334 PARK2 Polymorphisms with Risk of Parkinson's Disease: A Case-Control Study in South India and Meta-Analysis." *Meta Gene* 10 (December): 32–38. <https://doi.org/10.1016/J.MGENE.2016.09.007>.
- Rentzsch, Philipp, Daniela Witten, Gregory M Cooper, Jay Shendure, and Martin Kircher. 2019. "CADD: Predicting the Deleteriousness of Variants throughout the Human Genome." *Nucleic*

- Acids Research* 47 (D1): D886–94. <https://doi.org/10.1093/nar/gky1016>.
- Robinson, James T, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. 2011. “Integrative Genomics Viewer.” *Nature Biotechnology* 29 (1): 24–26. <https://doi.org/10.1038/nbt.1754>.
- Rocca, Walter A., Shannon K. McDonnell, Kari J. Strain, James H. Bower, J. Eric Ahlskog, Alexis Elbaz, Daniel J. Schaid, and Demetrius M. Maraganore. 2004. “Familial Aggregation of Parkinson’s Disease: The Mayo Clinic Family Study.” *Annals of Neurology* 56 (4): 495–502. <https://doi.org/10.1002/ana.20228>.
- Sano, Hiroyuki, Grantley R. Peck, Arminja N. Kettenbach, Scott A. Gerber, and Gustav E. Lienhard. 2011. “Insulin-Stimulated GLUT4 Protein Translocation in Adipocytes Requires the Rab10 Guanine Nucleotide Exchange Factor Dennd4C.” *Journal of Biological Chemistry* 286 (19): 16541–45. <https://doi.org/10.1074/jbc.C111.228908>.
- Schmitt, Michael W., Scott R. Kennedy, Jesse J. Salk, Edward J. Fox, Joseph B. Hiatt, and Lawrence A. Loeb. 2012. “Detection of Ultra-Rare Mutations by next-Generation Sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (36): 14508–13. <https://doi.org/10.1073/pnas.1208715109>.
- Smith, Y., R. M. Villalba, and D. V. Raju. 2009. “Striatal Spine Plasticity in Parkinson’s Disease: Pathological or Not?” *Parkinsonism and Related Disorders* 15 (SUPPL. 3): S156–61. [https://doi.org/10.1016/S1353-8020\(09\)70805-3](https://doi.org/10.1016/S1353-8020(09)70805-3).
- Song, Gwan Gyu, and Young Ho Lee. 2013. “Pathway Analysis of Genome-Wide Association Studies for Parkinson’s Disease.” *Molecular Biology Reports* 40 (3): 2599–2607. <https://doi.org/10.1007/s11033-012-2346-9>.
- Spataro, Nino, Francesc Calafell, Laura Cervera-Carles, Ferran Casals, Javier Pagonabarraga, Berta Pascual-Sedano, Antònia Campolongo, et al. 2015. “Mendelian Genes for Parkinson’s Disease Contribute to the Sporadic Forms of the Disease†.” *Human Molecular Genetics* 24 (7): 2023–34. <https://doi.org/10.1093/hmg/ddu616>.
- Szklarczyk, Damian, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, et al. 2019. “STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets.” *Nucleic Acids Research* 47 (D1): D607–13. <https://doi.org/10.1093/nar/gky1131>.
- Tan, E-K, H-H Kwok, H-K Kwok, L C Tan, W-T Zhao, K M Prakash, W-L Au, et al. 2010. “Analysis of GWAS-Linked Loci in Parkinson Disease Reaffirms PARK16 as a Susceptibility Locus.” *Neurology* 75 (6): 508–12. <https://doi.org/10.1212/WNL.0b013e3181eccfd>.
- Wang, K. S., J. E. Mullersman, and X. F. Liu. 2010. “Family-Based Association Analysis of TheMAPT Gene in Parkinson.” *Journal of Applied Genetics* 51 (4): 509–14. <https://doi.org/10.1007/BF03208881>.
- Watson, Caroline J., A. L. Papula, Gladys Y.P. Poon, Wing H. Wong, Andrew L. Young, Todd E. Druley, Daniel S. Fisher, and Jamie R. Blundell. 2020. “The Evolutionary Dynamics and Fitness Landscape of Clonal Hematopoiesis.” *Science* 367 (6485): 1449–54. <https://doi.org/10.1126/science.aay9333>.
- Williams-Gray, C. H., A. Goris, M. Saiki, T. Foltynie, D. A. S. Compston, S. J. Sawcer, and R. A. Barker. 2009. “Apolipoprotein E Genotype as a Risk Factor for Susceptibility to and Dementia in Parkinson’s Disease.” *Journal of Neurology* 256 (3): 493–98. <https://doi.org/10.1007/s00415-009-0119-8>.
- Xu, Chang. 2018. “A Review of Somatic Single Nucleotide Variant Calling Algorithms for Next-Generation Sequencing Data.” *Computational and Structural Biotechnology Journal* 16: 15. <https://doi.org/10.1016/J.CSBJ.2018.01.003>.
- Yemni, Eman Al, Dorota Monies, Thamer Alkhairallah, Saeed Bohlega, Mohamed Abouelhoda, Amna Magrashi, Abeer Mustafa, et al. 2019. “Integrated Analysis of Whole Exome Sequencing and Copy Number Evaluation in Parkinson’s Disease.” *Scientific Reports* 9 (1): 1–9. <https://doi.org/10.1038/s41598-019-40102-x>.
- Yoshida, Kenichi, Kate H.C. Gowers, Henry Lee-Six, Deepak P. Chandrasekharan, Tim Coorens, Elizabeth F. Maughan, Kathryn Beal, et al. 2020. “Tobacco Smoking and Somatic Mutations in Human Bronchial Epithelium.” *Nature* 578 (7794): 266–72. <https://doi.org/10.1038/s41586-020-1961-1>.
- Zhang, B., S. Kirov, and J. Snoddy. 2005. “WebGestalt: An Integrated System for Exploring Gene Sets in Various Biological Contexts.” *Nucleic Acids Research* 33 (Web Server): W741–48. <https://doi.org/10.1093/nar/gki475>.