Whole-genome sequencing of 50 LRRK2 G2019S carriers discordant for Parkinson's disease

C.Y. McLean[1], E. Drabant[1], E. Harrington[1], C.B. Do[1], N. Eriksson[1]

[1]23andMe, Inc, Mountain View, CA.

Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative disorders, affecting over six million people worldwide. PD symptoms include resting tremor, muscular rigidity, bradykinesia, and postural instability. The pathology of PD is characterized by the loss of dopaminergic neurons in the substantia nigra and is usually accompanied by Lewy bodies, abnormal protein aggregates comprised mainly of alpha synuclein, ubiquitin, and tau proteins. Over 20 single nucleotide variants (SNVs) have been associated with PD susceptibility[1,2]. Of particular interest is the LRRK2 G2019S mutation, with a population frequency in PD patients ranging from under 1% in Asians to 20% in Ashkenazi Jews and 40% in North African Arabs[3]. G2019S is present in both familial autosomal dominant and sporadic PD and has an estimated penetrance of 28% at age 59 and 74% at age 79 (ref. 4). Higher penetrance estimates in familial cases suggest the presence of additional genetic or environmental penetrance modifiers.

Here we present the results of whole-genome sequencing (WGS) of 50 unrelated European individuals concordant for both the presence of the LRRK2 G2019S mutation and the absence of the PD-associated GBA 84GG, N370S, V394L, and R496H mutations. The cohort, drawn from consenting 23andMe customers, comprises 37 individuals affected by PD and 13 healthy controls with no family history of PD. The goals of the project are three-fold: 1) To establish the feasibility of whole-genome sequencing from 23andMe's biobanked saliva samples, 2) Search for high-penetrance genetic modifier mutations that confer protection against PD in G2019S carriers, and 3) Identify additional 23andMe customers for WGS to study both PD genetics and Ashkenazi ancestry.

Methods

At least five µg of DNA was extracted from each biobanked saliva sample. Illumina sequencing generated 100 bp paired-end reads. Read processing followed the Broad Institute's "best practices" guidelines, using BWA[5], GATK[6], Picard[7], and Pindel[8] tools to align and call variants (Fig. 1). Variant annotation was performed with SnpEff[9] and RegulomeDB[10].

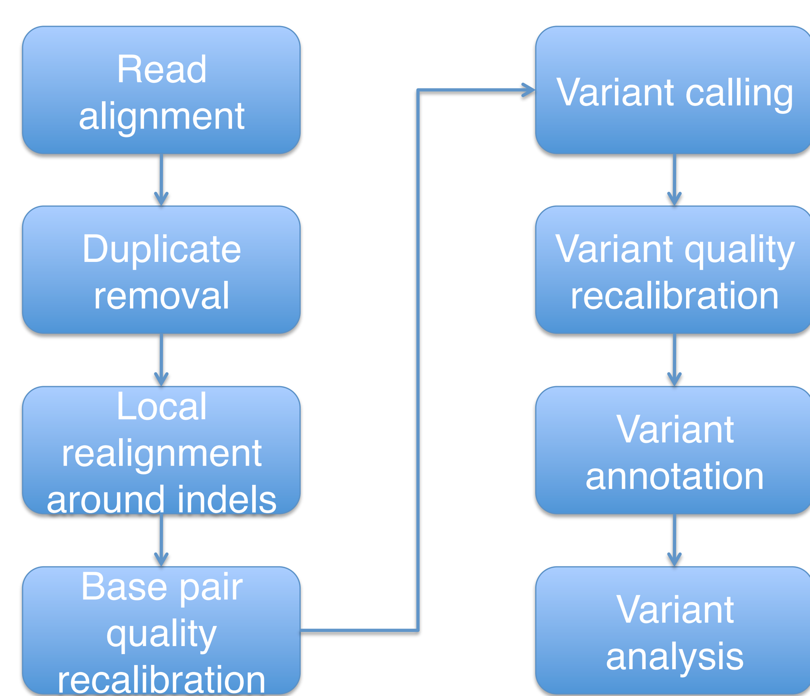

Figure 1. Sequence alignment and variant detection pipeline used.

Haplotype analysis was performed on the 23andMe chip genotypes phased using BEAGLE[11].

Results

The age of onset in PD cases ranges from 38-85 with a median of 56 (Fig. 2A). Individuals were sequenced to a median mapped depth of 44.9-fold coverage (Fig. 2B) with the percentage of genome sequenced ranging from 97.8-98.2%. The number of SNVs and insertion/deletion variants (indels) in each genome agree with previous estimates of individual human variation[12] (Fig. 2C, Table 1). Concordance of WGS-called SNVs and indels with 23andMe custom genotyping chip calls ranged from 99.91-99.97% per individual.

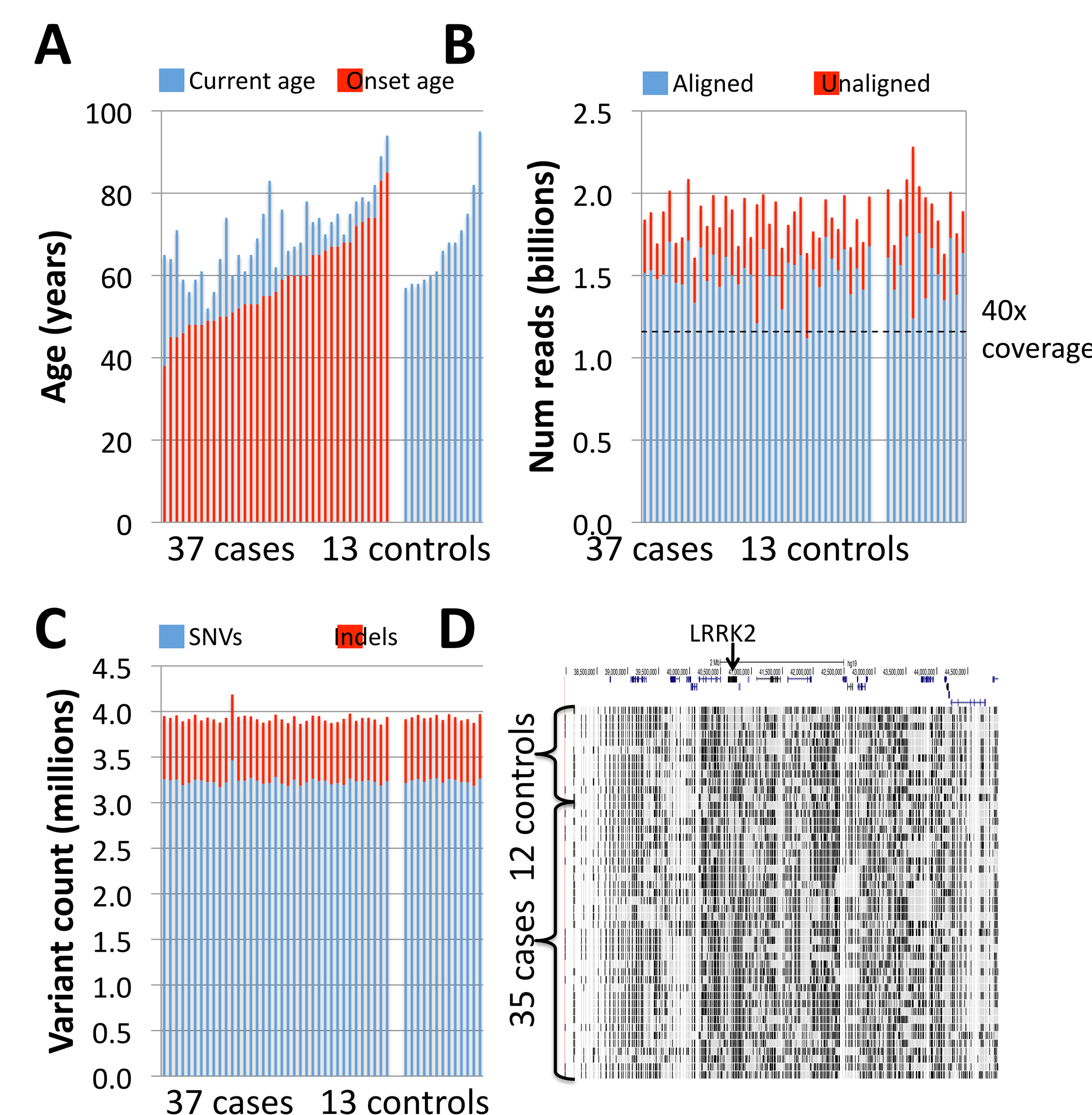

Figure 2. PD cohort statistics. A) Cohort age distribution. B) Sequencing depth and accuracy. C) SNV and indel counts. D) Phased non-G2019S haplotypes of a 7 Mb region surrounding *LRRK2*. Reference (alternate) nucleotides shown in gray (black).

Table 1. Variation counts identified in PD cohort.

Variation type	Unique	Mean per genome	Stdev per genome	Range
SNVs	11,019,180	3,236,705	41,761	3,172,144-3,463,691
Indels	2,313,321	691,649	16,212	658,454-724,046
Inversions	7,076	728	77	531-870
Long insertions	15,171	1,221	114	960-1,541
Tandem duplications	33,070	4,333	535	3,083-5,203

Because LRRK2 autophosphorylates itself both *in cis* and *in trans*[13], we examined the non-G2019S haplotype of LRRK2 (Fig. 2D) for its ability to discriminate PD cases from controls in 195 European individuals (81 cases, 114 controls) including 47 of the WGS cohort samples. However, neither logistic regression nor a support vector machine using a haplotype-based string kernel (Durand, E., unpublished) could reliably discriminate cases from controls.

PD-associated genes are interesting candidates for harboring modifier mutations. However, the mutational burden in protein-coding exons of PD-associated genes is similar in cases and controls in this cohort ($P > 0.05$ for all genes using Fisher's Exact test) (Table 2).

Table 2. Coding exon SNVs + indels in PD-associated genes[1,2].

Gene name	Vars per case	Vars per control		Gene name	Vars per case	Vars per control
ACMSD	0	0		*RAI1*	2.83	3.08
ATP13A2	4.43	3.38		*RIT2*	0.80	0.69
BST1	0.63	0.31		*SCARB2*	0	0
CCDC62	0.14	0.15		*SGK1*	0.34	0.23
FAM47e	2.09	1.85		*SLC41A1*	1.23	1.08
GAK	3.34	3.46		*SNCA*	0	0
GBA	0.03	0		*SREBF1*	0.11	0.15
LRRK2	9.69	9.62		*STK39*	0.09	0.15
MAPT	3.89	4.85		*SYT11*	1.97	2.00
MCCC1	2.00	2.00		*UCHL1*	0.26	0.31
NUCKS1	0	0		*UCP2*	0.51	0.85
PARK2	0.51	0.92		*UCP4*	1.26	1.08
PARK7	0	0.15		*USP24*	2.26	2.38
PINK1	0.80	0.85		*VPS35*	1.00	1.00

Results (cont.)

A genome-wide search in 339 Europeans possessing LRRK2 G2019S for SNVs associated with PD status yielded no significant results. Nonetheless, high-penetrance variants present only in controls are candidates for counteracting the effects of the gain-of-function LRRK2 G2019S mutation. However, no premature stop codon mutations absent in all 37 WGS cases are present in more than one control. Likely-regulatory mutations absent in all cases are present in at most three controls (Table 3).

Table 3. Variants with both expression quantitative trait locus and transcription factor binding evidence for regulatory function[10] absent in all 37 cases and present in 3 controls.

rsid	Hg19 position	Affected gene	Variant location	Notes
rs2304335	chr2:225751103	*DOCK10*	Intronic	Upregulated in idiopathic PD[14]
rs9882524	chr3:66492452	*SLC25A26*	Intronic	
rs35600708	chr9:97356243	*FBP1*	Upstream	Accumulates in postmortem PD brain[15]
rs7963928	chr12:92706348	*EEA1*	Intergenic	
rs3027288	chr17:8007416	*KCTD11*	Intronic	

The LRRK2 G2019S mutation is present at high frequency in Ashkenazi Jews[3]. The 23andMe database includes 283 Ashkenazi individuals who possess at least one of the LRRK2 G2019S or GBA 84GG, N370S, V394L, or R496H mutations. Identity-by-descent segments between these 283 sequencing candidates and the ~8,000 Ashkenazi individuals in the database shows the effective genome information obtained through imputation (Fig. 3).

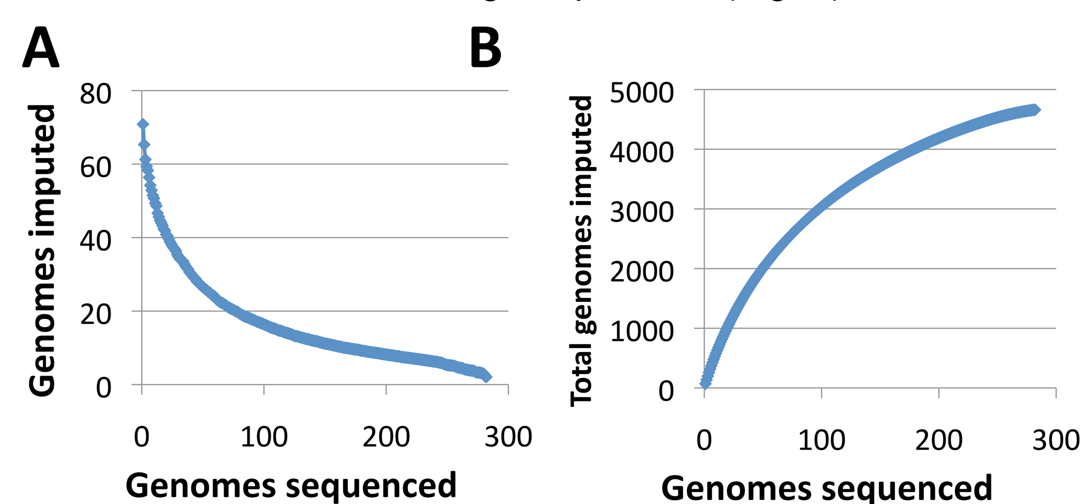

Figure 3. Imputation of genomic sequence data for Ashkenazi Jewish individuals. A) The number of additional genomes that can be imputed per genome sequenced. Calculations used a greedy algorithm to maximize imputed coverage, accounting for previously-sequenced individuals. B) Cumulative number of genomes imputed using the strategy in A.

Discussion

The high concordance of variants identified by WGS with the 23andMe custom genotyping chip and the similarity of the detected variation spectrum with known human variation demonstrate the feasibility of WGS from biobanked saliva samples. While this study did not identify genome-wide significant variants modifying the effects of LRRK2 G2019S in PD, a small number of variants are suggestive of PD-related function. Targeted WGS of Ashkenazi individuals can be used to increase power for future PD research and provide insight into genetic variation in the Ashkenazi population.

Acknowledgments

We thank 23andMe customers who consented to participate in research for enabling this study. We also thank employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

References and Resources

1. Do, C.B. *et al.*, *PLoS Genetics*, 7:e1002141, 2011.
2. International Parkinson's Disease Genomics Consortium and Wellcome Trust Case Control Consortium 2, *PLoS Genetics*, 7:e1002142, 2011.
3. Correia Guedes, L. *et al.*, *Parkinsonism and Related Disorders*, 16:237-242, 2010.
4. Healy, D.G. *et al.*, *Lancet Neurology,* 7:583-590, 2008.
5. Li, H. and Durbin, R., *Bioinformatics*, 25:1754-1760, 2009.
6. DePristo, M. *et al.*, *Nature Genetics*, 43(5):491-498, 2011.
7. Picard website, http://picard.sourceforge.net
8. Ye, K. *et al.*, *Bioinformatics*, 25:2865-2871, 2009.
9. Cingolani, P. *et al.*, *Fly*, 6:1-13, 2012.
10. Boyle, A.P. *et al.*, *Genome Research*, 22:1790-1797, 2012.
11. Browning, B.L. and Browning, S.R., *American Journal of Human Genetics*, 88:173-182, 2011.
12. The 1000 Genomes Project Consortium, *Nature*, 467:1061-1073, 2010.
13. Luzón-Toro, B. et *al.*, *Human Molecular Genetics*, 16:2031-2039, 2007.
14. Durrenberger, P.F. *et al.*, *Parkinson's Disease*, 2012:214714, 2012.
15. Ko, H.S. *et al.*, *Journal of Neuroscience*, 25:7968-7978, 2005.

Copyright © 2012 23andMe, Inc. All rights reserved.